

MADS JUHL STORR-HANSEN

IMPROVEMENT OF PHOTON PID THROUGH MACHINE LEARNING

Copyright © 2022 Mads Juhl Storr-Hansen
Msc. Thesis

NIELS BOHR INSTITUTE, UNIVERSITY OF COPENHAGEN

First printing, January 2022

Abstract

This thesis presents tree-based Machine Learning models to improve the cut based selection of $H \rightarrow Z(\rightarrow ll)\gamma$ in [1] *A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector* and $H \rightarrow \gamma\gamma$ in [40] *Measurement of the properties of Higgs boson production at $\sqrt{s}13\text{TeV}$ in the $H \rightarrow \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment.*

Several models have been created. Particle identification (PID) models for the single leptons. electrons and muons, and photons had an improvement of 19.0%(1.1%) for the electrons(muons) compared to the Loose (Medium) Working Points used by ATLAS. The photons had an improvement of 40.0% when compared to the Tight WP. The isolation models for the same particles obtained an improvement of 21.6%(15.3%) for the electrons(muons) compared to the cuts used by ATLAS. The photon isolation model showed an improvement of 39.9% compared to the Tight WP.

The models for $Z \rightarrow ll$ had an improvement of their True Positive Rate (TPR) of 66.8%(12.6%) for $Z \rightarrow ee(Z \rightarrow \mu\mu)$ when evaluated against the selection of [1] - requiring the same False Positive Rate (FPR). The model for $H \rightarrow \gamma\gamma$ obtained an improvement of 2.0% when evaluated against the selection cuts of [40].

Last of the models $Z \rightarrow ll\gamma$ had an improvement of 178.3%(189.2%) when compared to the cuts used in [1].

The models were also compared to the ATLAS cuts on a MC data with known signal and background, requiring the models and the ATLAS cuts having the same amount of background. It was here found that the $Z \rightarrow ee\gamma(\mu\mu\gamma)$ models had an improvement of 19.1%(8.8%) using the di-lepton models, and an improvement of 24.8%(36.1%) using the photon PID model for the cuts. If these models were combined would an improvement of 64.9%(68.0%) be achieved, while using the full models for $ll\gamma$ would result in an improvement of 128.1%(137.3%).

Doing the same on a $Z \rightarrow \mu\mu\gamma$ data set with real data resulted in an improvement of 0.7%, while the photon model performed 0.5% worse than the ATLAS cuts. Using both of these models resulted in 5.0% worse results than ATLAS, and using the combined model gave 2.5% more events than ATLAS.

The models for $H \rightarrow \gamma\gamma$ were also compared to the ATLAS cuts given known MC data, where the model using a photon pair had an improvement of 21.7% compared to ATLAS. If the single photon models were used, was an improvement 10.5%(9.6%) using different(same) PID scores for the two photons.

Introduction

This thesis is written in an attempt to improve the photon identification used by ATLAS. The current method ATLAS uses is a cut based method. This method requires single particles to pass loose, medium or tight working points (WP), the working points restrict the signal efficiency - a tight WP has a lower signal efficiency than a medium or loose WP¹. When reconstructing particles, it will be required of the reconstructed particles to pass several variable cuts.

These cuts allows ATLAS to quickly select events and remove background, improvements of these selections are expected when using Machine Learning (ML) models instead of cuts, this was found by previous students (among others) Helle Leerberg[32] and Sara Dahl[45]. Before attempts at improving the Particle IDentification (PID), an introduction to Particle Physics, the ATLAS detector and Machine Learning (ML) will be in order. The goal of these chapters is to give an introduction which can be read and understood by a physics student. After the introduction of these topic, the models created will be introduced. These models are trained using Boosted Decision Trees (BDT) algorithm called Light Gradient Boosting Machine (LightGBM, or LGBM). The models trained with this algorithm are isolation and identification models for electrons, muons and photons, and models that combine these single particle models for reconstruction of events: $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, $H \rightarrow \gamma\gamma$, $Z \rightarrow ee\gamma$ and $Z \rightarrow \mu\mu\gamma$. These models are evaluated against the working point used by ATLAS, and requiring the same False Positive Rate (FPR).

Lastly, the model are compared to the ATLAS cuts by comparing the results obtained by applying them to a set of known data. The trained models are compared to the ATLAS cuts by requiring the same amount of background for the ATLAS cuts and the models.

A main reason for these improvements is to improve the measurements of rare Higgs decays, for this thesis $H \rightarrow Z\gamma$ and $H \rightarrow \gamma^{(*)}\gamma$.

¹ The idea is to remove more background than signal.

Disclaimer

This thesis is a successor of the work done by previous Master Theses, each with a different focus. Since the previous students had a larger focus on the leptons will the focus of this thesis be on the photons, more precisely the identification model.

To do this I have trained new models for electrons, muons and photons, as well as their pairs and $ll\gamma$. The work done in this thesis has relied on the analysis of the previous master students, with some pre existing code borrowed and changed to fit with my own data analysis. This thesis is the conclusion of my Masters in High Energy Physics. I had little to no experience with machine learning and decision trees before I started working on this thesis. It has been fun and challenging experience working on this.

Acknowledgment

I would like to first and foremost thank my supervisor Troels Petersen for all his brilliant ideas and enthusiasm on which ideas to follow. I would also thank him for supporting me throughout this thesis and allowing me the time I needed during the many months spent on this thesis. I would also like to thank Benjamin Henckel, Helle Leerberg and Sara Dahl for helping me get started and during my thesis. Lastly, I would like to thank Aske Rosted, Malte Algren and Rasmus Ørsøe.

Contents

1	<i>Particle Physics Theory</i>	1
1.1	<i>The Standard Model</i>	1
1.2	<i>Constituents of baryons and mesons</i>	6
1.3	<i>Interactions of interest</i>	7
1.4	<i>Feynman Diagrams</i>	11
1.5	<i>Cross section</i>	11
1.6	<i>Luminosity</i>	13
1.7	<i>Pileup</i>	14
1.8	<i>Exploration of $H \rightarrow Z\gamma$ and $H \rightarrow \gamma^*\gamma$ and improvement of $H \rightarrow \gamma\gamma$</i>	15
2	<i>The ATLAS detector</i>	16
2.1	<i>ATLAS coordinate system</i>	17
2.2	<i>Particle interaction in matter</i>	18
2.3	<i>Inner Detector</i>	21
2.4	<i>Calorimeter</i>	25
2.5	<i>Muon Spectrometer</i>	27
2.6	<i>Magnet system</i>	28
2.7	<i>Trigger and Data Acquisition System</i>	28
3	<i>Machine Learning</i>	30
3.1	<i>Learning Problem</i>	30
3.2	<i>Learning Algorithms</i>	32
3.3	<i>Machine Learning models</i>	35
4	<i>Data processing and selection</i>	39
4.1	<i>Data files and derivation</i>	39
4.2	<i>Ntuple production</i>	39

5	<i>Monte Carlo models</i>	41
5.1	<i>Single lepton and photons</i>	45
5.2	<i>Lepton and photon pairs</i>	58
5.3	<i>llγ model and data selection</i>	68
6	<i>Reconstruction</i>	73
6.1	<i>Z → llγ MC models</i>	75
6.2	<i>H → $\gamma\gamma$ MC models</i>	78
6.3	<i>Z → $\mu\mu\gamma$ Data models</i>	80
7	<i>Conclusion</i>	82
7.1	<i>Future work</i>	83
A	<i>Appendix</i>	90
A.1	<i>Datasets</i>	90
A.2	<i>Invariant mass and transverse energy/momenta plots</i>	90
A.3	<i>Reweighted single particle plots</i>	90
A.4	<i>Reweighted particle pair weights</i>	90
A.5	<i>Background types for Z → llγ</i>	90
A.6	<i>Reweigh for Z → llγ</i>	98

List of Figures

- 1.1 The Standard Model particles, source <https://cds.cern.ch/record/1473657>. 1
- 1.2 Visualization of the protons contents and corresponding PDF, as a function of Q^2 . Taken from [25]. 6
- 1.3 PDF of a proton given at two energies, the first example is given at a center of mass energy of slightly more than 3 GeV, the second example is at a 100 GeV. In both examples can it be seen that gluons carries most of the momentum at low momentum fractions. Where the distribution at higher momentum fractions are dominated by the valence quarks. 7
- 1.4 The branching ratios for the different Z boson decays. It should here be noted that the "down" type quarks all have a branching ratio of 15.2%, while the "up" type doesn't include the top quark, and the up and charm has a branching ratio of 11.8%. There are also other decays with much lower branching ratios than what is shown here, those of interest for this thesis is $Z \rightarrow ll\gamma$ with a branching ratio of $\Gamma_{Z \rightarrow ee\gamma} = 0.052\%$ and $\Gamma_{Z \rightarrow \mu\mu\gamma} = 0.056\%$. 8
- 1.5 The different branching ratios for the Higgs particle, see also tab 1.3. 9
- 1.6 Feynman diagrams of the four main production processes, at the LHC. 10
- 1.7 Predicted branching ratio for the Higgs particle as a function of its mass. Keep in mind that the mass of the Higgs boson is 124.97 ± 0.24 GeV. Noted with a black line. 10
- 1.8 Feynman diagram of the annihilation of a lepton pair, which creates a virtual Z-boson or a virtual photon which decays to a new lepton pair. 11
- 1.9 The different particles and propagators in a Feynman diagram. 11
- 1.10 Feynman diagrams for s-, t- and u-channel processes. It should be noted that the u-channel is only applicable then there are identical particles in the final state. 13
- 1.11 Example of pileup in an interaction, where the interaction of interest is at $t=0$. This image is from the CMS experiment, but this would also be applicable for the ATLAS experiment[14]. 15
- 2.1 The Large Hadron Collider and its experiments. 16

- 2.2 Left: The ATLAS detector including its' different subdetector parts. Right: The two coordinate systems of the ATLAS detector[4]. 17
- 2.3 Material budget for the ATLAS ID, with distinctions between which subdetector the electron passes through. The forward direction ($|\eta| > 0.8$) has more material, and as a result, more photons convert here. 21
- 2.4 The Inner detector with its subdetectors: The Pixel detector, Transition radiation tracker and semiconductor tracker. 22
- 2.5 The high threshold probability for different particles in test beam data - as a function of Lorentz gamma factor[7]. 24
- 2.6 Sketch of the Electromagnetic Calorimeter. 25
- 2.7 An Electromagnetic shower created due to pair production and bremsstrahlung. 26
- 2.8 The ATLAS LAr calorimeter with simulated electron shower in red. 26
- 2.9 Graphical representation of the muon spectrometer, and detector parts. 27

- 3.1 How Machine Learning can be divided into Supervised and Unsupervised learning, and how Supervised learning can be divided into classification and regression. 32
- 3.2 Graphical representation of the difference between level-wise and leaf-wise tree growth. Examples of algorithms using these methods of growing would be XGBoost using the level-wise tree growth and LightGBM using the leaf-wise tree growth. 37

- 5.1 The ROC-curve for the different training tests. "phoPidReg40" is the one called test 0 in table 5.1. The x-axis shows the range $0.5 - 1$ and the y-axis show the range $0 - 0.5$. 44
- 5.2 Comparison of the ROC-curves of the initial guess model and test 2 model without the pho_core57cellsEnergyCorrection variable. 45
- 5.3 The transverse energy for the electrons chosen with the Pid method. The signal is displayed in blue, while the background is displayed in orange. 46
- 5.4 Transverse momentum of Pid muons, with the signal displayed in blue and background displayed in orange. 47
- 5.5 Transverse momentum of Iso muons, with the signal displayed in blue and background displayed in orange. 47
- 5.6 Transverse energy of Pid photons. 48
- 5.7 Signal and background distribution of $\langle \mu \rangle$, e_T and η for identified electron reweighed using different number of estimators. 50
- 5.8 shap values obtained from training the PID electron models. 51
- 5.9 Roc curves plotted for test sets of the electron models, with FPR and TPR values from the the ATLAS cut and their corresponding values in the the models. 53
- 5.10 SHAP values obtained from training the muon models. 55

5.11	Roc curves and the FPR and TPR values from ATLAS for the muon models.	56
5.12	SHAP values obtained from training the photon models.	56
5.13	Roc curves and the FPR and TPR values obtained from ATLAS cuts, for ISO (left) and PID (right) photons.	58
5.14	Distribution of invariant mass for the di-leptons.	59
5.15	The invariant mass of the photon pairs. With the Higgs peak at 125GeV, most of the background has a low mass, the small peak at ≈ 0 GeV arises due to conversions.	60
5.16	Plots of the signal and background distributions for the pair selections, and their reweighed values in $\langle \mu \rangle$, E_T/p_T , μ and $\mathcal{M}_{ll/\gamma\gamma}$. Larger plots of this can be seen in the appendix in section A.4.	62
5.17	Electron pair SHAP values, with the PID score for the secondary electron being the most impactful.	64
5.18	Muon pair SHAP values, with the PID score for the secondary muon being the most impactful.	65
5.19	Photon pair SHAP values, with the PID score for the secondary photon being the most impactful.	65
5.20	Visual presentation of where the ATLAS FPR and TPR values are positioned compared to the ROC-curve generated by the prediction of photon pairs.	67
5.21	ROC-curves for the electron and muon pairs, including the ATLAS working points.	67
5.22	The two $ll\gamma$ invariant mass distributions. With signal in blue and background in orange. The huge peak at 100GeV for $ee\gamma$ is due to electron pairs matched with a random photon.	68
5.23	Logarithmic plot of the invariant mass of $ee\gamma$ extended to masses to take an extended look at the signal/background distributions.	69
5.24	Plots of the signal and background distributions for the isolation selections, and their reweighed values in $\langle \mu \rangle$, E_T/p_T and μ . Larger plots of this can be seen in the appendix in section A.6.	70
5.25	SHAP values for the lepton pairs and their corresponding photon using combined variables from lepton pairs and the isolation and identification scores from the single photon models.	72
5.26	ROC-curve for the two $ll\gamma$ models, $ee\gamma$ left and $\mu\mu\gamma$ right.	72
6.1	The two distributions used for the ATLAS cuts and the corresponding selections using the models generated in the previous chapter.	76
6.2	The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background for the models and ATLAS.	77
6.3	The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background for the models and ATLAS.	77

- 6.4 The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background in the energy range 121 – 129GeV for the models and ATLAS. 79
- 6.5 The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background in the energy range 60 – 80GeV and 100 – 140GeV for the models and ATLAS. 81

List of Tables

1.1	The three different colours in QCD and their corresponding colour isospin (I_3^C) and colour hypercharge (Y^C) values.	4
1.2	The relationship between S.I. and natural units.	5
1.3	Branching ratio of the Higgs particle.	10
1.4	Values of particles and propagators in Feynman diagrams, $u(p)$ for a fermion is the spinor while ϵ_μ for the photons and the gluons are polarization.	11
1.5	Production rate of the different Higgs production channels during Run-2 at the LHC.	14
4.1	Different derivations used in this thesis. EGAM derivations are produced for specific decays, while MUON and HIGG derivations are using different skimming events.	39
4.2	The different processes, the derivations used for them, and which models and reconstructions they are used for.	40
5.1	The variables used for the various versions of the photon Pid model. Test # 0 is the initial guess.	43
5.2	The number of signal and background particle in the lepton and photon data sets, before and after removing the leptons (photons) with less than 4.5GeV (9.5GeV). It should be noted here that the background for the isolated muons have been increased by a factor of 93, from 101.331 to 9.423.783 before the cuts.	49
5.3	Isolation variables for the electron[48].	51
5.4	Pid variables for the electrons[17].	52
5.5	The false positive and true positive rates for ATLAS cuts and the models used in this model (referred to as LGBM). Besides the values of the FPR and TPR are the numeric difference between this thesis' models and the ATLAS cut in FPR given, and the percentage difference between the TPR are also given.	53
5.6	The unweighted and weighted AUC for the two electron models.	53
5.7	Isolation variables for the muon[48].	54
5.8	Pid variables for the muon[17].	54
5.9	AUC values for the muon and isolation models.	55

5.10 The false positive and true positive rates for ATLAS cuts and the models used in this model (referred to as LGBM). Besides the values of the FPR and TPR are the numeric difference between this thesis' models and the ATLAS cut in FPR given, and the percentage difference between the TPR are also given.	55
5.11 Isolation variables for the photon[48].	56
5.12 Pid variables for the photon[17].	57
5.13 Weighted and Unweighted AUC values for the two photon models.	58
5.14 The TPR and FPR for the ATLAS cuts and the improvements achieved by the photon models.	58
5.15 The number of signal and background particle in the different data set, before and after using a minimum invariant mass cut of 50GeV (80GeV) for leptons (photons), and after using a maximum invariant mass cut of 150GeV.	61
5.16 Variables used for the electron pairs.	63
5.17 variables used for the muon pairs.	64
5.18 Variables used for the photon pairs.	64
5.19 TPR and FPR of the pair models, and the difference between the values obtained with the models and the ones obtained by ATLAS.	66
5.20 The Area under the curve, with and without weights for the pair models.	68
5.21 The number of signal and background events in the $Z \rightarrow ee\gamma$ and $Z \rightarrow \mu\mu\gamma$ data sets, before and after removing the events with an invariant mass less than 50GeV, or higher than 150GeV .	69
5.22 Variables used for the $ee\gamma$ model.	71
5.23 variables used for the $\mu\mu\gamma$ model.	71
5.24 FPR and TPR for the two $ll\gamma$ models, and the improvements achieved by the LGBM models compared to the ATLAS cuts.	71
5.25 AUC values for the $ee\gamma$ and $\mu\mu\gamma$ models.	72
 6.1 Selections from [1] used to evaluate ATLAS.	73
6.2 Selections from [40] used to evaluate ATLAS.	74
6.3 Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.	78
6.4 Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.	79
6.5 Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.	80

1

Particle Physics Theory

1.1 The Standard Model

Most of this chapter will be based on the following three sources[54][36][53]. If the general public was asked, what the universe consists of, an answer similar to what the physics community gave in the 1930's would be given. That the matter in the universe is made of three types of particles: protons, neutrons and electrons. Some people might also mention light, also known as photons, and even fewer people would mention neutrinos, where they would only think of the electron neutrino.

However, our knowledge of the universe has been expanded in the last 90 years, it is now known that protons and neutrons are divisible. The universe as it is known today, is best described using what is called the Standard Model, the particle in this model is shown in fig 1.1.

As it can be seen in fig 1.1 there are no protons or neutrons included.

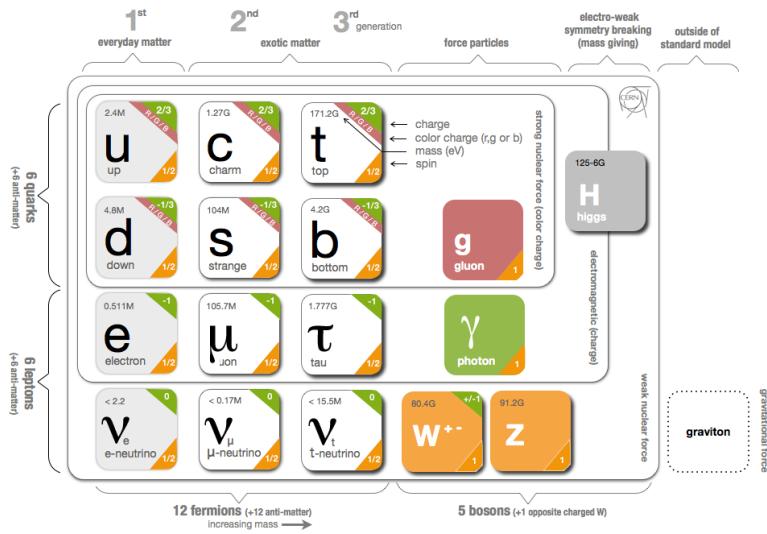


Figure 1.1: The Standard Model particles, source <https://cds.cern.ch/record/1473657>.

The photon and the electron and its corresponding neutrino are the

only ones included from the answer above. There are 12 fermions, five bosons, and a theoretical graviton. The graviton cannot be included since it so far has been impossible describe gravity using the Standard Model. Lastly, three forces are included: The strong nuclear force, the electromagnetic force and the weak nuclear force.

The different elementary particles can be described using their mass, their electric charge and their spin, which is a permanent angular momentum possessed by all particles. The latter is, however, only able to take certain values, the maximum value the spin angular momentum can have about any axis is given by $s\hbar$, with $\hbar \equiv h/2\pi = 1.0545718 \cdot 10^{-34} \text{m}^2\text{kg}/\text{s}$. The value of s is restricted to half-integer values $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$, these are the maximum values s can take; the spin can take any value of s from $-s$ to s jumping with integer steps. This gives rise to two kinds of elementary particles, elementary particles with half-integer spin and elementary particles with integer spin.

FERMIONS: The first group is the one with half-integer spin, which consists of two subgroups: quarks and leptons. A distinct feature for fermions compared to the bosons is the exclusion principle: which dictates that no two fermions can be described by the same quantum numbers. And since the particles we are made of are consisting of three fermions¹ these particles are also fermions. The total spin of a proton can be given by $\frac{1}{2} \pm \frac{1}{2} \pm \frac{1}{2} = \frac{1}{2} \vee \frac{3}{2}$ ², this principle also applies to us. An example could be that we are unable to walk through a wall without either moving or destroying the wall in the process. Besides hadrons, there are also particles called mesons, e.g. a pion, where these particle consist of two quarks³, and their total spin angular momentum is integer, and these particles are bosonic.

The twelve fermions, which are elementary particles can be divided into quarks and leptons. For both of these classes exist an upper and a lower state, of three pairs (generations). With the "up-type" quarks consisting of, in order of their generation, up, charm and top. The "down-type" quarks include down, strange and bottom. The leptonic version is similar, with the up-type being electrons, muons and tau; the down-types are here named similar to the up-types except for them including the name neutrino, e.g. electron neutrino.

QUARKS: This type of particle tends to form other particles - hadrons and mesons, but there is slightly more to it than that. Due to the nature of the quarks they are not individually observable; they can only be found in their bound state⁴. As noted, particles are distinct from each other in their attributes. For all fermions this results in a higher mass for higher generation - the mass ordering for up-type quarks are given as $m_u < m_c < m_t$ and likewise for all the other fermions⁵. The charge of a particle or object is normally measured in Coulombs; this is not practical here since the charge of a single particle is of the order of 10^{-19} Coulomb. So the elementary charge is defined as $e = 1.602 \cdot 10^{-19} \text{C}$ and is used to measure the charge of

¹ this could be either quarks or anti quarks. But since the universe is dominated by matter, these quarks would tend to be "real" quarks.

² this is true for all hadrons, particles which are consisting of 3 quarks.

³ a quark q and an anti quark \bar{q} .

⁴ Continued when discussing the strong nuclear force.

⁵ For context is the weight of $m_u = 2.2 \text{MeV}/c^2$ and $m_d = 4.7 \text{MeV}/c^2$

single particles. Up-type quarks have a charge of $+2/3$, while down-type quarks have a charge of $-1/3$. When these combine to form hadrons and mesons, they will always combine to an integer total charge. For hadrons this can be done by, for example, two up and one down for the proton, or two down and one up for the neutron. Likewise, can a meson be constructed by combining a quark and an anti quark⁶ see pions, here π^0 can be constructed by $u\bar{u}$ or $d\bar{d}$.

LEPTONS: These particles tend to be much lighter than their corresponding quarks - from family to family. The electron weighs $0.5\text{MeV}/c^2$, only a factor of 4 – 8 times lower than the two lightest quarks. However, as previously noted, it is not possible to find free quarks, so a comparison with the lightest baryon would be more apt. The lightest baryon is the proton with a weight of $938\text{MeV}/c^2$ and thus weighs 1836 times as much as the electron. Another difference between quarks and leptons is that leptons do not interact through the strong nuclear force, which allows us to observe them freely. However, this observation can be more or less difficult, depending on which types of particle being observed. Since they don't interact with the strong force, and gravitational forces are generally negligible compared to the other forces on this scale, this leaves the electromagnetic force as one of the few methods they could be measured with. However, this is much easier if the particle has a charge, which the neutrinos does not have. This makes it extremely hard to measure neutrinos. In general, neutrinos will be found in ATLAS by measuring the missing transverse energy(E_T) - since energy conservation requires the transverse energy to be the same before and after an interaction, and the other particles can be found in other places of the detector - I will return to this in later sections. While the mass of an electron is low compared to other particles it is still massive compared to the neutrinos. It was previously thought that neutrinos were massless; it is now known that they do indeed have a mass and that this mass is very low and the specific value is not known.

The charge of the leptons are integer values of the elementary charge, unlike the quarks which had fractions of the elementary charge.

GAUGE BOSONS AND THE FORCES: The bosons are different to the fermions. While the fermions had half-integer spin, the bosons have integer spin. Furthermore, while the fermions are the building blocks of the universe, the bosons are the mediators of the different interactions. There are four forces, oh which three are included in the standard model. The forces of interest are the strong nuclear force, the weak nuclear force and the electromagnetic force, where each of these is coupling through different properties of the particles. The last force, which is not included in the standard model, is the gravitational force, which is theorized to be propagated by the theoretical graviton. However, this has not been proven, and it is yet to be reconciled with the rest of the forces.

⁶ An anti particle is identical to the particle in everything but its charge.

The electromagnetic force: Is the most well known of the three forces, it is the first one students are taught of. It couples to the charge of particles, so it couples to the quarks and the electron, muon and tau⁷. It can for example be seen as the visible light, since it is propagated by photons. Since photons are massless, and they don't have any charge, the propagation range of this force unlimited, it does however become weaker with distance $\propto \frac{1}{r^2}$.

The strong nuclear force: The strong nuclear force is what binds quarks together in pair or groups of threes - mesons and hadrons⁸. The bosons which are mediating this force are the gluons, and these couple to the colour charge of particles - including their own colour charge.

Quarks	I_3^C	Y^C	Antiquarks	I_3^C	Y^C
r	1/2	1/3	\bar{r}	-1/2	-1/3
g	-1/2	1/3	\bar{g}	1/2	-1/3
b	0	-2/3	\bar{b}	0	2/3

As it can be seen in table 1.1 there are three different colour charges for a given particle. These colours are characterized by their two colour charges: The colour isospin (I_3^C) and colour hypercharge (Y^C), which can be compared to the electrical charge in electromagnetism. However, it is not possible to measure any particles with colour values different from 0, this phenomena is called colour confinement. There are 8 different gluons, with different combinations of colour.

$$r\bar{g}, g\bar{r}, r\bar{b}, b\bar{r}, g\bar{b}, b\bar{g}, \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}) \quad \text{and} \quad \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$$

And since gluons couples to particles with a colour charge, they will also couple to themselves, this leads to a production of jets. An example of this is the interaction $e^+e^- \rightarrow q\bar{q}$ where quark anti-quark pair initially created in this interaction will separate at high velocities. As they separate, the colour field will be restricted to a tube with an energy density of 1 GeV fm^{-1} . As the quarks separate further, the colour field will at some point reach an amount of energy where it is favourable to creat a new $q\bar{q}$ pair, which will break the colour field into smaller "strings". This process will continue until all $q\bar{q}$ pairs have low enough energy to combine into colourless hadrons. This will lead to two jets of hadrons, one following the initial quark and the other following the initial anti-quark.

The weak nuclear force: Is the force responsible for phenomena such as beta decay and the fusion of hydrogen to helium in the core of the sun. Where the electromagnetic force coupled to the charge of the particle and the strong force coupled to the colour of the particles, then the weak force couples to the weak isospin of the particles - hence it interacts with all 12 elementary particles. The propagator of this interaction are the three bosons W^+ , W^- and Z , which due to

⁷ It does also couple to baryons with charge such as protons, and mesons.

⁸ It should be noted that exotic particles could be discovered, which are combinations of four or five quarks, $qq\bar{q}\bar{q}$ or $qqqq\bar{q}$

Table 1.1: The three different colours in QCD and their corresponding colour isospin (I_3^C) and colour hypercharge (Y^C) values.

their high mass, for W it is 80 GeV and for the Z it is 91 GeV, have a very short range.

1.1.1 Natural units:

In our everyday life, there are some units of measure which are more useful than others. The mass of most objects are not given in solar masses, it is given in kg, height is given in meters not lightyears. However, these units are not optimal for particle physics, since the mass of an electron is $9.11 \cdot 10^{-31}$ kg, and lengths are not always given in meters, since the root-mean-square charge radius of a proton is given as $\langle r^2 \rangle^{1/2} = 0.8 \cdot 10^{-15}$ m, the rest mass of this proton is $1.67 \cdot 10^{-27}$ kg. The unit for area has also been changed, since most particle physics has interaction cross sections in the order of nano- or pico-"barn", where a barn is defined as

$$\text{barn} \equiv 10^{-28} \text{m}^2$$

The examples could continue, but to reduce the number large exponents and work with units that reflects on the scale of the physics, natural units have been introduced.

These natural units are the fundamental constants of quantum mechanics, which replaces [kg, m, s] with [\hbar , c, GeV].

With $\hbar = 1.055 \cdot 10^{-34}$ J·s, c being the speed of light in vacuum $c = 2.998 \cdot 10^8$ m/s, and lastly $\text{GeV} = 1.602 \cdot 10^{-10}$ J, which is much more size relevant, since that is approximately the rest mass of a proton.

However, this can be simplified by setting $\hbar = c = 1$, which can be seen in tab 1.2. This simplifies calculations, since everything will now be given in units of energy. It does also simplify for example the Einstein energy-momentum relation

$$E^2 = p^2 c^2 + m^2 c^4 \quad \text{reduces to} \quad E^2 = p^2 + m^2 \quad (1.1)$$

Where the SI units can be regained by doing dimensional analysis - $\hbar \cdot c \approx 200$ MeV fm is used.

A special case of setting the speed of light to one, is the simplifications of a lot of relativistic calculations - since the original units can be restored by multiplying the result with \hbar and c, by differing amounts depending on the desired unit. But it also makes it easier to write up the γ -factor, which for example describes how the time moves slower at high velocities:

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \quad (1.2)$$

Where $\beta = \frac{v}{c}$. During this thesis there will be references to two different γ , one of which is this factor, and the other one is the photon. Normally when γ is used in a formula, it will be the gamma factor, while the γ in the written interactions is the photon - unless otherwise stated.

Unit	[kg, m, s]	$[\hbar, c, \text{GeV}]$	$\hbar = c = 1$
E	$\text{kg m}^2 \text{s}^{-2}$	GeV	GeV
p	kg m s^{-1}	GeV/c	GeV
m	kg	GeV/c^2	GeV
t	s	$(\text{GeV}/\hbar)^{-1}$	GeV^{-1}
l	m	$(\text{GeV}/\hbar c)^{-1}$	GeV^{-1}
A	m^2	$(\text{GeV}/\hbar c)^{-2}$	GeV^{-2}

Table 1.2: The relationship between S.I. and natural units.

1.2 Constituents of baryons and mesons

In the previous section, it was stated that hadrons had constituents, which are the quarks. But how did physicists realise this earlier? Given the example of a proton, which consists of 2 up quarks and a down quark, where the proton weighs approximately 938 MeV, this is however not the sum of the up quarks, which weigh 2.3 MeV, and the down quark, which weigh 4.8 MeV. The mass is off by a factor of a 100.

As mentioned earlier the (valence) quarks interact through the strong nuclear force, through gluons. These gluons interact with each other and create sea quarks. This leads to a Parton Distribution Functions (PDF). These PDFs describe the probability density for finding a parton with a certain part of the longitudinal momentum. Given the example of a proton, the proton had only consisted of a single particle, which has all the momenta, this would have resulted at a PDF distribution which would be a dirac delta function at $x = 1$ where x is the fraction of momenta the given quark has. If the proton consists of three quarks, each with a third of the momenta, the PDFs would be delta functions at $x = 1/3$. If these three quarks interacted with each other these distributions would be smeared out. Lastly if processes of higher order such as the production of sea quarks which are produced from the gluons, then this would tend to increase the PDFs value at low x -values.

These four examples can be found through electron-proton scattering, which gives us information about the structure of the photon. The first example is at very low energies, where the wavelength of the virtual photon, which is exchanged between the electron and the proton, is large compared to the size of the proton. In this case the elastic scattering will give information regarding the global properties of the proton such as charge radius. For the second example, the energy of the virtual photon will be larger, so the wavelength of the photon is approximately the same size as the proton. In this case the proton will no longer be seen as point like, and it will have extended charge and magnetic momentum distribution. For the third example the energy will be increased further so the wavelength of the virtual photon is relatively small compared to the protons size. In this case the dominant process will be inelastic scattering, where the photon interacts with the constituents of the proton. Lastly if the energy are much higher, so the wavelength of the virtual photon is much shorter than the size of the proton, will the proton appear as a sea of interacting quarks and gluons[54].

Visualization of this can be seen in figure 1.2, where Q^2 is the absolute value of the squared four momentum of the virtual photon (since the four momentum squared of the virtual photon is negative).

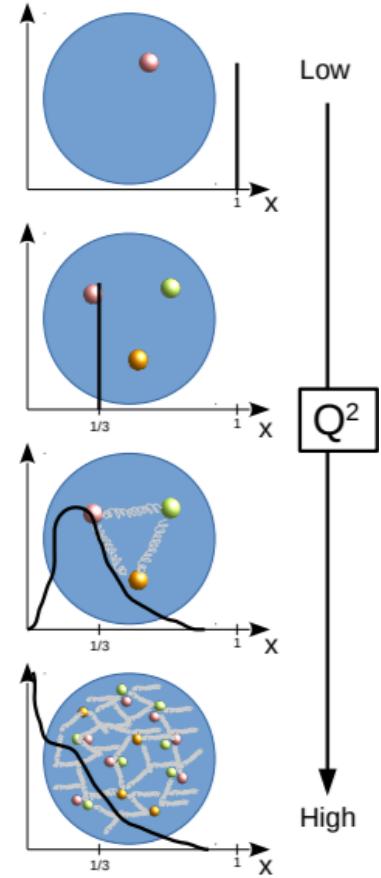


Figure 1.2: Visualization of the proton's contents and corresponding PDF, as a function of Q^2 . Taken from [25].

EXAMPLE: PARTON DISTRIBUTION FUNCTION OF A PROTON:

In figure 1.2 it is only plotted for a single parton. But as mentioned this is however not the case for a hadron, in the case of the proton

which consists of three valence quarks (two up quarks and a down quark), there are also sea quark-pairs and gluons, which also has a part of the momentum.

There are, as it can be seen in figure 1.3, many constituents of the

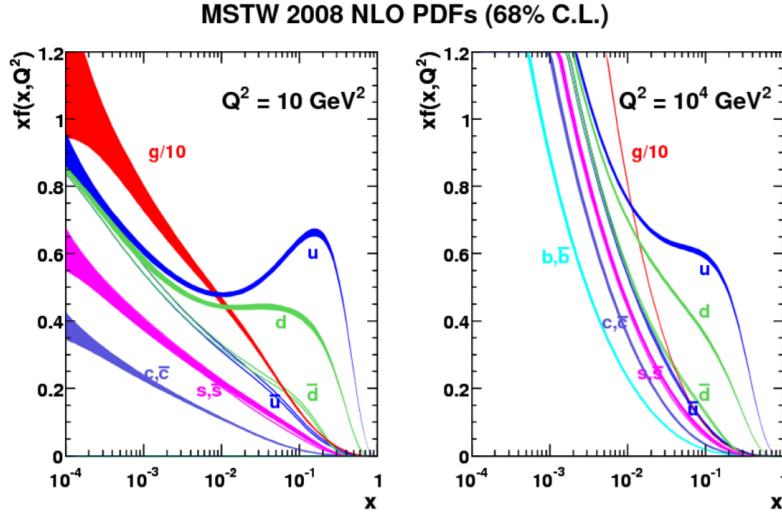


Figure 1.3: PDF of a proton given at two energies, the first example is given at a center of mass energy of slightly more than 3 GeV, the second example is at a 100 GeV. In both examples can it be seen that gluons carries most of the momentum at low momentum fractions. Where the distribution at higher momentum fractions are dominated by the valence quarks.

proton, where the fraction of momentum the different constituents has, is determined by energy. At low momentum fractions, the most common parton is a gluon, while at higher momentum fractions the most probable particles are the valence quarks, which for protons are the up and down quarks, which shouldn't be confused with their corresponding anti-quarks. Figure 1.3 was found at[10].

1.3 Interactions of interest

For this thesis there will be a focus on the Higgs boson and its decays, with a special interest in the photons that are included here. But due to low production rates of Higgs bosons, similar channels from the Z boson will be used⁹.

1.3.1 The Z boson

As mentioned in the section regarding the weak interaction, the Z boson is one of the propagators of this interaction. The total decay width of the Z boson is $\Gamma_Z \approx 2.5 \text{ GeV}$, which gives a half-life of approximately $\frac{\hbar}{2.5 \text{ GeV}} = 2.6 \times 10^{-25} \text{ s}$ [54]. The Z-boson would therefore be able to travel approximately 1 fm at the speed of light before it would decay. It is therefore not possible to do measurements on the Z boson, and there will instead be done measurements on what it decays to. From these can the Z boson be reconstructed. Where W^\pm has a charge, the Z boson has no charge, this results in the sum of the electric charge of the Z bosons decay products having to be 0. If further laws of conservation are included, the decay products will be required to be a particle and their corresponding anti particle.

⁹ Which is especially useful since a Higgs boson may decay to a Z boson.

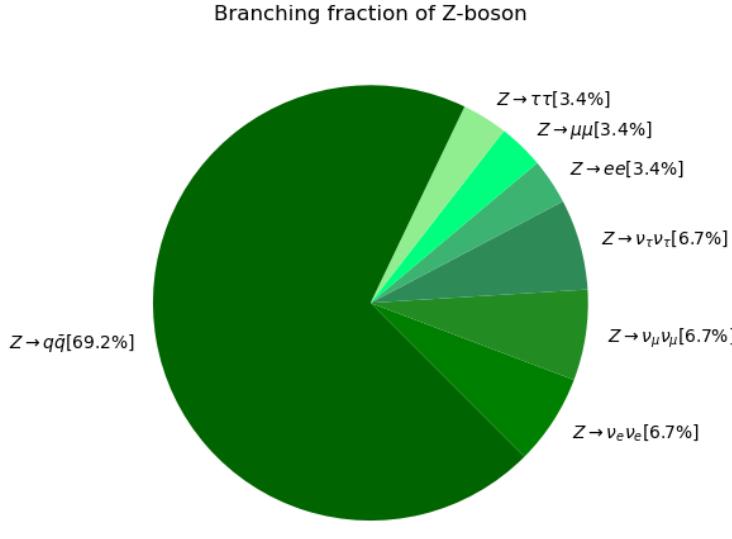


Figure 1.4: The branching ratios for the different Z boson decays. It should here be noted that the "down" type quarks all have a branching ratio of 15.2%, while the "up" type doesn't include the top quark, and the up and charm has a branching ratio of 11.8%. There are also other decays with much lower branching ratios than what is shown here, those of interest for this thesis is $Z \rightarrow ll\gamma$ with a branching ratio of $\Gamma_{Z \rightarrow ee\gamma} = 0.052\%$ and $\Gamma_{Z \rightarrow \mu\mu\gamma} = 0.056\%$.

BRANCHING RATIO: Unstable particles may decay to particles with lower mass; this decay is not always the same decay, a particle may decay to a number of different particles restricted by conservation laws. For example there are lepton, colour, flavour, charge conservation, so if the decay ends with a lepton, the corresponding anti lepton will also have to be present. This leads us to a few possible decay channels[54],[23][57].

1. A Z boson will decay to a charged lepton pair, so either an electron pair, muon pair or tau pair, so three possible decays. This happens approximately at the same rate for all the generations, and happens 10.2% of the time.
2. The Z boson may also decay to the elusive neutral leptons, the neutrinos, of all three generations. These are however extremely hard to measure since they rarely interact with anything, due to their lack of charge. The Z boson decays equally often to these three particles, and it decays to these approximately 20.1% of the time
3. Lastly, the Z boson can decay to a quark pair, where each of these may have one of three colours, this has to be taken into account while doing theoretical calculations. The decays into quarks differ from the leptonic, since they aren't all equal. The Z boson may decay to any of the "down" type quarks at 15.2% each, while of the "up" type, it only decays to the up and charm at 11.8% of the time each. This leads to a total branching ratio for the decay to a pair of quark and anti-quark particles at a 69.2% chance. However, quarks are due to their colour not free, and can only be observed in their bound states - as mentioned earlier, so these will be measured as jets of hadrons.

This is also shown in fig 1.4[57]. These decays are unfortunately not all equally easy to detect. The neutrinos are notoriously hard to detect, and the quarks create jets and form hadrons. Furthermore the decay into a τ pair can also be hard to measure since the lifetime of a tau particle is 2.9×10^{-13} s, and they will decay again before they can be measured[66].

ATLAS and this thesis will therefore focus on the

$$Z \rightarrow l + \bar{l} \quad \text{and} \quad Z \rightarrow l + \bar{l} + \gamma \quad (1.3)$$

Where l can be either e or μ and the γ is a photon.

1.3.2 The Higgs boson

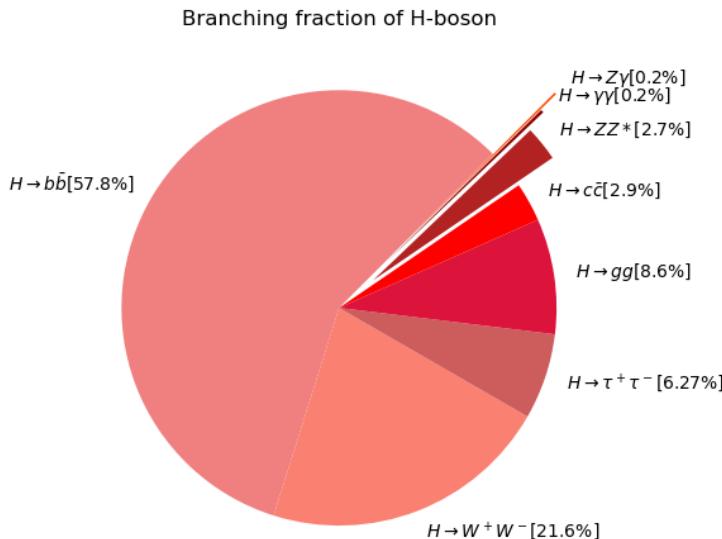


Figure 1.5: The different branching ratios for the Higgs particle, see also tab 1.3.

During the 1970's a connection between two of the four fundamental forces was established - the weak force and the electromagnetic force. This implied an underlying force which unified electricity, magnetism, light and certain types of radioactivity - which was called the electroweak force.

However, there was a small issue with this theory, as it predicted that the mass of the force-carrying particles, the photon and the W^\pm and Z bosons, were massless. This is wrong in the case of the W^\pm and Z bosons, since they are almost a 100 times heavier than the proton.

To solve this issue Robert Brout, François Englert and Peter Higgs proposed what is now known as the Brout-Englert-Higgs mechanism[52]. This mechanism allowed the W^\pm and Z bosons to interact with an invisible field, from which they obtained mass. This does also extend past these two bosons, all subatomic particles get their mass due to interactions with the Higgs field, their mass is determined by how much they interact with the Higgs field¹⁰.

The Higgs field acts slightly different to the other fields, where other

¹⁰ Neutrinos may be an exception to this, this is yet to be confirmed.

fields are measured by either letting particles interact with the field, think an electron being deflected in an electromagnetic field, or by producing a quantum particle - which in this case is the photon. But it is not possible to measure particles interacting with the Higgs field, which leaves us with producing the Higgs particle and learn of its properties[51]. So far has it been found that the mass of the Higgs particle is 124.97 ± 0.24 GeV, its width is found to be less than 14.4 MeV and lastly it has spin 0[27].

PRODUCTION CHANNELS: When discussing the Higgs production mechanisms, there are four processes normally taken into account. These can all be seen in fig 1.6[28]¹¹.

The largest of these four production channels are the gluon-gluon fusion (ggF), which produces the Higgs particle through quantum loops. The next most common production channel is the vector boson fusion (VBF), this production channel is characterized by two forward jets. Then there are vector boson associated production (VH), where a Higgs particle is produced in association with either a W^\pm or Z boson (also known as Higgs strahlung). Lastly there are t̄t associated production (t̄tH), this process is however very rare. For this thesis only ggF and VBF production channels will be taken into account.

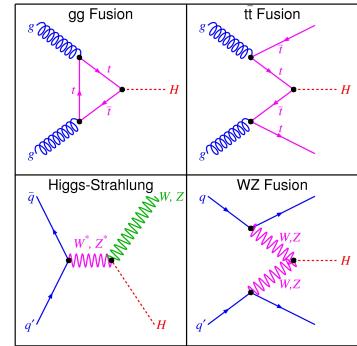


Figure 1.6: Feynman diagrams of the four main production processes, at the LHC.

¹¹ Feynman diagrams will be introduced in the next section

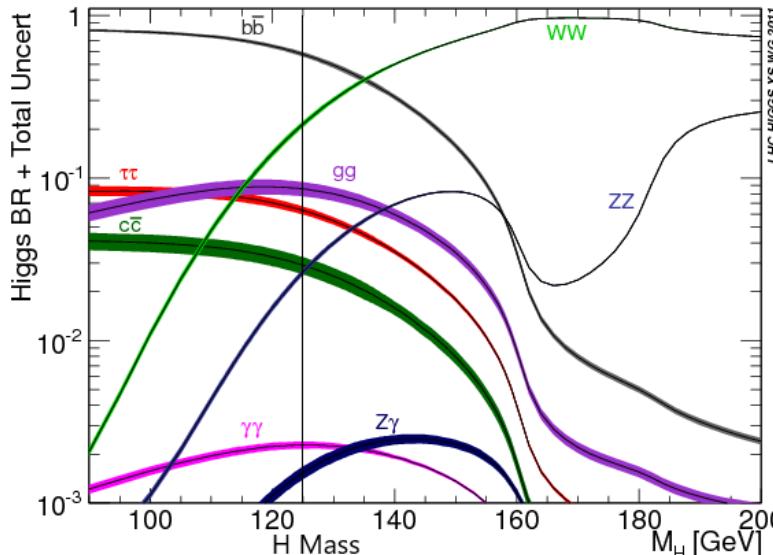


Figure 1.7: Predicted branching ratio for the Higgs particle as a function of its mass. Keep in mind that the mass of the Higgs boson is 124.97 ± 0.24 GeV. Noted with a black line.

BRANCHING RATIO: The branching ratio for the Higgs particle was predicted to vary with its mass, the predicted distribution is shown in fig 1.7[28].

Given the Higgs mass of ≈ 125 GeV the branching ratio can be found as in table 1.3 and fig 1.5[54][28][27]. Though the branching ratio for $H \rightarrow \gamma\gamma$ and $H \rightarrow Z\gamma$ are the two lowest values, of those displayed, are these the ones focused on in this thesis.

Decay mode	Branching ratio
$H \rightarrow b\bar{b}$	57.8%
$H \rightarrow WW^*$	21.6%
$H \rightarrow \tau^+\tau^-$	6.4%
$H \rightarrow gg$	8.6%
$H \rightarrow c\bar{c}$	2.9%
$H \rightarrow ZZ^*$	2.7%
$H \rightarrow \gamma\gamma$	0.2%
$H \rightarrow Z\gamma$	0.2%

Table 1.3: Branching ratio of the Higgs particle.

1.4 Feynman Diagrams

It is in particle physics only possible to describe an interaction to a certain extend by writing e.g.

$$a + b \rightarrow c + d$$

What is the propagator? What is the connection between initial and final state? This isn't too obvious. These issues have lead physicists to create a pictorial description of the underlying mechanisms. These diagrams are however not just a given picture the writer finds pretty, the diagrams has to follow the Feynman rules, which describes the entire interaction.

An example of one of these diagrams can be seen in fig (1.8), which shows the annihilation of a lepton pair which creates a new lepton pair, with a mediator of a Z-boson or a photon¹².

But how are these diagrams read? Unless otherwise stated the horizontal axis is time and vertical axis is space. The axes don't necessarily give specific time or distance differences, they are there to make it possible for the reader to distinguish different particles and whether they are there before or after the interaction.

Each of the lines symbolize either a particle or a propagator, as can be seen in fig 1.9.

A straight line with an arrow on it is a fermion, where the direction of the arrow tells the reader whether it is a the fermion (pointing right) or the anti-fermion (pointing left). A wavy line is an electroweak boson, the squiggly line is a gluon and the dashed line is the Higgs boson.

The diagram can then be described using the four momentum of the particles, the value for the propagator and the vertices.

The values of the particles and propagators can be seen in tab (1.4). There are apart from these also values for the vertices, which depends on what kind of interaction it is - e.g. Weak or QCD. For all of these vertices, it is required that the energy and momentum are conserved. It should be noted that one Feynman diagram is rarely enough to satisfy a calculation. There are most of the time several different possible time-ordered diagrams, and diagrams of higher order. The order is determined by the number of vertices, where due to the value of the coupling constants, the higher order terms will be suppressed. There are however still possibilities of higher order terms, e.g. the next to leading order term.

1.5 Cross section

The cross section of an interaction is a way of describing the probability of an interaction, but unlike normal probability the cross section has the unit of area. For example imagine if a beam with type "a" particles, with a given flux ϕ_a , which crosses a volume of space with n_b particle "b" per volume. The interaction rate of this example would

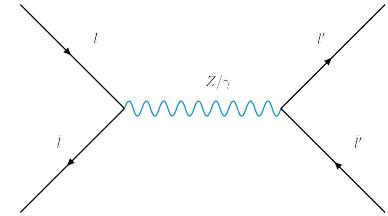


Figure 1.8: Feynman diagram of the annihilation of a lepton pair, which creates a virtual Z-boson or a virtual photon which decays to a new lepton pair.

¹² A virtual particle in a Feynman diagram is a particle which is not present as a real particle in the final state.

	Particle	Propagator
Fermion	$u(p)$	$\frac{i}{\gamma^\mu q_\mu - m}$
Photon	ϵ_μ	$\frac{-ig_{\mu\nu}}{q^2}$
W/Z	-	$\frac{-i(g_{\mu\nu} - q_\mu q_\nu/m^2)}{q^2 - m^2}$
Gluon	ϵ_μ	$\frac{-ig_{\mu\nu}}{q^2} \delta^{ab}$
Higgs	-	$\frac{i}{q^2 - m^2}$

Table 1.4: Values of particles and propagators in Feynman diagrams, $u(p)$ for a fermion is the spinor while ϵ_μ for the photons and the gluons are polarization.

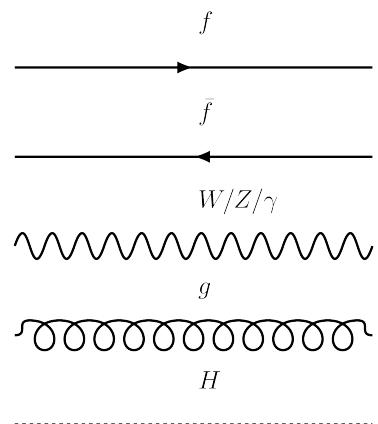


Figure 1.9: The different particles and propagators in a Feynman diagram.

be given as

$$r_b = \sigma \phi_a \quad (1.4)$$

with σ being the cross section, which has dimensions of area, with a flux given as particles per area per second, will r_b gain a dimension of particles per second, which coincides with our wish for a rate. It can be useful to think of σ as the effective cross section for the given target particle, imagine throwing golf ball towards a tennis ball, the tennis ball would get hit, if the distance between the center of the golf ball and the tennis ball is smaller than their combined radius. This is however rarely the case, and there will often be a significant difference between the effective cross section, and what is found in theory. The value which is found in theory, is an expression of the underlying quantum mechanical probability, for the interaction to occur.

To avoid this section going beyond the scope of this thesis and dragging on, will some assumptions regarding the cross section be made. First of all the idea of an effective cross section will not be used any more, and a focus on the quantum mechanical probability will be used instead. Secondly the Drell-Yan interaction will be used a two to two interaction, For example $a + b \rightarrow 1 + 2$. Third of all will a center of mass frame be used for simplicity, which simplifies the math.

The special case of the center of mass frame gives us

$$\mathbf{p}_a = -\mathbf{p}_b = \mathbf{p}_i^*$$

And

$$\mathbf{p}_1 = -\mathbf{p}_2 = \mathbf{p}_f^*$$

Which gives us an easier time describing the interaction.

To simplify the math in these interactions the Mandelstam variables are introduced:

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2 \\ t &= (p_1 - p_3)^2 = (p_2 - p_4)^2 \\ u &= (p_1 - p_4)^2 = (p_2 - p_3)^2 \end{aligned}$$

These three quantities are equivalent to the four-momentum squared of the exchanged boson, in their respective diagram¹³.

The Feynman diagram for these three variables also have their own processes, as given in a Feynman diagram, see fig 1.10[64].

Mandelstam variables are four-vector scalar products, and are thus Lorentz invariant, which simplifies calculations since they can be evaluated in any frame. If a center of mass frame is used, where there is no net momentum, and $p_1 = (E_1^*, \mathbf{p}^*)$ and $p_2 = (E_2^*, -\mathbf{p}^*)$ can s be written as

$$s = (p_1 + p_2)^2 = (E_1^* + E_2^*)^2 - (\mathbf{p}^* - \mathbf{p}^*)^2 = (E_1^* + E_2^*)^2 \quad (1.5)$$

So the Mandelstahm variable s is the total energy available in the system squared. Another feature of the Mandelstam variables gives a

¹³ It should be noted that p_1 and the other momenta are four momenta and is therefore given in Minkowski space as $p_1 = (E_1, \mathbf{p}_1)$, with p_1 being the three-momenta - what is normally thought of as momentum

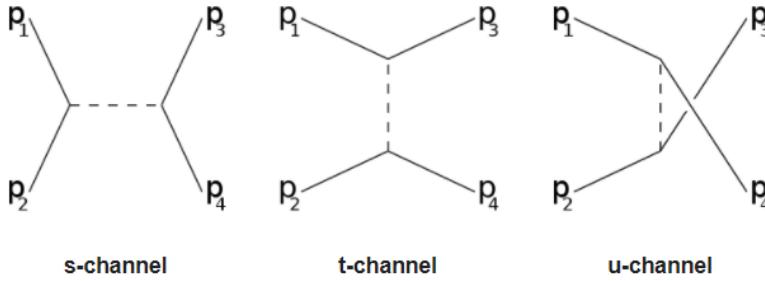


Figure 1.10: Feynman diagrams for s-, t- and u-channel processes. It should be noted that the u-channel is only applicable when there are identical particles in the final state.

specific value for a given interaction.

$$s + t + u = \sum_{i=1}^4 m_i^2 \quad (1.6)$$

Now using the Mandelstam variable and some math, which is omitted due to being outside the scope of this project, the cross section can be written as

$$\sigma = \frac{1}{64\pi^2 s} \frac{p_f^*}{p_i^*} \int |\mathcal{M}_{fi}|^2 d\Omega^* \quad (1.7)$$

Here, $d\Omega^*$ is the solid angle element in the center of mass frame. Lastly, there is $|\mathcal{M}_{fi}|$ which is the matrix element of the interaction, and can be calculated by using the Feynman diagram, the value is determined by which particles are in the interaction and the mediated boson¹⁴[54].

1.6 Luminosity

In the previous section was a concept of cross section introduced, but more information is required to calculate the production rate of a given particle. Given a production cross section σ_p can the production rate $\frac{dR}{dt}$ be acquired by using a proportionality factor.

$$\frac{dR}{dt} = \mathcal{L} \sigma_p \quad (1.8)$$

Where \mathcal{L} is the luminosity, which has unit of $\text{cm}^{-2} \text{s}^{-1}$. The description of the Luminosity varies depending on whether the target is fixed or if it is two colliding beams¹⁵. The fixed target luminosity is given by $\mathcal{L} = \Phi \rho_T l$, where Φ is the flux of the beam, ρ_T is the density of the target, and lastly l is the length of the target in the beam direction. If it is two colliding beam, the case will be slightly more advanced, since both beams will now be thought of as being both the beam and the target. The general idea is to include both the number of particles or bunches in each beam N_i and their density $\rho_i(x, y, s, \pm s_0)$, with $s_0 = c \cdot t$ is the distance to the central collision.

Especially in the case of two beams can the luminosity vary with time. Integrating the luminosity with time allows for relation between the luminosity and the observed events to be established. The integrated

¹⁴ These calculations are beyond the scope of this thesis, and manual calculations are mainly done while learning the material. After having learned about it, it is a lot more common to calculate the values using various programs.

¹⁵ Normally two colliding beams would be used, since this can obtain a higher center of mass energy.

luminosity is defined as.

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') dt' \quad (1.9)$$

This can be used to obtain the events of interest

$$\mathcal{L}_{\text{int}} \cdot \sigma_p = \text{events of interest} \quad (1.10)$$

The time which this integrated luminosity is integrated over is the sensitive time - a.k.a. the time where the measuring is taking place[12].

1.6.1 Expected values of the cross section

The expected production rate of the different decay processes for the Higgs particle can be calculated using the total production of Higgs particles times the branching ratio. Using the values found in table

Name of process	Cross section [pb]	#H bosons
Gluon gluon fusion	49	6.9M
Vector Boson Fusion	3.8	520k
Vector Boson associated production	2.3	320k
Top quark associated production	0.5	70 k

Table 1.5: Production rate of the different Higgs production channels during Run-2 at the LHC.

1.5 from [54][28][27], can the the total number of Higgs particles be found produced. If only the GGF and VBF production channels are used, then the total number of Higgs produced during Run2 will be 7.4M. Using a few more decimals than in tab 1.3 a branching ratio of $B(H \rightarrow \gamma\gamma) = 0.23\%$ and $B(H \rightarrow Z\gamma) = 0.154\%$ can be obtained. This leads to approximate 17k $H \rightarrow \gamma\gamma$ and approximately 11.4k $H \rightarrow Z\gamma$. Using the branching ratios from fig 1.4 a production of both $H \rightarrow Z(ee/\mu\mu)\gamma$ can be calculated to ≈ 800 , during Run2.

1.7 Pileup

In the ideal world would only one interaction produce an interesting particle, with no other interactions creating particles, when running the particle accelerators. However, this is not the case, due to the low production cross sections for e.g. the Higgs boson are bunches of protons used instead of single protons. This does however give us the small problem of a lot of additional interactions, which rarely contains interactions of interest, the cross section for inelastic scattering between two protons at 13 TeV is $78.1 \pm 3.0\text{mb}$ [31], while the production cross section for the Higgs boson is 52.8pb (only taking gg and VBF into account, from tab 1.5), this is roughly a factor of $1.5 \cdot 10^9$ in difference. All these particles, which are produced in the additional interactions are pileup.

One may ask, how many pileup interactions there are during a run at the LHC? During run 2 at the LHC was the average 32 pileup interactions, and they had more than 50 pileup interactions in short periods of time, this increased to 90 pileup interactions in the test

period.

An example of pileup, can be seen in fig 1.11 from CMS, where multi-

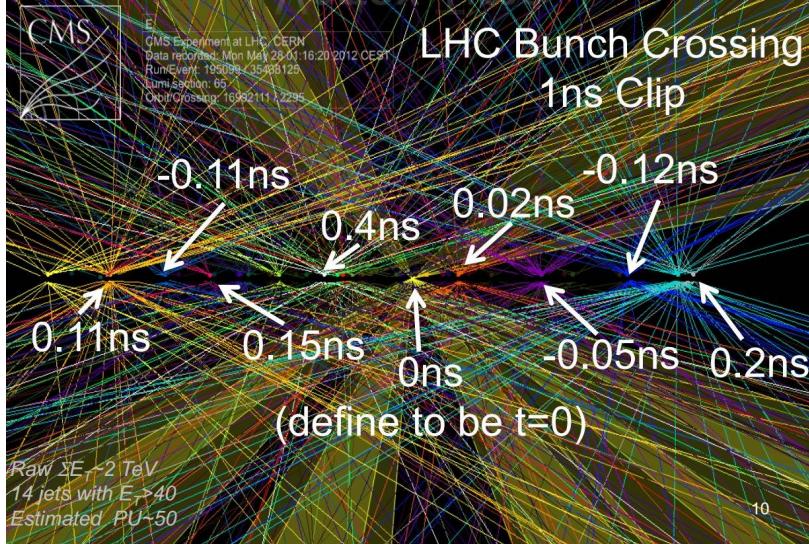


Figure 1.11: Example of pileup in an interaction, where the interaction of interest is at $t=0$. This image is from the CMS experiment, but this would also be applicable for the ATLAS experiment[14].

ple pileup interactions are shown to create a lot of particles. The issue with these pileup events is that they in some circumstances looks quite similar to what would be expected from the signal. E.g. jets of pileup particles, pileup jets, mimics the Higgs boson and therefore represents a background which needs to be rejected, if possible[29][47].

1.8 Exploration of $H \rightarrow Z\gamma$ and $H \rightarrow \gamma^*\gamma$ and improvement of $H \rightarrow \gamma\gamma$

One may ask why the photon identification specifically should be improved. The reason for this is to improve our chance of finding $H \rightarrow Z\gamma$ and $H \rightarrow \gamma^*\gamma$ ¹⁶ which so far haven't been discovered. Improvement of the photon identification will also improve our $H \rightarrow \gamma\gamma$ reconstruction. Improvement of these channels, since they are through the Standard Model theorised to be common decay products of the $H \rightarrow HH$, which is yet to be discovered. Improvement in detection of its decay products which will improve our ability to measure the interaction.

¹⁶ 3.2σ has been observed for $H \rightarrow \gamma^*\gamma$, and thus would be to improve the measurement.

2

The ATLAS detector

It is now time for an introduction to the ATLAS experiment, now that particle physics has been introduced. The detector in question is one of the four detectors at the LHC (Large Hadron Collider), which is a particle accelerator at CERN (The European Organization for Nuclear Research), on the swiss side of the France-Switzerland border near Geneva.

The collider has a circumference of 27 km and is buried underground. The layout of the collider is shown in figure 2.1, where the four main experiments are written in yellow: A Toroidal LHC Apparatus (ATLAS), Compact Muon Solenoid (CMS), LHC-beauty (LHCb) and A Large Ion Collider Experiment (ALICE). Where the goal of these experiments is to test particle physics theories, such as measuring the properties of the Higgs boson, families of new particles predicted by supersymmetric theories and other unanswered questions in particle physics[62]. For the sake of this thesis will the details on how CMS, LHCb and ALICE are build, and what they are researching, not be discussed since they are not the focus of this thesis.

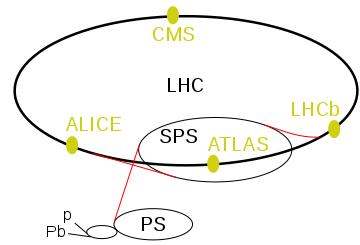


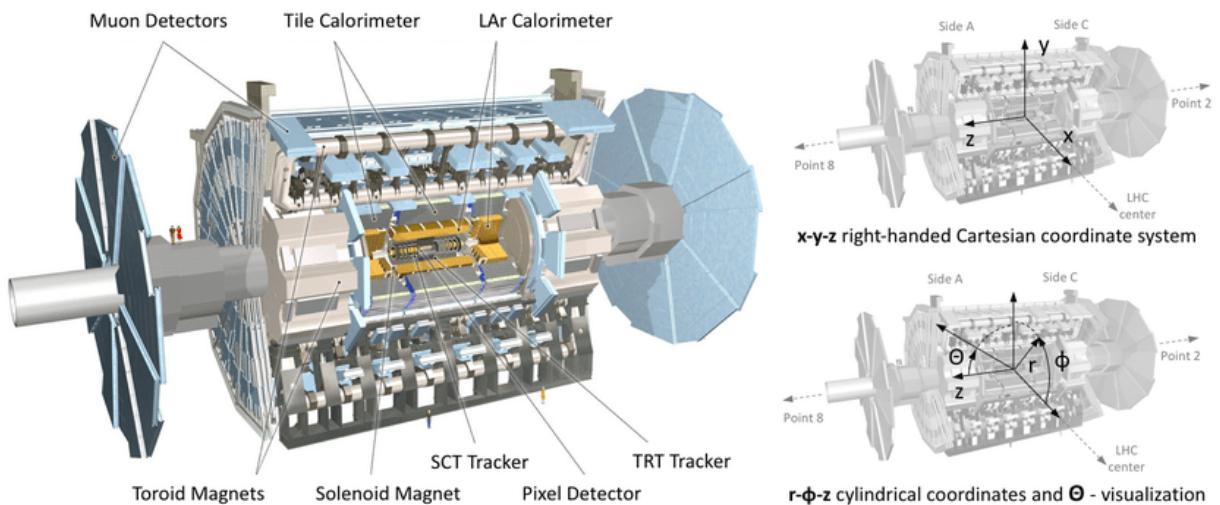
Figure 2.1: The Large Hadron Collider and its experiments.

THE ATLAS DETECTOR is the largest detector constructed for particle collisions - regarding its volume. Where LHC is a ring in the ground, where the particles can accelerate, the ATLAS detector is a cylinder, where they can collide. The ATLAS detector is 46m long and has a diameter of 25 meters, with a staggering weight of 7000 tonnes, this weight is similar to that of the Eiffel Tower.

It is however not only its volume which is large, the data size is even larger. The beams of particles are accelerated to energies of 6.5 trillion electron-volts, with an equivalent speed of 99.99999% of the speed of light. These beams collide at the center of the ATLAS detector, which results in new particles being created and their decay products (in the case they are unstable - as most particles are) flying in all directions. This number of interactions are at a staggering amount of a billion interactions per second, this data rate is similar to 20 simultaneous telephone conversations held by every person in the world. ATLAS handles this amount of data by only making use of a small fraction of it, since most of it is of little to no interest, and only about one

in a million collisions are flagged as interesting and are therefore recorded.

But how are all these measurements made? The detector consists of many layers of instruments, with six different detecting subsystems, to measure the kinematics and the type of particles. Furthermore there is a magnet system which bends the charged particles path, which allows us to measure their momenta[5]. These subsystems are what will be introduced in the following sections after a brief introduction to the coordinate system that ATLAS uses and some other important variables and phenomena.



2.1 ATLAS coordinate system

The coordinate system of the ATLAS detector can be described both using a cartesian coordinate system and a cylindrical coordinate system. Where the cartesian coordinate system is defined in such a manner that the interaction point is the origin of the coordinate system. With the z-axis pointing along the beam axis, with the positive being towards side A and the negative direction being towards side C, see figure 2.2 for clarification of side A and side C. The x-y plane is chosen to be transverse to the beam direction, with the positive x-axis pointing towards the center of the LHC ring, and positive y-axis pointing up.

The cylindrical coordinate system has a similar z-axis as the cartesian coordinate system. Where the azimuthal angle and the polar angle, ϕ and θ , are measured around the beam axis and the angle from the beam axis respectively[21]. It is usually the cylindrical coordinates which are used.

The reason for using the cylindrical coordinates rather than cartesian coordinates, is due to the cylindrical coordinates being directly translated into other important variables such as pseudorapidity, rapidity

Figure 2.2: Left: The ATLAS detector including its' different subdetector parts. Right: The two coordinate systems of the ATLAS detector[4].

and the angular separation.

RAPIDITY: Is used to measure relativistic velocities. One of the strengths of rapidity is that differences in rapidity are Lorentz Invariant under boosts along the propagation axis. This is crucial, since the collision of two protons are essential collisions of many partons, which carry different momentum fractions of the protons in the Z-direction. This would result in them having different boosts in the rest frame, which doesn't give any problems when using rapidity. Rapidity is defined by

$$y \equiv \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) \quad (2.1)$$

PSEUDORAPIDITY: Is used as a spatial coordinate, which describes the angle of the particle relative to the beam axis. It is defined as

$$\eta \equiv -\log \left(\tan \left(\frac{\theta}{2} \right) \right) \quad (2.2)$$

Where it can be noted that

$$y \approx \eta - \frac{\cos(\theta)}{2} \left(\frac{m}{p_\perp} \right)^2$$

In the limit where $p_\perp \gg m$ pseudorapidity becomes equal to rapidity. Since the mass is unknown, then pseudorapidity is used instead of rapidity since $\Delta\eta$ is invariant[65].

ANGULAR SEPARATION: Is Lorentz invariant under a boost in the z direction. The angular separation can be described purely through angular quantities by

$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.3)$$

2.2 Particle interaction in matter

It is important to know how particles interact in matter, since it is through these interactions they can be detected[36]. A certain distinction will have to be made, when working with the particles interactions in matter. The distinction is between charged and neutral particles, e.g. a proton vs a neutron. This distinction will have to be made since the fundamental forces couples to different particles - the electromagnetic force couples to charged particles, but not neutral particles. Lastly there is also a difference between how hadrons and leptons interact in matter.

2.2.1 Interactions of charged particles in matter

The charged particles can loose energy through two types of processes: Energy loss due to ionization and energy loss due to radiation.

For most cases, all except electrons, ionization energy loss will be dominating the energy loss compared to radiation energy loss, at all energies but the highest energies.

THE IONIZATION ENERGY LOSS arises from Coulomb scattering from the atomic electrons. The theory behind this was worked out first by Bohr, and then in more detail by Bethe, Bloch and others in the 1930's. The Bethe-Bloch formula, which describes the energy loss of the particle per length, is given by:

$$\frac{dE}{dx} = \frac{4\pi N_A r_e^2 m_e c^2 z^2 \rho Z}{A \beta^2} \left(\frac{1}{2} \log \left(\frac{2m_e c^2 \gamma^2 \beta^2 W_{max}}{I} \right) - \beta^2 - \frac{\delta}{2} + \frac{C}{Z} \right) \quad (2.4)$$

Where the constant $4\pi N_A r_e^2 m_e c^2 = 0.3071 \text{ MeV g}^{-1} \text{ cm}^2$. r_e is the classical electron radius $r_e = e^2 / (4\pi\epsilon_0 m_e c^2) = 2.817 \times 10^{-13} \text{ cm}$. Z and A are the atomic number and weight of the absorber, whereas z refers to the charge of the projectile.

N_A is Avogadro's number, $6.022 \cdot 10^{23}$. ρ is the mass density of the absorber. W_{max} is the maximum momentum transfer, defined as

$$W_{max} = \frac{2\gamma^2 \beta^2 m_e c^2}{1 + 2\gamma^2 \left(\frac{m_e}{M} \right) + \left(\frac{m_e}{M} \right)^2} \quad (2.5)$$

I is the average ionization potential, and can be approximated as $I = 16Z^{0.9} \text{ eV}$. δ is the so called density effect, which is a screening parameter, this is caused in the case of extremely relativistic projectiles polarizing nearby atoms. Lastly there is C which is a quantum mechanical shell correction, which is ignored except for very slow projectiles.

This energy loss only depends on the mass of the projectile and very weakly through W_{max} . It is furthermore common to normalizing the target material, by dividing through with ρ , this allows us to look at the energy loss in an equivalent thickness (given in g cm^{-2}).

RADIATION ENERGY LOSS is due to radiative collisions between a charged projectile particle and the nuclei of the target material. This leads to the radiation of photons, so called bremsstrahlung¹. This energy loss is however only dominant for electrons - except for energies at TeV, which is due to the fractional energy loss for an electron being $\frac{dE}{E} \frac{Zze^2}{m_e^2}$, where it would result in a factor of $\frac{m_e^2}{m_i^2}$ if it should be calculated for other particles than the electron, with the factor of m_i being the mass of the projectile[36][24].

¹ German for breaking radiation

2.2.2 Interactions of neutral particles in matter

Neutral particles in matter can be divided into three types: The photon, particles like the neutron and lastly neutrinos.

PHOTONS are the most important of the three for this thesis, and will therefore be covered first. Photons can interact with atoms in multiple ways, which one depends on the energy of the photon. At energies below eV the main interaction is Thomson/Raleigh scattering, if the energies are in the range from eV to keV the main interaction will be photoelectric effect, and if the energy is in the range keV to MeV the main interaction will be Compton scattering. These forms of interactions are however of little to no interest to us, since it is the photons this thesis will focus on, will have energies of 10 GeV and above, which leads us to the last form of interaction: Pair production (MeV and above). At this energy, the photon can fluctuate into a virtual electron pair, if a virtual photon is exchanged between either of these electrons and the nucleus, both electrons will be brought on mass shell - they are no longer virtual. This effect is proportional to the material the photons are traversing. This pair production leads however to multiple electrons, which are able to radiate off photons due to Bremsstrahlung. This results in both photons and electrons creating electromagnetic showers, when they traverse a medium. The only noticeable difference between the shower from an electron or a photon, is that the shower from a photon starts slightly later.

NEUTRONS, due to being neutral, do not ionize the material which they traverse. At energies of sub-MeV they will behave differently compared to protons - in regards to nuclear interactions. They have at these energies a high cross-section for nuclear interactions with especially light nuclei, which allows for absorption of the neutron. At multi-MeV energies protons and neutrons will have rather similar cross-sections, which results in their difference being whether or not they can be seen in the tracking detector.

NEUTRINOS only interact with other particles through the weak interaction, which results in them being extremely hard to detect. They will normally pass through all the sub detectors without leaving any signal, and are thus normally taken into account through missing transverse momentum²[36][24].

2.2.3 Material budget

The amount of interactions a particle has with a material differs, from material to material depending on the properties of the material. This is normally described using a variable called the radiation length. This radiation length is the average thickness of a material required to reduce the mean energy of an electron by a factor of e. It can mathematically be calculated using:

$$\frac{1}{X_0} = 4 \left(\frac{\hbar}{mc} \right)^2 Z(Z+1) \alpha^3 \frac{N_A}{A} \log \left(\frac{183}{Z^{1/3}} \right)$$

with Z being the atomic number, A is the atomic mass³, N_A is

² Since momentum has to be conserved in the transverse plane (should be zero). If the momentum is non-zero is the missing transverse momentum normally assumed to origin from neutrinos.

³ Both of these are for the material.

avogadro's number and α being the finestructure constant. It is wise to describe distances in radiation lengths instead of normal lengths, since the radiation length varies from material to material. This can give us an idea of how much energy is lost for a given particle passing through the medium. An example of this can be seen in fig 2.3[8]. As it can be seen here the material budget is given as a function of pseudo rapidity - it has rotational symmetry so no ϕ dependency⁴. Furthermore a distinction between the subdetectors has been included, so it is easier to determine where the particle loses its energy.

⁴ This is assuming that the detector is rotational symmetric.

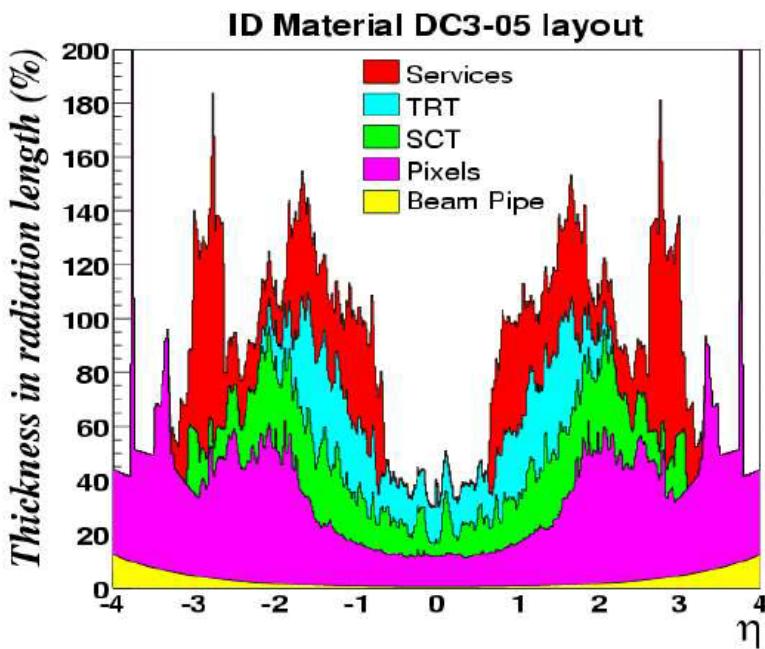


Figure 2.3: Material budget for the ATLAS ID, with distinctions between which subdetector the electron passes through. The forward direction ($|\eta| > 0.8$) has more material, and as a result, more photons convert here.

2.3 Inner Detector

As the name suggests, the inner detector is the first part of the detector to make measurements on the particles. Where it measures the directions, momentum and charge of the electrically-charged particles produced in the collisions - neutral particles can not be observed by this detector. This is done by detecting the interactions between the charged particles and the material of the detector at discrete points in the detector, where the total size of this sub detector is 6.2 m and it has a diameter of 2.1 m. To do these measurements the entirety of the inner detector is enveloped in a 2 T magnetic field, which causes the charged particles to curve - which reveals the particles charge depending on the direction of the curvature. Furthermore, the degree of curvature will give information about its momentum.

But the inner detector can be divided into even smaller pieces: The Pixel Detector, The Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

Tracking of particles in the Inner Detector are only performed if they have a transverse momentum higher than 0.5GeV. This tracking requires the particles to have $|\eta| < 2.5$, if electron identification is to be performed on the particle will $|\eta| < 2.0$ [5][21][61].

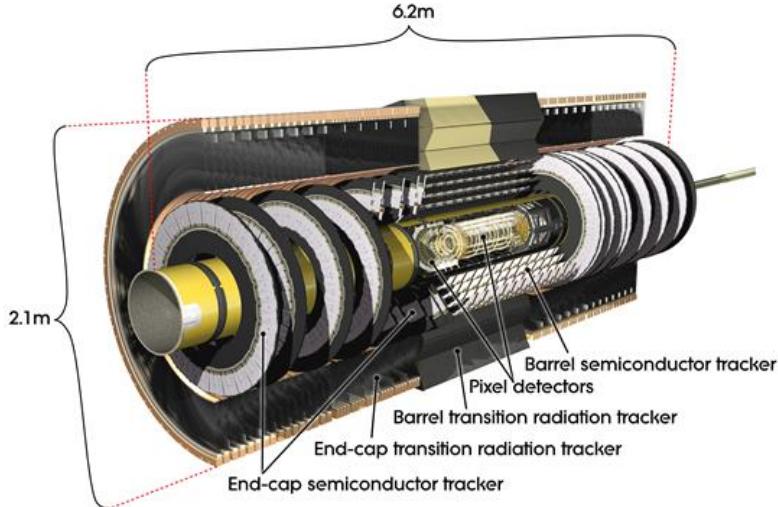


Figure 2.4: The Inner detector with its subdetectors: The Pixel detector, Transition radiation tracker and semiconductor tracker.

2.3.1 Pixel Detector

The Pixel Detector is the innermost part of the Inner Detector, and is therefore designed to be able to receive significant amount of radiation. The Pixel detector contains 92 million pixels, over a Silicon area of 1.9 m^2 , each of these pixels have a size of $50 \times 400 \mu\text{m}^2$ for the external layers, and $50 \times 250 \mu\text{m}^2$ for the innermost layer. It can be divided up into two areas, the barrel, which has 4 layers and 1736 sensor modules and 3 discs at each end-cap which has 288 modules.

The general idea of the pixel detector is to make an inverse camera. Where a normal camera would focus light onto a single point, the pixel detector absorbs the light/particles from a single point. This leads to a cylindrical construction, with each pixel being 50 by $400 \mu\text{m}$. After the interaction have happened all the particles fly out in all directions, and therefore hit the pixel detector. These particles have energies in the range of 100 MeV to 100 GeV and can hence knock out thousands of electrons in the Silicon strips. These electrons and their corresponding holes drift in different directions due to an applied electric field. This will however not stop the incident particle, but it will track its course through the detector. From here the charge can be read out and the data can be saved if deemed important[46][61].

2.3.2 Semiconductor Tracker (SCT)

A semiconductor tracker uses, that electrons in the valence band of a solid can be excited to the conduction band, due to an incident particle. This leads to a "hole" in the valence band, which can be understood

as a lack of an electron, in a place where there is supposed to be an electron - it acts as a positron. If an electric field is applied to the semiconductor can the electron and the hole be separated - since the hole behaves as a positron it will accelerate in the opposite direction compared to an electron. The signal from this current, will depend on the energy loss of the incident particle - more energy deposited in the SCT results in a larger signal. If this is used in conjunction with a material which have small band gaps⁵ (as small as 1 eV), pairs of electron-holes can be produced by depositing on average 3-4 eV. This will result in a large number of electron-hole pairs with only small statistical fluctuations, from a low-energy particle - thus a SCT is good at detecting low energy particles[36].

The SCT in the ATLAS detector is the middle component of the inner detector. It is somewhat similar to the Pixel Detector in concept, but where the Pixel Detector had small pixels, the SCT has long, narrow strips of silicon (The width of these strips are 80 μm , which allows for an accuracy of 17 μm when recording particles (in the transverse direction of the strips))[5]. These strips have a slight angle between them, which allows the SCT to measure in 3D. There have been used 60 m^2 of silicon for the microstrips, which is distributed over 4 cylindrical barrel layers and 18 planar endcap discs. This combined with each microstrip tracker consisting of 4,088 two-sided modules and over 6 million implanted readout strips, allows the SCT to record charged particles with an accuracy of 17 μm . The SCT is the most critical part of the inner detectors ability to track particles in the plane perpendicular to the beam direction. This is due to the SCT measuring particles over a larger area than what the Pixel Detector does[61].

2.3.3 Transition Radiation Tracker (TRT)

Lastly there is the Transition Radiation Tracker, which is the outermost component of the inner detector. The TRT detector has 350,000 read-out channels distributed over a volume of 12 m^3 . The basic detector element is a straw tube, with a diameter of 4 mm, and a length of 144 cm in the barrel and a length of 39 cm in the endcaps, with a core of 0.03 mm gold-plated tungsten wire. The TRT consists of 50,000 of these straws in the barrel - where each end of the straws are read out separately. There are an additional 125,000 straws in each of the endcaps - for a total of 300,00 straws in the TRT. The TRT has a slightly worse position resolution compared to the Pixel Detector and the SCT - it obtains a precision of 0.17 mm, but it is indispensable due to its ability to enhance the momentum resolution, providing a fast level-2 trigger and track measurement points up to the radius of the solenoid magnet - this is described in further details later. It is furthermore able to provide a stand-alone electron/pion separation[50].

To do this separation, ATLAS is using a phenomenon called Transition radiation. This is a phenomena which occurs when a relativistic

⁵ The energy difference between the valence and conduction bands.

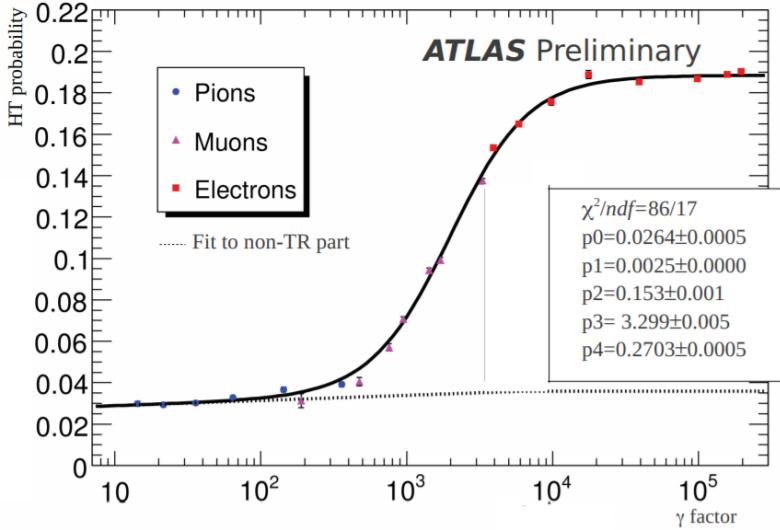


Figure 2.5: The high threshold probability for different particles in test beam data - as a function of Lorentz gamma factor[7].

charged particle crosses the interface of two media with different dielectric constants. The loss of energy due to this transition depends on the lorentz factor of the factor, with the emitted photon directed forward with peak intensity at an angle of $1/\gamma$ compared the the path of the incident particle. The Lorentz factor dependency allows us to discriminate between particles - especially electrons from other particles. Given a reasonable example of particles with momentum in the range from 1 to 100 GeV, this results in TR photons produced by the electron, with energy in the range of 5 to 15 keV, the x-ray range. But since a particle with $\gamma = 2 \cdot 10^3$ produces on average 0.8 x-ray photons per transition[24][63]. To increase the number of x-ray photons is the TRT detector created with a lot of layers of alternating inert and detector material, which results in an increase in the number of TR photons - due to a lot more transitions from one medium to another[61][63].

But how does this look for ATLAS? ATLAS uses the generic onset curve, which is given as

$$p_{HT}(\gamma) = p_0 + p_1 \log_{10}(\gamma) + \frac{p_2}{1 + \exp(-(\log_{10}(\gamma) - p_3)/p_4)} \quad (2.6)$$

Atlas combines 30-35 hits in this likelihood to assert whehter the particle is an electron or not. Example of this is displayed in figure. 2.5. As it can be seen, both pions and muons have a probability value lower than the electrons - which is due to the Lorentz gamma factor of the different particles. Given a certain amount of energy the electron will always have a higher γ factor, than the muon (pion) since they weight 200 (280) times as much as the electron.

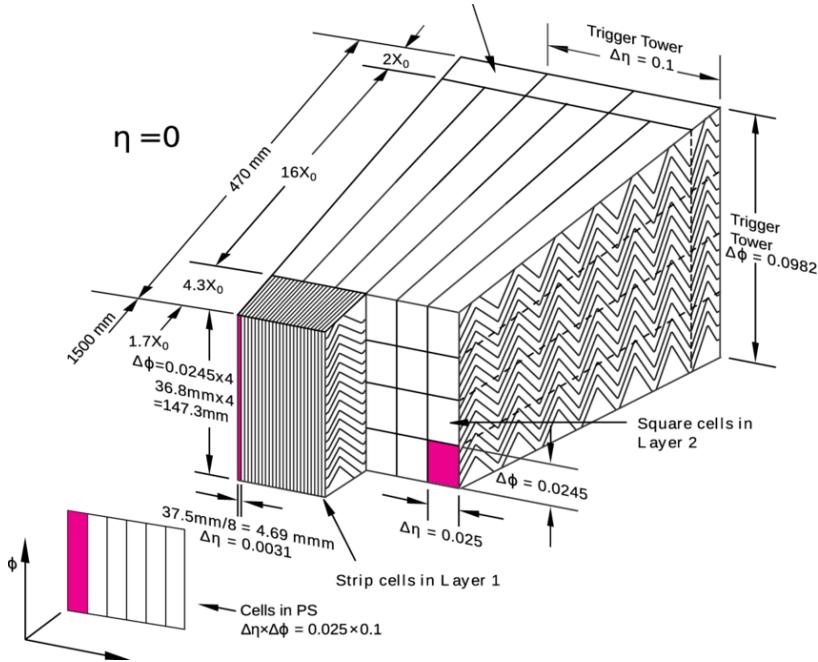


Figure 2.6: Sketch of the Electromagnetic Calorimeter.

2.4 Calorimeter

A calorimeter is a device used to measure the total energy and position of hadrons, electrons and photons, by total absorption of the incident particle - or more precisely it measures the energy loss of a particle being stopped in the detector - all but muons are stopped. This is however not as simple as it might seem, the incident particle creates a shower, from which it has to be reconstructed from. Furthermore does particles not interact in the same way, there are two main ways. The electromagnetic cascades, which include electrons, photons and also π^0 which all interact electromagnetically - the decay of the π^0 is electromagnetically $\pi^0 \rightarrow \gamma\gamma$, these are mainly measured in the Electromagnetic Calorimeter, sketched in fig 2.6, taken from [39]. The other kind of cascades are hadronic, these are called jets and include all the hadrons except π^0 . however, calorimeters are not able to stop muons and neutrinos, and are therefore unable to measure those particles.[24][9][5].

2.4.1 Electromagnetic Calorimeters

A relativistic electron travelling through a medium, loses the majority of its energy to bremsstrahlung⁶, which is proportional to its current energy

$$\frac{dE}{dx} = -\frac{1}{X_0} E \quad (2.7)$$

This gives a depth dependence of the energy of $E(x) = \exp(-x/X_0)$. The bremsstrahlung photon will after a mean flight distance of $\frac{9}{7}X_0$ be converted into an electron pair, e^+e^- . This leads to what is called an electromagnetic shower, since these electron will radiate off more

⁶ See section 2.2.1

bremsstrahlung photons, which will be converted to electron pairs until the critical energy is reached $E_c = \frac{1194 m_e c^2}{Z+1.24}$. Such a shower can be seen in fig. 2.7[18].

THE EM CALORIMETER USED BY ATLAS is the Liquid Argon(LAr) calorimeter, the barrel has a length of 6.4m and a thickness of 53 cm, which contains 110,000 channels. The LAr has been made in such a way, so it doesn't have any so called dead spaces. Dead spaces are areas of the detector where it can't measure due to the transportation of electronic or optical signals. These dead spaces have been avoided by crafting the absorbing material in an accordian shape, as seen in fig 2.8. The calorimeter consists in addition to this of 4 layers, with different metals, such as tungsten, copper and lead[5].

In the LAr the electronic signal is transported by copper electrodes which are sitting on the lead absorber plates, this electronic signal arises from the ionization of the liquid argon, which fills out the space between the lead plates[5][24].

2.4.2 Hadronic Calorimeters

The hadronic calorimeter won't be covered to the same extend as the Electromagnetic calorimeter, due to the difference in significance between these two sub detector parts for this thesis. But the idea of them are that the hadronic particles are much harder to stop than EM particles. HCAL are therefore required to have a lot more material to stop the particles. Since hadrons may sometimes deposit only the minimum ionization energy in the EM calorimeter, there will be the additional need for an outer layer of a hadronic calorimeter - they may at other times do more interactions in the EM calorimeter.

The hadronic calorimeter consists of a central barrel with 64 wedges, where each of these are 5.6 m long and weigh 20,000 kg. There is in addition to this two extended barrels, with each 64 wedges - which are 2.6m long and weigh 9,600 kg. All of this ends up being approximately half a million plastic scintillator tiles, and a total weight of 2900 tons[5].

The hadronic calorimeter is build as a sampling type detector, to keep the cost down. A sampling calorimeter combines layers of high-Z (i.e. dense) passive absorber layers, which generates cascades, with layers of active detectors, which counts cascade. The active detector part can be made of things like plastic scintillators, gaseous wire chambers or noble liquid ionization chambers. This is relatively inexpensive, and also relatively easy to manufacture with a fine granularity.

But when working with the hadronic calorimeter you run into the problem of having to translate the particle energy into an electronic signal - since it is done differently for π^0 and the other hadrons. Especially the hadronic components have many ways to dissipate the energy in nuclear reactions - some of which are not detectable. This results in hadronic energy calibration being harder to do. The hadronic calorimeter used by ATLAS is called the Tile Hadronic Calorimeter

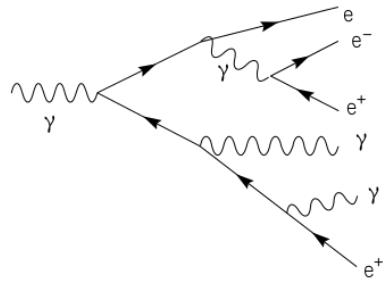


Figure 2.7: An Electromagnetic shower created due to pair production and bremsstrahlung.

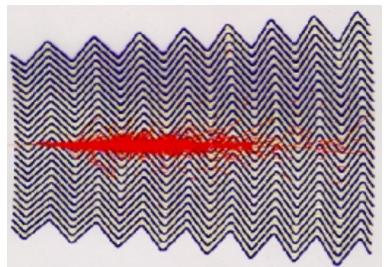


Figure 2.8: The ATLAS LAr calorimeter with simulated electron shower in red.

(TileCal), and works by sampling the energy of the hadrons[24].

2.5 Muon Spectrometer

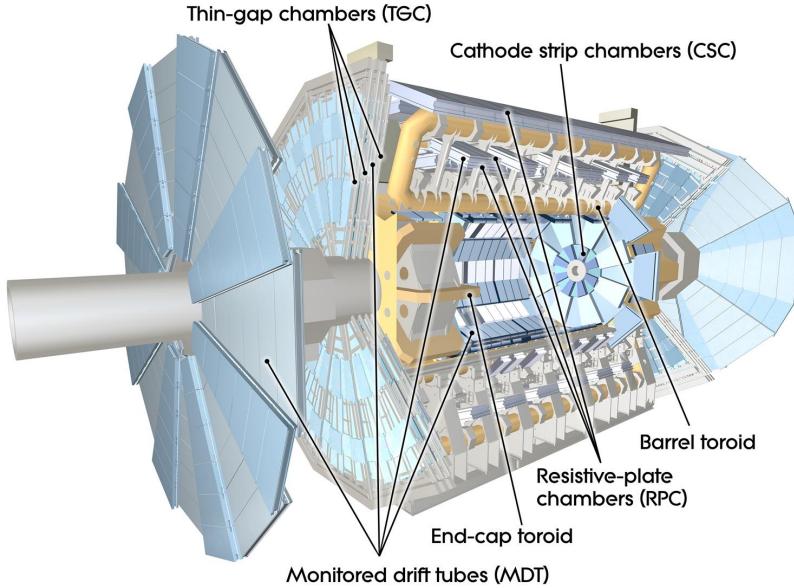


Figure 2.9: Graphical representation of the muon spectrometer, and detector parts.

The outer most part of the ATLAS detector has now been reached. The Muon Spectrometer consists of four subsystems, as seen in fig 2.9[5]: Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC). The Thin Gap Chambers, with their 440,000 channels, are used for triggering and the second coordinate measurement, in the non-bending direction, at the ends of the detector. The Resistive Plate Chambers, with their 380,000 channels and an electric field of 5,000 V/mm, is also used for triggering and second coordinate measurements, however this is for the central region of the detector. The Monitored Drift Tubes, with their 1,171 chambers and 354,240 tubes, are used to measure the curves of the track, where it obtains a resolution of 80 μm . Lastly the Cathode Strip Chambers, with their 70,000 channels, are used to measure precision coordinates at the ends of the detector, where it obtains a resolution of 60 μm [5].

The principle of the Muon Spectrometer is somewhat similar to the inner detector - have the charged particle move through a magnetic field, so the momentum can be measured. But due to how little the muons interact with other materials (the muons loses 1-3 MeV in the calorimeter), the sub-detector extends from a radius of 4.25 m to 11 m (the full radius of the detector). This allows the Muon spectrometer to measure the momentum of muons to a good accuracy. A muon with an energy of 100 GeV would have its momentum measured to an accuracy of 3%, while a muon with 1 TeV would have its momentum measured to an accuracy of 10%.

This part of the detector system is important - especially for the case

of interactions which produces muons as a decay product. If the Muon Spectrometer had not been implemented, it would have been near impossible to measure the total energy of the particles in the event - since the muons would not have been measured[61]. This would be due to the curvature of the muons path in the inner detector would be too small to measure and high p_T values.

2.6 Magnet system

The magnet system at ATLAS consists of two different superconducting magnet systems - a solenoidal and a toroidal. These magnets are needed to determine the momentum and the charge of the particles in the field. The more a particles path is curved due to the effect of the magnetic field, the lower the momentum of the particle. The direction of this curvature is determined by the charge of the particle.

The solenoid magnet surrounds the inner detector and provides a 2 T magnetic field, this field is strong enough to force particles with a momenta lower than 400 MeV to repeatedly loop and therefore not be measured. The magnetic field from the solenoid magnet has an almost uniform direction, this combined with the strength of the field, allows for very precise measurements.

The toroidal magnet consists of eight air-core superconducting barrel loops, and two end-caps toroidal magents. Where the solenoid surrounds the inner detector, the toroidal is outside the calorimeters and within the muon system. It is furthermore able to create an even stronger magnetic field, than what the solenoid is able to produce, the strength can vary between 2 and 8 T. Unfortunately the magnetic field is not uniform, since a solenoid magnet with the size required, would be too expensive to build[5][61].

2.7 Trigger and Data Acquisition System

All these subdetector systems need a way to select which particles to save and which to discard, the system used for this is the Trigger and Data acquisition system. Data acquisition is a daunting task, since ATLAS can observe up to 1.7 billion (1.7GHz) proton-proton collisions per second, this data volume is equivalent to more than 60 terabytes per second. This amount of data is not sustainable to save, a lot of this is also unwanted data. To solve this problem ATLAS uses a two stage triggering system, the level-1 hardware trigger and the High-Level Trigger (HLT).

LEVEL-1 HARDWARE TRIGGER: The first stage uses information from calorimeter and the muon spectrometer, to decide whether or not the data is kept for further analysis. The choice of whether the data should be saved or not is done within 2.5 μ s of the event, this is done through a buffer of a 100 events in each detector part. This Hardware trigger saves up to 100,000 (100kHz) events per second, which the

High-Level Trigger can work on further.

HIGH-LEVEL TRIGGER: The HLT bases its trigger on the results from a large CPU farm, which runs an analysis of either the whole event, for the selected layers of the detector, or by using the data in smaller isolated regions of the detector. This results in only 1000 events per second being selected by the HLT, and are thus fully reassembled into event record, and saved for further analysis[5][6][49].

3

Machine Learning

The ATLAS experiment is currently selecting signal and background by applying a cut based method. Here the signal is separated from the background by setting threshold in variables such as the invariant mass, the energy, the charge, or the pseudorapidity. This is conceptually easy to understand and apply, but it comes at a cost. The signal will also be reduced if the background is reduced using these cuts. This can be a crucial problem when working on interactions with very low interaction rates - such as interactions with the Higgs particle. This thesis will make use of Machine Learning (ML) methods to increase the signal efficiency compared to the cut based method applied by ATLAS. This chapter introduces some different machine learning theory and methods, used in this thesis.

3.1 Learning Problem

An event contains information about several particles, each with multiple variables. If a person was to decide whether one of the particles indeed were e.g. a photon, that person would have to take several of these variables into account. The person would also have to find all the other relevant particles, and discard the ones irrelevant for the interaction of interest. This would be a daunting task, and even more so when the person realised that the influx of events is 1kHz.

To avoid this issue machine learning will be used to categorize the particles, and do reconstruction of the interactions. In the following sections how the data can be divided into subsets to improve predictions, which biases might arise in the models, what algorithms might be useful and the risk of over- or underfitting will be discussed. This should give the reader a better idea of what machine learning is, why it is so useful, and thus why it is used for this thesis. Covering all details regarding machine learning would be too long for this thesis, so some aspects of Machine learning has been left out[3].

3.1.1 Training, validation and test set

When making a model it will be required to separate data, so the model can be evaluated on a dataset it hasn't trained on before. The dataset is normally split into three subsets:

Training dataset: The subset of data used to train the model.

Validation dataset: Used for fine-tuning the hyperparameters and evaluation of the model. The model will occasionally use the subset, but it won't learn anything from it.

Test dataset: Used when the model is done training, to evaluate how well the model performs, can also be used to compare models.

How the separation of data should be performed varies from model to model. Model using only a few hyperparameters will require a smaller validation set, than a model with a lot of hyperparameters. The amount of data required to train a model also varies, and it can thus be required to have a larger training sample in these cases. If there is little to no validation set, it can be created by using cross validation. This is when a random subset of the training data is selected as the validation set, this can be repeated several times, and is a good method to reduce the risk of over fitting [2].

For this thesis the test size has been selected to be at 20%, thus the training set is 80%, where this training set has been split into a training and validation set, with the validation set being 20% of the training set, and the real training set being the remaining 80%.

3.1.2 Bias in the data set

It is important to be aware of the content in the training sample, since it may affect the model in ways it wasn't supposed to. For example, given a training set which is 80% signal, would result in a model more inclined to predict candidates as signal, so it may predict that 95% of the test sample is signal - which would be wrong.

This is called a bias, where the model chooses to focus on details it is not supposed to focus on. I'll now introduce some few examples of bias in the data sets.

Sample bias: A model performs the best when the training set it trained on resembles the test set it is predicting on. For our case of reconstruction of particle, will better results in data be obtained, if the Monte Carlo data has similar distributions to the real distribution found in the ATLAS detector.

Prejudice bias: This bias may arise, when the distribution of signal and background differs a lot. For this thesis the variables η (pseudorapidity), E_T , p_T , $|\mathcal{M}|$ and $\langle \mu \rangle$ (pileup) differs from signal to background. To avoid any bias these have been reweighed so their distributions are the same in signal and background.

Confirmation bias: Is when you intentionally manipulate the process to make the model, to confirm your already existing hypothesis. This can be avoided, by giving the model as few inputs from the users side as possible, and try to blind the user, so the user is unable to see the difference between signal and background, and thus cant favour one over the other [35].

3.2 Learning Algorithms

Machine learning is not a hammer, it is a screwdriver which needs to be used for the correct screw to be used optimally. This section and the following section will give an introduction to the different kinds of algorithms. With this section containing the crude difference, and the following section containing descriptions of different algorithms, which can be used to solve problems.

The first distinction made in the ML models is the difference between

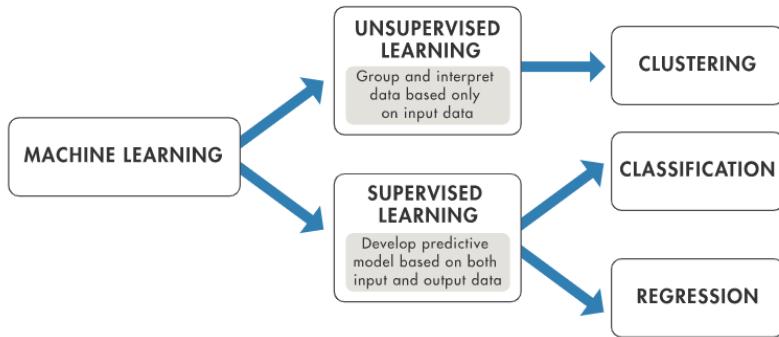


Figure 3.1: How Machine Learning can be divided into Supervised and Unsupervised learning, and how Supervised learning can be divided into classification and regression.

supervised and unsupervised learning. Where supervised learning is used for making prediction on data, and is characterized by the input being given labels, so the algorithm would know the desired output. Unsupervised learning doesn't get any labels to determine the desired results from, it will have to find structures in the input by itself[60][34].

3.2.1 Classification, regression and clustering

Supervised learning can be divided into two subgroups, classification and regression, this work done in this thesis has been done by using classification. The last learning method is clustering, which is not a part of this thesis and therefore will only the definition be given. These three types of learning methods can be seen in 3.1.

Classification: This technique is used for the prediction of discrete values, such as whether a tumor is cancerous or benign, or what iris flower you have based on its petal and sepal width and length (Multivariate data set introduced by statistician and biologist Ronald Fisher in 1936¹).

Regression: Where classification used discrete results, regression will make predictions with continuous values. This could be the predictions of temperature or any result which includes real numbers.

Clustering: Lastly there is clustering which technique used for unsupervised learning. This technique is used to find patterns or groups hidden in data, to explore how a data set might be clustered into smaller groups[60][34].

¹ https://en.wikipedia.org/wiki/Iris_flower_data_set

3.2.2 Loss function

The method used to evaluate a model is a loss function. The loss function defines an objective of the model, so the parameters may be tuned to minimizing the loss function. Similar to the models are the loss functions specialized for different problems, so a given loss function might perform better on one kind of problem rather than another problem.

Where there are several types of loss functions for regression problems, such as Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Bias Error (MBE)², there are also some different types of loss functions for classification problems. Where students are quite familiar with MAE and MSE they are normally not as familiar with the following loss functions.

Binary Cross Entropy Loss (BCEL) is the most common loss function for two class classification problems. This loss function is used to measure the difference in randomness between two random variables.

Where the loss is defined as

$$\text{BCEL} = - \sum_{i=1}^n y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i)) \quad (3.1)$$

With n being the number of data points, y_i being the measured value and $h_\theta(x_i)$ being the predicted value. This loss function also have the neat detail, that if y_i is either 0 or 1, one of the terms becomes 0 - which is rather useful for binary classification.

This method can be expanded to cases of multiple classes with the **Categorical Cross Entropy Loss (CCEL)**, which requires that there will only be one non-zero element. But is otherwise similar to BCEL. So when working on a problem, you will have to choose the loss function, which works the best for you[56].

² These won't be described any further since regression hasn't been used during this thesis

3.2.3 Gradient descent

Gradient descent is an optimization method used to find the optimized coefficients for a model, or something similar. Gradient descent tries to minimize a given cost function (f), for the best set of coefficients so $f(\text{coefficients})$ would be a global minima of the cost function. The only known values the gradient descent knows is the $f(\text{coefficients})$ of its nodes, and the slope at these locations - so it can reduce the value from that point. The size of the next step for a given node is determined by a parameter called learning rate.

Batch Gradient Descent: This method is the most common type of gradient descent used in Machine Learning. It evaluates the cost function for the entire training set for each iteration³, it then compares this prediction to the true values and calculates a sum or average error, which can be minimized. From the given cost function can the derivative for each coefficient be calculated and the new coefficient values can be determined.

Stochastic Gradient Descent: Where batch gradient descent suffers

³ With each iteration being called a batch, hence the name.

from poor speed when performed on a large dataset, this issue can be avoided by using stochastic gradient descent. This method updates the coefficients after each training instance, rather than at the end of the batch instance. Updating coefficients after each training instance results in noise, which requires randomization of the order of the training data. Doing this randomization utilises the random walk of the updates after each training instance, which will allow the algorithm to avoid getting stuck or distracted. The cost will then be calculated for one training pattern instead of summing over all the training patterns. So to reiterate will the batch gradient descent calculate all the derivates and then calculate the update, where the stochastic gradient descent will calculate the derivative and do the update immediately for each training instance[22].

3.2.4 Over- and Underfitting

The goal of supervised learning is to approximate a given target function, so the given model can be generalised from the training set to any unknown data from the problem in question. The idea is then to make approximation which follows the target function. This may however end in unwanted under- or overfitting, which both results in poor generalization.

Underfitting: A model which is neither capable of modelling the training data or generalize to new data is underfitted. This can be solved by using another machine learning algorithm.

Overfitting: A model which has learned both the signal and the noise is overfitted. This hurts the generalization since it may result in finding signal in places where there is no signal, due to being fixated on the noise. The risk of overfitting can be reduced by either using a resampling technique, such as k-fold cross validation⁴. It can also be done make validating against a validation set, instead of using only a training and test sample[43].

⁴ When you train and validate your model k times, each time on a different subset of the training data.

3.2.5 Occam's razor

Occam's razor is a simple principle which dictates that given two competing models, which perform equally well, the simpler of the two models should be used.

For machine learning this would imply that a simpler hypothesis (fewer assumptions, broader application) should be favoured when compared to an equally good complex model (more assumptions, narrower application).

An example of the complexity was discussed by Pedro Domingos in the late 1990's, where he describes Occam's two razors. With the first razor dictating that simplicity is desirable given the choice of two models with the same generalization error. The second razor dictates that two models with the same training-set error, then the simpler of the two should be preferred since it would be most likely to have a lower generalization error[19].

3.3 Machine Learning models

There exists a lot of different machine learning methods, this thesis will make use of the boosted decision tree method which is a supervised learning method, this algorithm is excellent for classification problems. The algorithm used here will be the LightGBM algorithm, due to its excellent speed when handling huge volumes of data, compared to other BDT algorithms such as XGBoost.

3.3.1 Decision tree

Tree-based algorithms are based on how trees are built, you start at the trunk and move out towards the leaves. They are as mentioned above a type of supervised learning and can be used for both categorical and continuous input/output variables.

The categorical decision trees give results that are equivalent to yes and no answers. Where the continuous variable decision tree can give results which can take a continuous value.

When discussing the decision trees some terminology will be required. A **Root Node** represents the entire sample, that can be divided into smaller sets. **Splitting** is the process that allows for nodes to divide into sub-nodes. **Decision Nodes** is the term for sub-nodes splitting into additional sub-nodes. When the tree does not split any further, a **Leaf** has been created. The algorithm may have created too many sub-nodes, which may have caused overfitting, these sub-nodes can be removed by the process of **pruning** - basically the opposite of splitting. Subsections of the tree are called **branches**. A node that creates sub-nodes is called a **parent node** while their sub-nodes are called **child nodes**.

The trees are prone to overfitting if they have been given no constraints by making too many leaves. There is also a risk of losing information when categorising continuous variables into different categories. However, decision trees are conceptual easy to understand and have intuitive graphical representations. When working with decision trees less data cleaning will be required compared to other modelling techniques, since they are less influenced by outliers or missing data. Another advantage is that there are several methods of determining the significance of variables in the models.

There are a couple of hyperparameters that can be used to control these decision trees.

The **minimum samples for node split** is a parameter that determines the minimum number of data points in a given leaf. If there were no lower bound on the number of data points the model would be prone to overfitting. The model would also be underfitted if this variable is too high. **Minimum samples for leaves** are similar in function to the minimum samples for node split and are used for the same. The **maximum depth of the tree** is also used to control over-fitting. The deeper the tree is allowed to grow the more details the algorithm learns. **Maximum number of leaves** is an alternate variable to the

max depth of the tree. **The maximum number of features considered in a split** allows the algorithm to check the relation between different variables, and more easily avoid over-fitting. **Learning rate** is used to determine the impact of a tree - where lower values are preferred since it allows the model to become more robust - at the cost of more trees required to finish the model.

A model created using only a single large tree is prone to being over-fitted and is thus inadvisable. An alternative method is random forest, which uses the average result from several different trees created at random. These forests are created by taking random samples from the training set, and use each of these samples for its own tree. A set of random variables is picked by each tree to train on, and the tree grows as large as possible - without pruning. By comparing the results from the trees the result can be found⁵. This is a very strong method, since it handles both classification and regression. Further more it is rather effective at handling large data sets, with many variables[55].

⁵ if it is a classification problem would the result be determined by majority vote, while it would be the average for a regression problem

3.3.2 Ensemble methods

An ensemble method improves the machine learning algorithms ability to predict outcome by the use of several models, where the result can be obtained by averaging over the results from these models. For this thesis two ensemble methods will be introduced, and how they are used for decision trees.

BAGGing: Is short for Bootstrap AGGregating, and the idea of it is to create multiple bootstrap subsamples from a sample of data. The most effective model can be obtained from combining decision trees trained on these subsamples.

Random Forests: This is somewhat similar to BAGGing, with a notable difference being that BAGGing had full disposal of the features for each of its decision trees. Random forests are build up using a random subset of the features used in the model. This allows for different splitting of the data, depending on which features are used. This will as a result give a greater ensemble to average over, and thus give a more accurate prediction for the final model[20].

3.3.3 Hyperparameter optimization

It is important to understand the difference between model- and hyperparameters. Model parameters are learned during the training of the model, such as the weight of a variable. Hyperparameters are set by the user and are used to set restrictions for the algorithm - such as number of leaves in a forest, its depth etc. The goal of hyperparameter optimization is to minimize the loss function, which can be obtained through a couple of different methods.

Manual Search: This requires the user to set all of the hyperparameters. The user would have to select a value, train the model and see the results and then change the hyperparameters accordingly. This process would be repeated until satisfying results have been obtained.

Random Search: The best hyperparameter values are obtained by

creating a grid in n-dimensions (with n being the number of hyperparameters), and train the model using random values in this grid⁶.

Grid Search: Similarly to random search a grid will be used to optimize the model. The initial value could be the best result from a random search, to see if a better result can be achieved in a local region of the gridspace.

Automated Hyper parameter tuning: The hyper parameters are determined through the use of a technique, this technique could be: Bayesian Optimization, Gradient Descent or Evolutionary Algorithms.

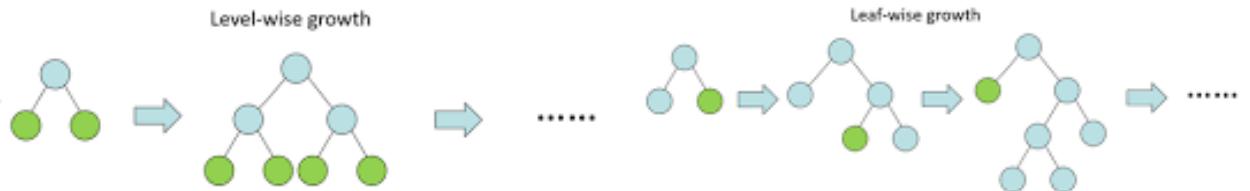
Bayesian Optimization makes use of probability to find the minimum value of the loss function. This method is more efficient than the Manual, random and grid search described above. It requires three parameters, the Objective function which is the loss function, domain space which defines the range of the hyperparameters, and the optimization algorithm which determines the search algorithm used to select the input values of the new iterations. **Gradient descent** have already been explained in section 3.2.3. **Evolutionary algorithm** makes use of the idea of natural selection. It creates N models with predefined values for the hyperparameters, where the accuracy of all the models are determined. The models with the best accuracy (normally 50% of the models) are kept and offspring are generated from combinations of these models[30].

⁶ The training process is run several times using a combination of new random hyperparameter settings

3.3.4 LightGBM

LightGBM is a gradient boosting framework which uses tree based machine learning algorithm. But it differs from other tree based algorithms, such as XGBoost in the fact that it doesn't grow the tree one level at a time.

LightGBM grows its tree by selecting the leaf with the highest delta-



loss and growing the tree from that point. Given a fixed number of leaves this will result in a better minimization of the loss function, compared to what an algorithm using level-wise growth would obtain. A visual representation of these two kinds of growth can be seen in fig 3.2[33].

This leaf-wise growth allows LGBM to run much faster with a lower memory usage. It does however have one disadvantage, the algorithm is prone to overfitting. It can thus be a favourable to use LGBM as long as you have more than 10.000 datapoints[59], and use another

Figure 3.2: Graphical representation of the difference between level-wise and leaf-wise tree growth. Examples of algorithms using these methods of growing would be XGBoost using the level-wise tree growth and LightGBM using the leaf-wise tree growth.

model with smaller datasets. Another way to counteract this overfitting would be to use one of LGBM's features: Max_depth. This parameter determines the maximum depth of the tree, reducing this variable may improve the model in case of overfitting.

The traditional method for split the leaves would be by using Feature Parallel splitting.

Feature Parallel in traditional algorithms: The partitions are traditionally made vertically⁷. The workers survey their local split to determine the best local split, they then proceed to communicate their findings to the others so the best global split may be found. This best split is then performed, and the other workers are notified, so they may split the data they received accordingly. This method suffers from speed and memory usage, since the speed of this process can't be sped up when larger amounts of data are used, the communication of the cuts are also costly at one bit per data.

Feature Parallel in LGBM: To counteract the computation issue LGBM has decided to have all the workers carry all the data, instead of doing vertical partitioning⁸. The feature parallel in LGBM proceeds through the following steps: A worker finds the local best split, they communicate the best split with each other and then they perform this split. Doing the features parallel still suffer from doing the splits overhead⁹, it can therefore be wise to use other methods given too large datasets.

LGBM would normally use data parallel splitting.

Data Parallel in Traditional algorithms: The aim of Data parallel is to parallelize the decision learning, which requires the data to be partitioned horizontally. After partitioning local histograms are constructed using the data given to the workers. These local histograms are merged to a global histogram, and the best split can be found from here. This method suffers from a high cost of communication, which may slow the process down.

Data Parallel in LGBM: In LGBM the communication cost has been reduced by using "Reduce scatter" to merge the histograms of different workers, for different non-overlapping features. From here the workers can find the local best split, and compare to the global best split. This histogram subtraction allows LGBM to only communicate histograms for one leaf, the neighbors histogram can be achieved through subtraction as well[33].

⁷ This may however vary from algorithm to algorithm

⁸ This saves the algorithm from communicating the partitioning between the different workers

⁹ "Any combination of excess or indirect computation time, memory, or other resources that are required to perform a specific task"[44].

4

Data processing and selection

As mentioned in section 2.7 there are no scarcity regarding the data volume in the ATLAS experiment. This will give a short introduction to how the ATLAS experiments limits its datafiles to what is relevant for the different groups.

4.1 Data files and derivation

During Run 1, the job of reducing the size of the data from PB to TB was done by individual users. This was a taxing system, both due to needing a lot grid jobs¹ to accomodate the downscaling, but also due to difficulties with cross-team analysis - since the different teams would be using different structure in their data format. In an attempt to fix this issue in Run 2, would this down scaling be made centrally[13]. This downscaling is called derivations, and would allow the users to just download the DxAOD files². These DxAOD files doesn't include all the data, which allows for a smoother and faster analysation. The main operations in derivations are the following[13][45]:

- **Skimming:** Removing the irrelevant events
- **Thinning:** Removing the irrelevant objects from within an event, but keeping the rest of the event
- **Slimming:** Removing the irrelevant information from within objects, but keeping the rest of the object

These new DxAOD files are tailor made by the different groups in ATLAS, which allows limitation of the events and variables in the DxAODs to the ones relevant to their analysis. The relevant DxAODs for this thesis are given in tab 4.1[16][38][26].

4.2 Ntuple production

The DxAODs created by the groups are accessible for all ATLAS members via Rucio, which is a data management system, from which they can be downloaded. These DxAOD files are however still quite big and has to be reduced to smaller files, which can be achieved by creating Ntuples. The process follows the same principle as the

name	Selection
EGAM ₃	$Z \rightarrow ee\gamma, Z \rightarrow eee$
EGAM ₄	$Z \rightarrow \mu\mu\gamma$
EGAM ₇	e and γ Background
MUON ₁	At least one CB muon with $p_T > 25\text{GeV}$ and $ \eta < 2.5$, with a second muon with $p_T > 4\text{GeV}$ and $\mathcal{M}_{\mu\mu} > 70\text{GeV}$.
HIGG1D ₁	Skim for $H \rightarrow \gamma\gamma$
HIGG1D ₂	Skim for $H \rightarrow Z\gamma$

Table 4.1: Different derivations used in this thesis. EGAM derivations are produced for specific decays, while MUON and HIGG derivations are using different skimming events.

¹ Jobs run on ATLAS server, which has access to vast computing power and direct access to all the datasets - no download required before running the job.

² Derived xAOD, where a xAOD file is the file type used to save all the events produced in ATLAS - both for Monte Carlo and Data)

ones used for DxAODs, however, this is done locally by the user. The ntuple production used in this thesis was developed by a previous student, Helle Leerberg [32]. The idea here is to only save the relevant particles and variables. So when looking at EGAM3 files, where you are interested in the interaction $Z \rightarrow ee\gamma$ will no muons be saved. A more detailed description of this process can be found in Helle Leerbergs[32] and Sara Dahls[45] theses. After the ntuples have been created, will these be converted into hdf5 files, from which the analysis are created. The derivation types and their processes are given in tab 4.2, a complete list of the files used can be found in appendix A.1.

Process	Derivation	Model
$Z \rightarrow ee$	EGAM3	eIso, ePid, ee, ee γ
$Z \rightarrow ee\gamma$	EGAM3	eIso, ePid, ee γ , ee, γ Iso, γ Pid
$Z \rightarrow \mu\mu$	MUON1	μ Iso, μ Pid, $\mu\mu$
$Z \rightarrow \mu\mu\gamma$	EGAM4	μ Iso, μ Pid, $\mu\mu\gamma$, $\mu\mu$, γ Iso, γ Pid
$H \rightarrow \gamma^*\gamma$	HIGG1D2	γ Iso, γ Pid, $\gamma\gamma$
$H \rightarrow \gamma\gamma$	HIGG1D1	γ Iso, γ Pid, $\gamma\gamma$
$H \rightarrow Z(l\bar{l})\gamma$	HIGG1D2	γ Iso, γ Pid, eIso, ePid
$W \rightarrow ev$	EGAM1	eIso, ePid
$W \rightarrow \mu\nu$	MUON1	μ Iso, μ Pid
Jets	EGAM7	eIso, ePid, γ Iso, γ Pid

Table 4.2: The different processes, the derivations used for them, and which models and reconstructions they are used for.

5

Monte Carlo models

This chapter will describe the method used to create the models. All of the models are created using the steps described below.

SIGNAL SELECTION: Determining which particles in the events are signal and background is the first step to create the models. Some particles may also be discarded if deemed irrelevant. The criteria for being a signal or background particle differs between models. The special case of the models with two leptons should be noted. In these cases "The tag and probe method"[15] is used, this is a data-driven technique used to measure the efficiency of particle detection. It uses the decays of known resonances to pair particles. The particle pairs consist of a tag and a probe particle, which differs by how well they are defined - the tag has tighter selection criteria than the probe. Doing these selections for MC data makes use of truth variables, these only exists in MC data and tells the user what a particle really is.

REWEIGHING: Some of the data selections may have the issue of lacking background. This was mainly a problem for the single particle data sets. To avoid training too skewed models has a minimum requirement for the background been set to 90% of the signal. If there is less background than this, then the background will be added until sufficient background is present.

After the selection and possibly adding additional background the reweighing is performed. The reason for reweighing is to avoid biases in the model - see section 3.1.2. The variables reweighed are η and $\langle \mu \rangle$, if it is a muon data set is p_T also reweighed otherwise e_T and lastly if the signal and background is a combination of multiple particles, such as electron pairs, is the invariant mass also reweighed.

This reweighing should result in a weighted distribution where signal and background is quite similar, which should lessen the risk of miss classification. The method used for reweighing was Gradient Boosted Reweighting (GBReweighting), which is based on an ensemble of regression trees. This method allows several variables to be reweighed, so distributions of similar signal and background may be acquired. A previous student - Helle Leerberg[32], did some experimentation in

this field, and found that of the input parameters: The number of estimators (nEst), maximum sample in leaf (max_sample_leaf), and maximum depth (max_depth), the one with the greatest influence on the output was nEst.

There are two options when doing the reweighing, the conventional way where background is reweighed to resemble the signal (Regular), or where signal is reweighed to resemble the background (Reverse). Both of these are trained and saved for the training, validation and test sets, performance of the different weights on the validation set is determined by the user. The user may then choose either the regular or reverse reweighing scheme, and the number of estimators, and use these as input for the training models.

The other parameters for the GBReweighter are learning rate set to 0.2, the max_depth is 3, the min_samples_leaf is 200 and the loss_regularization is 5 (the approximate number of events in each leaf to prevent the model from "exploding").

TRAINING AND VALIDATION: LGBM is used to train on the reweighed data, as described in chapter 3. For these model have the hyper parameters been set to num_leaves=30, max_depth=-1 (no limit), min_data_in_leaf=30, feature_fraction=1 (all the variables in the models will be used in each of the trees), bagging_fraction=1 (All the data will be used in each of the training iterations), the learning_rate=0.05, the number of boosting rounds is set to 2500 and the early stopping is set to 500. The selection of these are based on the work of Helle Leerberg[32].

CHOICE OF VARIABLES FOR THE MODELS: One of the most important things for a model are the variables used for training, it can be tough to determine which variables to use and which to leave be. Here is an example of the method used on the model for photon identification - the other models used similar method.

At first a training model with all the PID variables is trained, from which the SHAP (SHapley Additive exPlanations) values can be obtained. This is a game theoretical approach to explaining the output of a machine learning model[58], its use is to determine the importance of variables used in the model. After determining the SHAP values is the variable with the lowest score removed, and the model is trained again. These runs are compared to an initial guess of variables, all of these can be found in table 5.1.

It is now time to validate the possible PID models , which is done by creating ROC-curves and compare these, the ROC-curve shows the performance of a classification model, it is plotted with the false positive rate (background classified as signal) and true positive rate (signal classified as signal)[11].

As seen in 5.1 are most of the models fairly equal in their ability to predict on the data, except for the initial guess model, which achieves an AUC score which is 4.5% lower than what can be achieved by the

test # /Variables	0	1	2	3	4	5	6	7	8	9	10
pho_isPhotonEMLoose	-	x	x	x	x	x	x	-	-	-	-
pho_isPhotonEMTight	-	x	x	x	x	x	x	x	x	x	x
pho_Rhad1	x	x	x	x	x	x	x	x	x	x	x
pho_Rhad	x	x	x	x	x	x	x	x	x	x	x
pho_weta2	x	x	x	x	x	x	x	x	x	x	x
pho_Rphi	x	x	x	x	x	x	x	x	x	x	x
pho_Reta	x	x	x	x	x	x	x	x	x	x	x
pho_Eratio	x	x	x	x	x	x	x	x	x	x	x
pho_f1	x	x	x	x	x	x	x	x	x	x	x
pho_wtots1	x	x	x	x	x	x	x	x	x	x	x
pho_DeltaE	-	x	x	x	x	x	x	x	x	x	x
pho_weta1	-	x	x	x	x	x	x	x	x	x	-
pho_fracs1	-	x	x	x	x	x	x	x	x	-	-
pho_ConversionType	x	x	-	-	-	-	-	-	-	-	-
pho_ConversionRadius	x	x	x	x	x	x	-	-	-	-	-
pho_VertexConvEtOverPt	-	x	x	-	-	-	-	-	-	-	-
pho_VertexConvPtRatio	-	x	x	x	-	-	-	-	-	-	-
pho_zo	-	x	x	x	x	x	x	x	-	-	-
pho_zoSig	-	x	x	x	x	-	-	-	-	-	-
pho_core57cellsEnergyCorrection	-	x	x	x	x	x	x	x	x	x	x
pho_r33over37allcalo	x	x	x	x	x	x	x	x	x	x	x
NvtxReco	x	-	-	-	-	-	-	-	-	-	-
correctedScaledAverageMu	x	-	-	-	-	-	-	-	-	-	-

Table 5.1: The variables used for the various versions of the photon Pid model. Test # 0 is the initial guess.

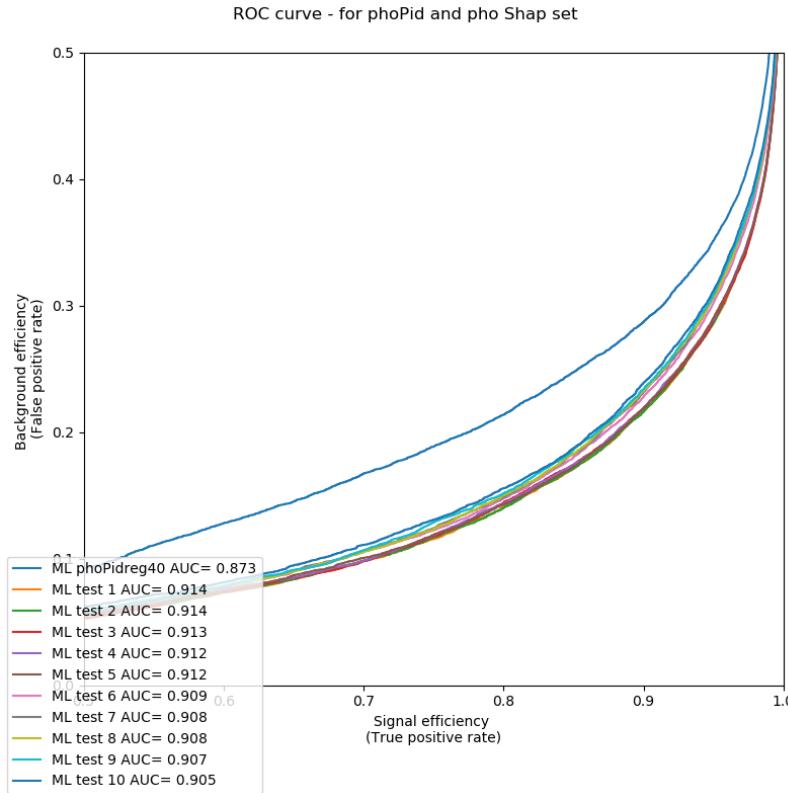


Figure 5.1: The ROC-curve for the different training tests. "phoPidReg40" is the one called test o in table 5.1. The x-axis shows the range 0.5 – 1 and the y-axis show the range 0 – 0.5.

models with the most variables(The AUC score is the "Area Under the Curve", and the perfect model has a value of 1 while a random model has a value of 0.5).

A small issue arose after having run the comparison, the variable `pho_core57cellsEnergyCorrection` is not found in the data files for $e\gamma\gamma$ and $\mu\mu\gamma$, it was also found that it had a poor performance in simulation, which would result in poor performance when used on real data.

Instead of rerunning all of the models the second test model was selected and `pho_core57cellsEnergyCorrection` was removed from it. This was compared to the initial guess¹. It was assumed there would be a similar dependence of `pho_core57cellsEnergyCorrection` for all the models, therefore were only the second model rerun for comparison.

This resulted in the initial guess having an AUC score of 0.873 and the revised test 2 having an AUC score of 0.884. For this thesis the test 2 model has been used, so the highest precision can be achieved. In hind side should the simpler of the two models have been used, as Occam's razor dictates, since the improvement of the AUC is too small compared to the added complexity.

¹I picked the second model since it shared a first place with the first test model for highest AUC score. One may argue that a model with fewer variables should have been used, since the difference in AUC score is negligible.

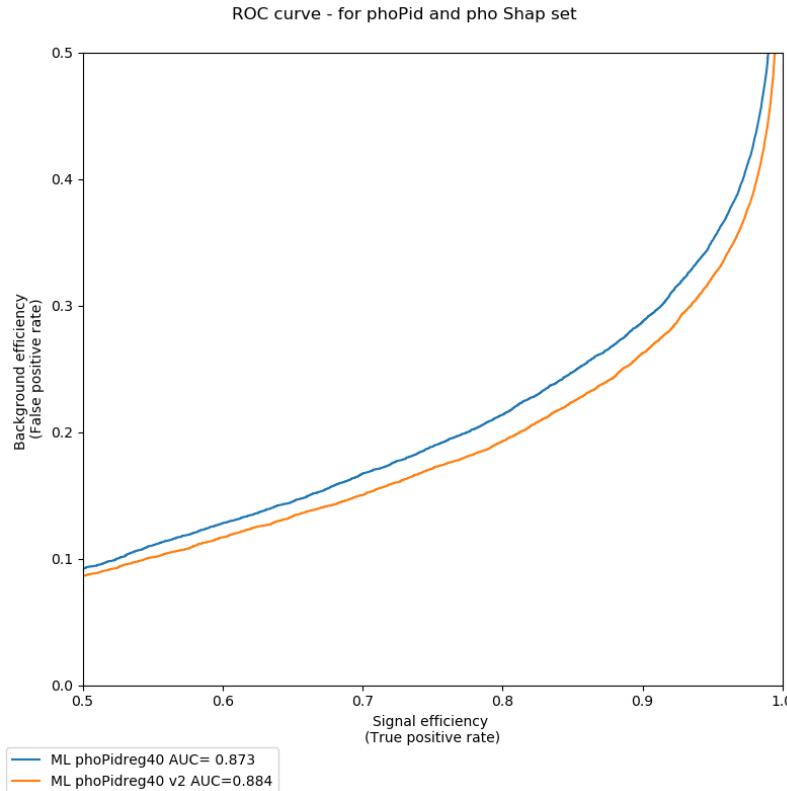


Figure 5.2: Comparison of the ROC-curves of the initial guess model and test 2 model without the pho_core57cellsEnergyCorrection variable.

5.1 Single lepton and photons

This section will go into more details regarding the method used to select lepton and photon candidates. There will be two models for each particle type, PID (Particle IDentification) and ISO(particle ISolation).

5.1.1 Data selection for leptons and photons

ELECTRON MODEL DATA SELECTION:

The data files used for the electron models are the $ee\gamma$, electron pairs, $H \rightarrow Z(ee)\gamma$, $W \rightarrow ev$ and some jets. The W decay and the jets are used to increase the amount of background found - which should improve the reweighting and training.

An electron candidate is taken into account if it has an "ele_truthType" value, after having found all of the candidates they are looped over, to determine whether they are signal, background or should be discarded.

Ele PID: For electron identification it is only required that the ele_truthPdgId_atlas value is 11 - this pdg value is the value for electrons[41], otherwise they are background.

The E_T distribution for this selection is shown in fig 5.3². The majority of the low energy candidates, sub 20GeV, are background, while signal dominates at energies higher than 20GeV.

² The distribution of isolated leptons and photons can be found in the appendix in section A.2

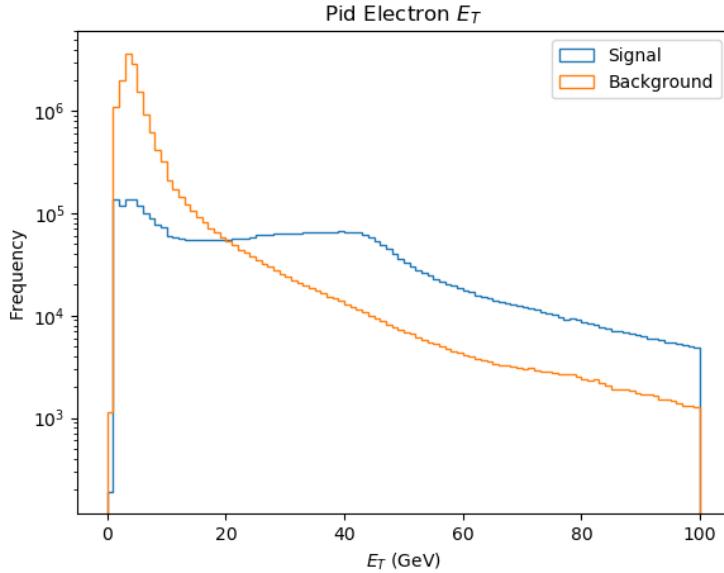


Figure 5.3: The transverse energy for the electrons chosen with the Pid method. The signal is displayed in blue, while the background is displayed in orange.

Ele ISO: The requirement for isolated electrons are determined through the use of the "truthtype" variable[37], which tells the user what particle it is. If the candidate is an isolated electron it is categorized as signal while non isolated electrons and unknown particles are the background.

The distribution of the transverse energy for the isolated electrons are like the PID distribution where the low energy candidates are dominated by background - candidates with transverse energy below 15GeV, while higher energies are dominated by signal.

MUON MODEL DATA SELECTION:

The data files used for the muon models are the $\mu\mu\gamma$, muon pairs, $H \rightarrow Z(\mu\mu)\gamma$, and lastly, $W \rightarrow \mu\nu$ files. Where the W decay is to improve the amount of background in the model.

Similar to the electrons, all candidates with a "muo_truthType" are taken into consideration.

PID muons: It is required that the candidates has a PdgId value of 13(muon) to be signal, otherwise it is background. The transverse momentum distribution in fig 5.4 shows the distribution being dominated by background for candidates with transverse momentum below 20GeV. The distribution is signal dominated in the transverse momentum range of 20 – 70GeV, for transverse momentum above 70GeV surpasses the background the signal.

Isolated muons Similarly to the electron is a muon candidate signal if it has a truthtype of isolated muon, and it is background if it has either a truthtype of non isolated muons or unknown. The transverse momentum distribution for the isolated muons shown in fig 5.5, are dominated by background for transverse momentum below

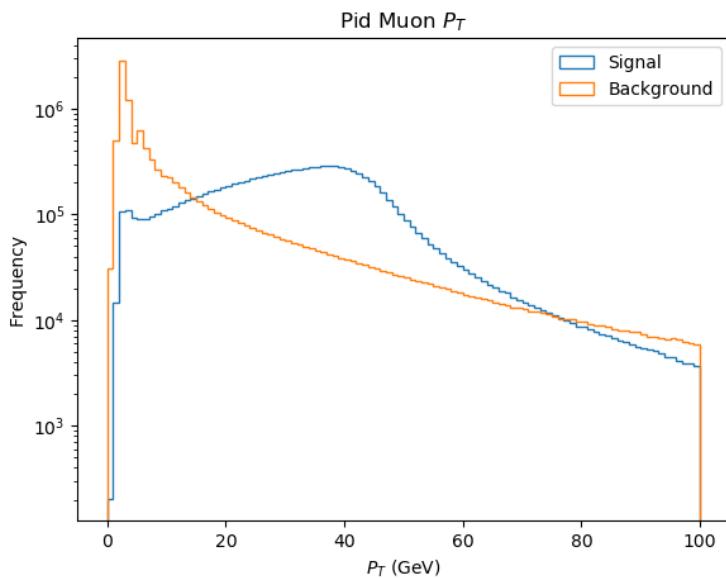


Figure 5.4: Transverse momentum of Pid muons, with the signal displayed in blue and background displayed in orange.

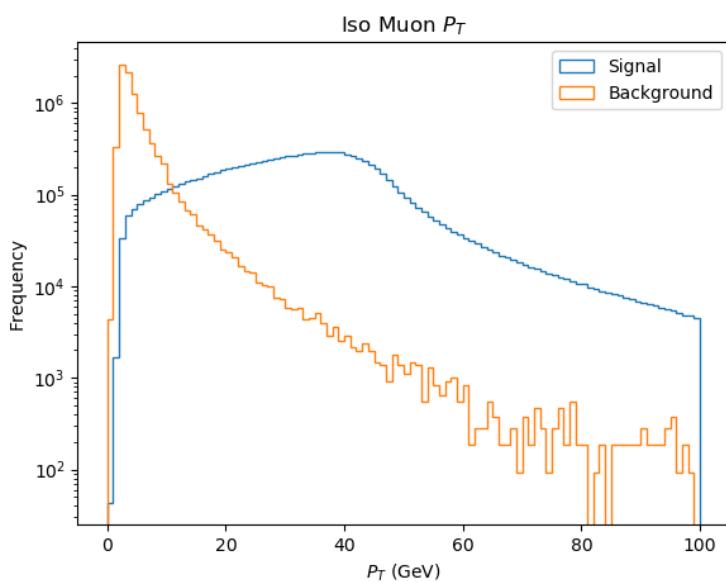


Figure 5.5: Transverse momentum of Iso muons, with the signal displayed in blue and background displayed in orange.

15GeV and dominated by signal at momentum higher than this. The background for the isolated muons drops off much quicker than the background of the identified muons.

PHOTON MODEL DATA SELECTION:

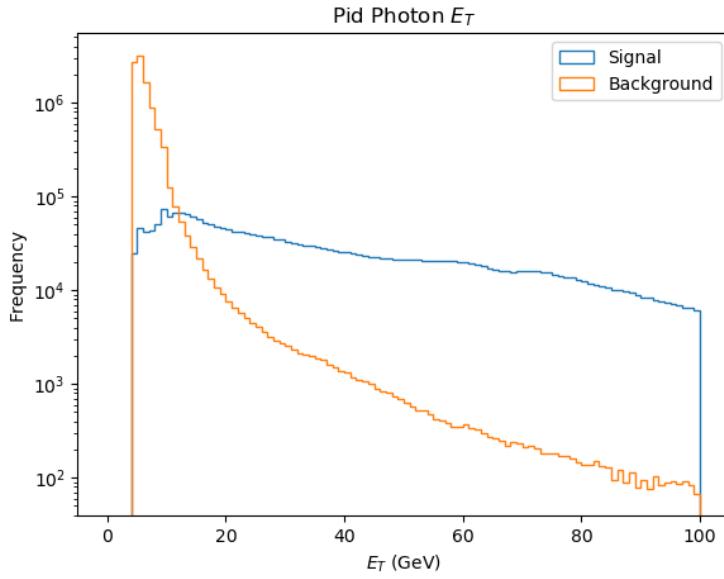


Figure 5.6: Transverse energy of Pid photons.

Like the leptonic selections the candidates are required to have a "pho_truthType" value, due to a lot of low energetic photon candidates is it required of the candidates to have at least 4.5GeV of transverse energy. This is done to save a lot of computation time.

PID photons: For a photon candidate to be considered signal it is required to be either an isolated or non-isolated photon. It is considered background if it is either unknown, isolated electrons, non-isolated electrons, background photons, L-, G- or Unknown-jets. This selection, as seen in fig 5.6, resulted in a broad range of energies for signal and a lot of low energy background. The distribution is completely dominated by background for candidates with less than 15GeV of transverse energy, after which the background quickly drops off. Whereas the signal has a much slower decline.

Isolated photons: Lastly, are the isolated photons, which is required to be either an isolated photon or electron. Lastly, for the single particle models are the isolated photons³. The photon candidate is background if it is either a non isolated electrons, unknown photons, non isolated photons, hadrons, and L-, G- or an Unknown-jets. The distribution in transverse energy for the isolated photons is basically identical to the distribution for identified photons, but can be seen in appendix A.2.

³ In this case isolated electrons are included since, they could origin from a photon which pair converted

5.1.2 Reweighting of leptons and photons

Before doing any reweighting it is required of the data to have a somewhat equal amount of signal and background. A minimum amount of background was required, if there was less background than 90% of the number of signal additional background would be added. The only data set which required this additional background was the isolated muons, which had their background increased by a factor of 93. After this a cut will be applied to remove all low transverse energy/momentun candidates, a cut of 4.5GeV for leptons and 9.5GeV for photons is applied - The number of signal and background candidates after this cut is displayed in 5.2.

Pid Electron sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 4.5\text{GeV}$	3.732.861	0.889	7.059.930	0.457
Iso Electron sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 4.5\text{GeV}$	2.840.473	0.960	2.641.337	0.333
Pid Muon sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 4.5\text{GeV}$	9.964.298	0.973	6.238.583	0.560
Iso Muon sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 4.5\text{GeV}$	10.338.372	0.988	3.551.670	0.377
Pid Photon sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 9.5\text{GeV}$	2.763.477	1.0	9.819.342	1.0
Iso Photon sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
$E_T > 9.5\text{GeV}$	2.846.571	1.0	4.001.945	1.0

After reweighting the MC files the reweighted distributions can be plotted, the weighted distributions for PID electrons are displayed in fig 5.7⁴.

The reweighted electron PID distributions with 10 estimators did poorly, especially for the pileup at low and high values. The weights with the other number of estimators did better, but it was found that the regular reweighting scheme with 40 estimators did better overall. It was found that using 100 estimators for the weights of the isolated electrons resulted in distributions that did quite poorly. 10 estimators did slightly better, but not as well as with 40 estimators. For this thesis, the weights found using the reverse reweighting scheme with 40 estimators will be used for the isolated electrons.

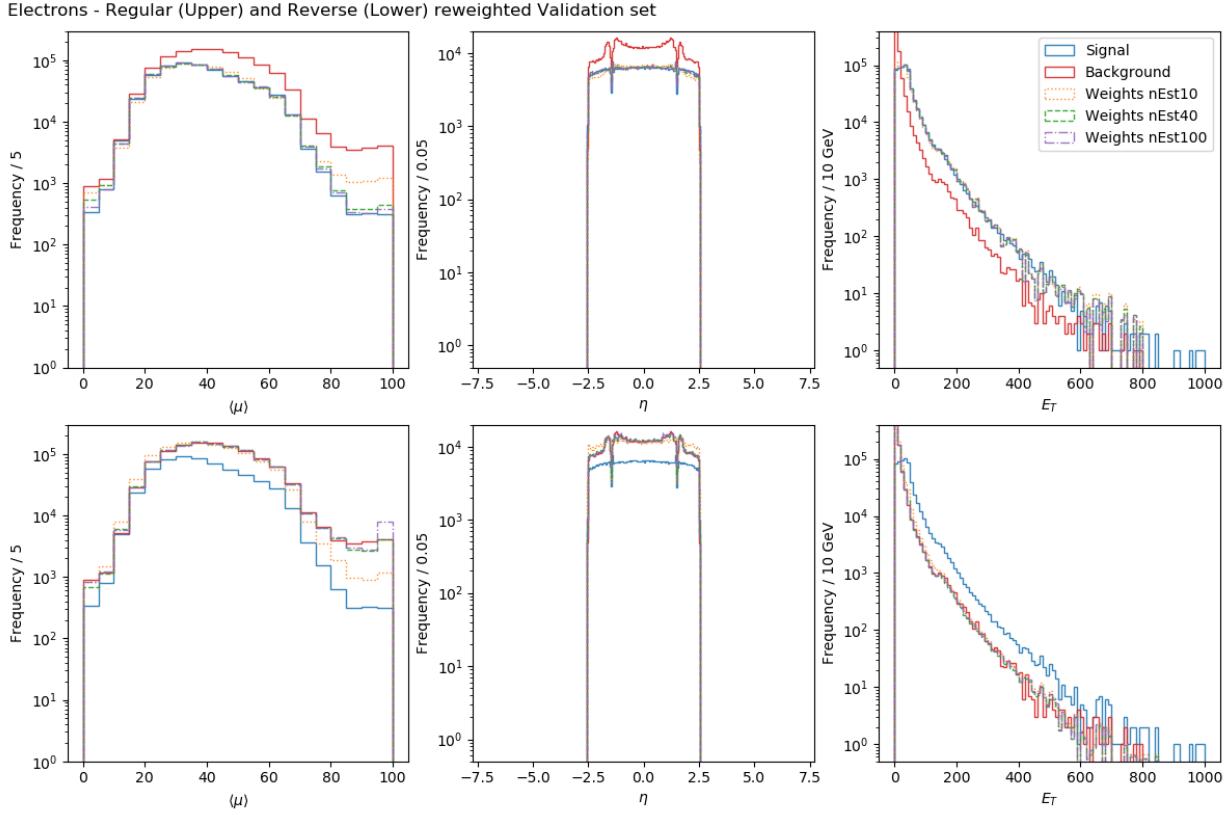
All the reweighting schemes for the isolated muons did quite well, though the regular reweighting scheme with 10 estimators did slightly worse than the others, for example see high pileup. By using Occam's razor, the weights for the isolated muons were found to be the ones from the reverse reweighting scheme with 10 estimators.

It was found that using 10 estimators for the identified muons resulted in poor performance. when looking at the performance of the rest of the weights, were the one using 40 estimators with the regular reweighting scheme chosen to be used for the training.

It was found that using the regular weights with 100 estimators resulted in extremely poor performance, with the only other outlier

Table 5.2: The number of signal and background particle in the lepton and photon data sets, before and after removing the leptons (photons) with less than 4.5GeV (9.5GeV). It should be noted here that the background for the isolated muons have been increased by a factor of 93, from 101.331 to 9.423.783 before the cuts.

⁴ The others can be found in Appendix A.3



being the reverse weights using 10 estimators which also performed poorly. For training of the model the regular weights using 10 estimators will be used for the isolated photons.

Using weights generated using 100 estimators for the identified photons resulted in poor results, after looking at the distributions it was that using the regular weights with 40 estimators performed the best, and will therefore be used for training.

5.1.3 Training and evaluation of Leptons and photons

ELECTRON TRAINING AND EVALUATION: The two electron models make use of different variables. The variables used for the isolation model are found in tab 5.3. They mainly rely on the ratio between the sum of the transverse momentum or energy tracks in a cone around the electron compared to the total transverse energy of the electron candidate.

The variables used for the identified electrons are found in tab 5.4. Where the variables focus on the particle candidates journey through the detector, and how it deposited its energy.

Using these two sets of variables allows for training of model which determines whether an electron candidate is identified as an electron

Figure 5.7: Signal and background distribution of $\langle\mu\rangle$, e_T and η for identified electron reweighed using different number of estimators.

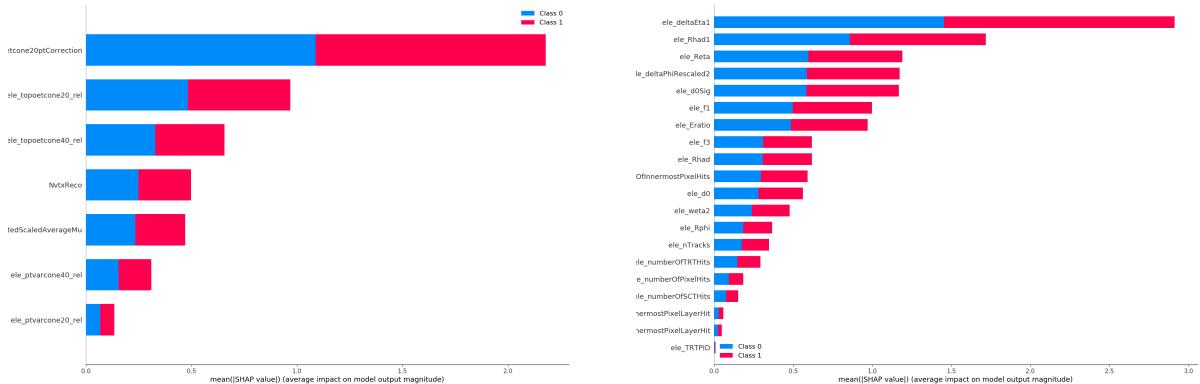
Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
$\langle \mu \rangle$	CorrectedScaledAverageMu	Pileup
	ele_ptvarcone20_rel	The sum of p_T tracks in a 0.20 cone around the object, divided by the transverse energy
	ele_ptvarcone40_rel	The sum of p_T tracks in a 0.40 cone around the object, divided by the transverse energy
	ele_topoetcone20_rel	The sum of E_T tracks in topoclusters in a 0.20 cone around the object, divided by the transverse energy
	ele_topoetcone40_rel	The sum of E_T tracks in topoclusters in a 0.40 cone around the object, divided by the transverse energy
	ele_topoetcone20opt-Correction	Correction to the topoetcone20

Table 5.3: Isolation variables for the electron[48].

and another model to determine whether the electron candidate is isolated.

Similarly to the example of the PID photons earlier in this chapter, SHAP was used to determine how significant a variable was in the training of the model. The SHAP values for the electron models are shown in fig 5.8. Where it can be seen that the variables describing the TRT and the closest pixels to the candidate, had little to no impact on the Pid model, where $\Delta\eta_1$ had a much larger significance on the model.

To evaluate how well these models are performing, it will be required



to have an ATLAS reference point. The isolation model is evaluated against the track isolation $p_T^{\text{varcone}0.2}/E_T < 0.15$ [17], while the PID

Figure 5.8: shap values obtained from training the PID electron models.

Name	Feature	Description
d_0	ele_do	The impact parameter - the distance from the particles closest point to the beam,
σ_{d_0}	ele_doSig	The uncertainty in the impact parameter
R_{had1}	ele_Rhad1	Ratio of E_T in the first layer in the hadronic calorimeter to the E_T in the EM cluster
R_{had}	ele_Rhad	Ratio of E_T in the hadronic calorimeter to E_T in the EM cluster
f_3	ele_f3	Ratio of energy in the back layer to the total energy in the EM accordian calorimeter(For particles with energies below 100GeV, otherwise known to be inefficient at higher energies)
$w_{\eta 2}$	ele_weta2	Lateral shower width in the ECAL
R_ϕ	ele_Rphi	Ratio of energy in 3x3 cells over energy in 3x7 cells
R_η	ele_Reta	Ratio of energy in 3x7 cells over energy in 7x7 cells
E_{ratio}	ele_Eratio	Ratio of energy difference between largest and second largest energy deposit.
f_1	ele_f1	Ratio of energy in the strip layer to the total energy in the ECAL
$\Delta p/p$	ele_dPOverP	Momentum loss through detector divided by original momentum
$\Delta\eta_1$	ele_deltaEta1	Difference in η from cluster position in the strip layer and the extrapolated track
$\Delta\phi_{res}$	ele_deltaPhiRescaled2	Difference in ϕ between cluster position in the middle layer and the extrapolated track
	ele_expectInnermostPixel-LayerHit	How many pixel are hit in the innermost layer
	ele_expectNextToInnermost-PixelLayerHit	How many pixel are hit in the next to innermost layer
	ele_core57cellsEnergyCor-rection	Correction to the calculated energy of 5x7 core cells in the electron calo cluster
	ele_nTracks	The number of tracks
	ele_numberOfInnermost PixelHits	Number of hits in the Pixel detector
	ele_numberOfPixelHits	Number of hits in SCT
	ele_numberOfSCTHits	Number of hits in the TRT
	ele_numberOfTRTHits	Likelihood probability based on transition radiation in the TRT
eProbabilityHT	ele_TRTPID	

Table 5.4: Pid variables for the electrons[17].

model are evaluated against the loose, medium and tight Working Points. The results from these ATLAS cuts will give an FPR and TPR value, which can be inserted into a roc-curve for the models. Doing this will allow us to find the models TPR, when the model has the same FPR as ATLAS.

The results from the cuts performed by ATLAS, and the correspond-

Model	FPR ATLAS	FPR LGBM	difference	TPR ATLAS	TPR LGBM	difference
eIso Tight	0.8096	0.8096	0	0.8136	0.9891	21.6%
ePid Loose	0.0131	0.0131	0	0.7318	0.8710	19.0%
ePid Medium	0.0043	0.0043	0	0.6697	0.7893	17.9%
ePid Tight	0.0010	0.0131	0.0121	0.5935	0.6674	12.5%

ing results from the models are shown in tab 5.5. The ISO model obtained an improvement in TPR of 21.6% compared to ATLAS, while the PID model achieved an improvement in (Loose, Medium, Tight) of (19.0%, 17.9%, 12.5%) compared to ATLAS.

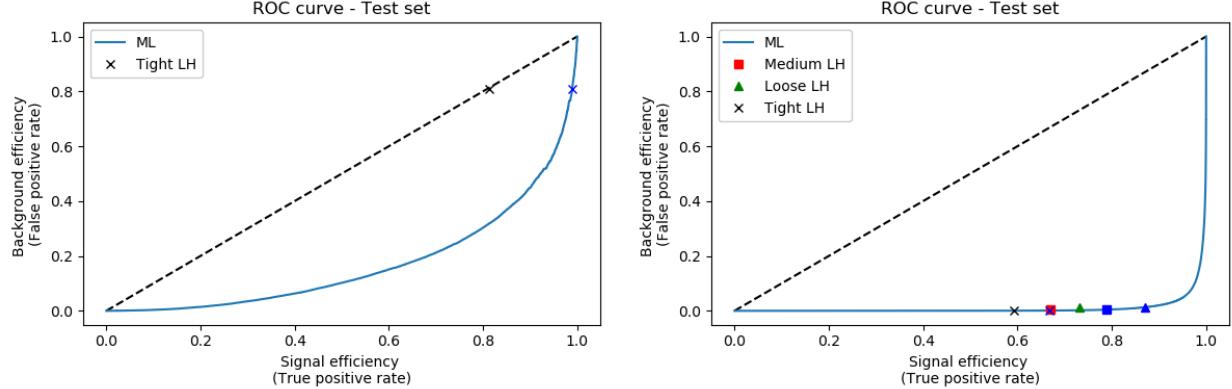
The PID model overall performed better than the ISO model, see tab 5.6 where the PID model achieved a significantly higher AUC score than the ISO model. The ISO model did however obtain a better improvement in % compared to the identification, which is due to it being easier to do identification of a candidate rather than isolation of it.

Using the track isolation to find the the candidates ATLAS deemed

Model	Unweighted AUC score	Weighted AUC score
eIso	0.823245	0.829532
ePid	0.981197	0.989837

Table 5.5: The false positive and true positive rates for ATLAS cuts and the models used in this model (referred to as LGBM). Besides the values of the FPR and TPR are the numeric difference between this thesis' models and the ATLAS cut in FPR given, and the percentage difference between the TPR are also given.

Table 5.6: The unweighted and weighted AUC for the two electron models.



isolated resulted in a result where the FPR and TPR values were basically equal to each other, fig 5.9, where as the similar TPR obtained by the model shows an improvement - and a result which isn't basically random.

When looking at the ROC-curve for the PID model, an overall improvement was found compared to ATLAS.

Figure 5.9: Roc curves plotted for test sets of the electron models, with FPR and TPR values from the the ATLAS cut and their corresponding values in the models.

MUON TRAINING AND EVALUATION: Similarly to the electron models are the two muon models using different variables. The isolation variables used are shown in tab 5.7, and the PID variables are shown in 5.8.

The PID variables for the muons differs slightly due to how the

Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
$\langle \mu \rangle$	CorrectedScaledAverageMu	Pileup
	muo_ptvarcone20_rel	The sum of p_T tracks in a 0.20 cone around the object, divided by the transverse energy
	muo_ptvarcone40_rel	The sum of p_T tracks in a 0.40 cone around the object, divided by the transverse energy
	muo_etconeCoreCone-EnergyCorrection	Correction to the energy loss for muons in all layers of calorimeter

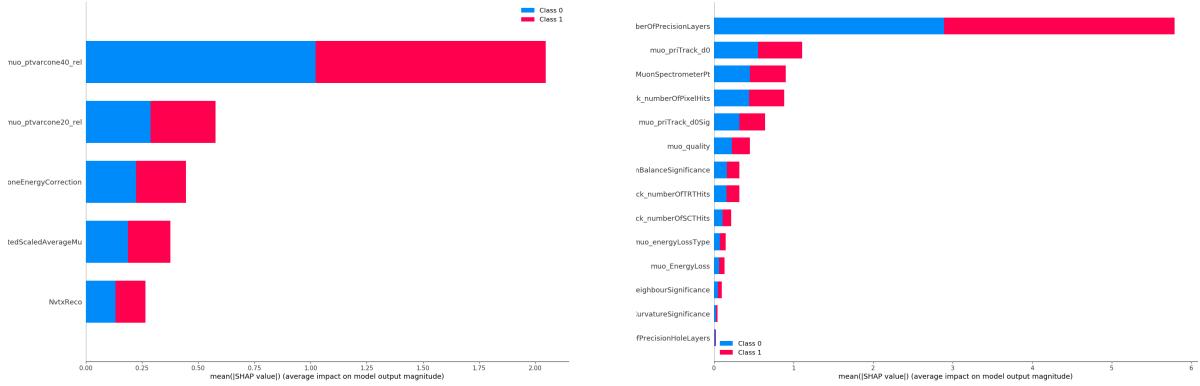
Table 5.7: Isolation variables for the muon[48].

muons interact in the detector compared to the electrons. There are no variables referring to the TRT, η , ϕ or number of hits in Pixel, SCT and TRT. After running the two models, the SHAP values are

Name	Feature	Description
d_0	muo_priTrack_do	The impact parameter - the distance from the particles closest point to the beam,
σ_{d_0}	muo_priTrack_doSig	The uncertainty in the impact parameter
	muo_numberOfPrecisionHoleLayers	
	muo_numberOfPrecisionLayers	
	muo_quality	
	muo_MuonSpectrometerPt	
	muo_scatteringCurvatureSignificance	
	muo_scatteringNeighbourSignificance	
	muo_momentumBalanceSignificance	
	muo_EnergyLoss	
	muo_energyLossType	

Table 5.8: Pid variables for the muon[17].

calculated in fig 5.10, with the most important isolation variable being the relative ptvarcone with 0.4 in radius, and the most important identification variable is the one regarding the number of precision layers. The AUC score for the muons are shown in tab 5.9, where it can be noted that the isolation model for muons had a better AUC when the weights had not been used, but otherwise performed better than the electron isolation model. The PID model for the muons performed almost as well as the electron PID model.



ATLAS considers a muon candidate isolated if it has $p_T^{varcone30} / p_T < 0.06$ [42], and the candidate is considered identified as a muon if it passes one of the working points. Using these cuts to determine the improvement achieved with the models, an improvement in isolation of 15.3% can be achieved, while identification has been improved by (3.6%, 1.1%, 9.1%) for (Loose, Medium, Tight). The smaller improvements of muons compared to electrons are due to ATLAS being better at identifying muon using the working points.

Model	Unweighted AUC score	Weighted AUC score
mIso	0.935665	0.894899
mPid	0.961305	0.966881

Figure 5.10: SHAP values obtained from training the muon models.

Table 5.9: AUC values for the muon and isolation models.

Model	FPR ATLAS	FPR LGBM	difference	TPR ATLAS	TPR LGBM	difference
mIso Tight	0.2523	0.2523	0	0.7711	0.8893	15.3%
mPid Loose	0.0972	0.0972	0	0.8917	0.9239	3.6%
mPid Medium	0.0826	0.0826	0	0.8802	0.8895	1.1%
mPid Tight	0.0725	0.0725	0	0.7967	0.8694	9.1%

The two ROC-curves for the muon models are displayed in fig 5.11. It was found that the AUC for the isolation model was lower than that of the identification model. The reason for the jumps in the PID roc-curve was not determined during the work on this thesis.

PHOTON TRAINING AND EVALUATION:

Last of the single particle models are the photon models, which has similar variables as the electron, with a few differences - which was expected since photons can convert to electrons through pair production, and electrons can radiate photons through bremsstrahlung. The variables used for the isolation model for the photons are found in tab 5.11, while the variables used for the photon identification model can be found in tab 5.12.

The photon models differ slightly to the other single particle models, since the ATLAS cuts have been determined by whether they

Table 5.10: The false positive and true positive rates for ATLAS cuts and the models used in this model (referred to as LGBM). Besides the values of the FPR and TPR are the numeric difference between this thesis' models and the ATLAS cut in FPR given, and the percentage difference between the TPR are also given.

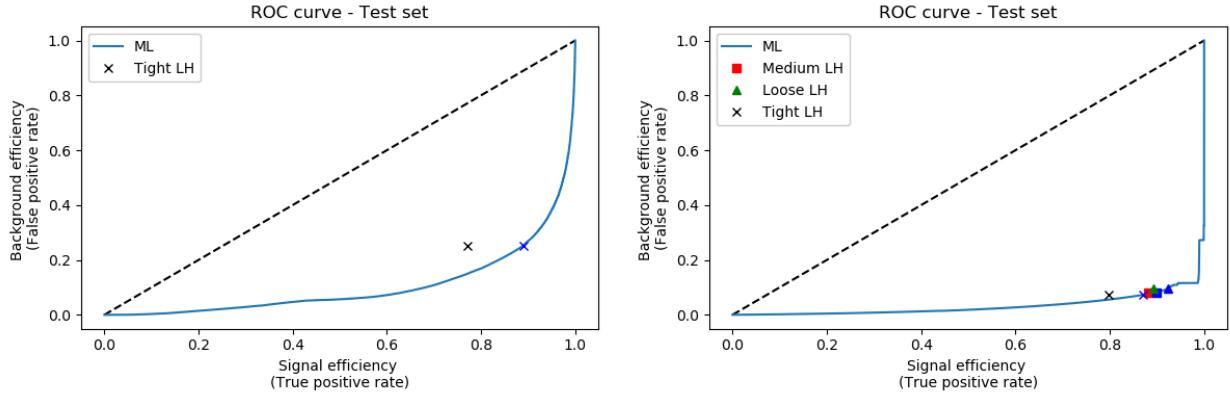


Figure 5.11: Roc curves and the FPR and TPR values from ATLAS for the muon models.

Table 5.11: Isolation variables for the photon[48].

Name	Feature	Description
NvtxReco		Number of reconstructed vertices
$\langle \mu \rangle$	CorrectedScaledAverageMu	Pileup
pho_et		
pho_ptvarcone20		The sum of p_T tracks in a 0.20 cone around the object, divided by the transverse energy
pho_topoetcone20_rel		The sum of E_T tracks in topoclusters in a 0.20 cone around the object, divided by the transverse energy
pho_topoetcone40_rel		The sum of E_T tracks in topoclusters in a 0.40 cone around the object, divided by the transverse energy

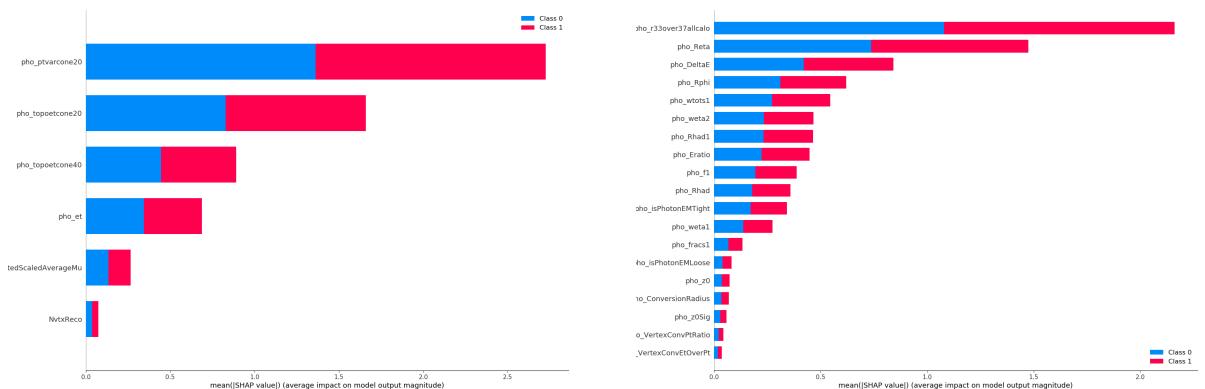


Figure 5.12: SHAP values obtained from training the photon models.

Name	Feature	Description
R_{had1}	pho_Rhad1	Ratio of E_T in the first layer in the hadronic calorimeter to the E_T in the EM cluster
R_{had}	pho_Rhad	Ratio of E_T in the hadronic calorimeter to E_T in the EM cluster
$w_{\eta 2}$	pho_weta2	Lateral shower width in the ECAL
R_ϕ	pho_Rphi	Ratio of energy in 3x3 cells over energy in 3x7 cells
R_η	pho_Reta	Ratio of energy in 3x7 cells over energy in 7x7 cells
E_{ratio}	pho_Eratio	Ratio of energy difference between largest and second largest energy deposit.
f_1	pho_f1	Ratio of energy in the strip layer to the total energy in the ECAL
	pho_isPhotonEMLoose	Whether the photon passes the loose criteria
	pho_isPhotonEMTight	Whether the photon passes the tight criteria
	pho_wtots1	Lateral shower width in the EM strip layer
	pho_DeltaE	Uncertainty in the energy
	pho_weta1	Lateral shower width in the ECAL
	pho_fracs1	
	pho_ConversionRadius	The radius of conversion
	pho_VertexConvEtOverPt	
	pho_VertexConvPtRatio	
	pho_zo	Origins displacement from interaction point
	pho_zoSig	Uncertainty in origins displacement
	pho_r33over37allcalo	

Table 5.12: Pid variables for the photon[17].

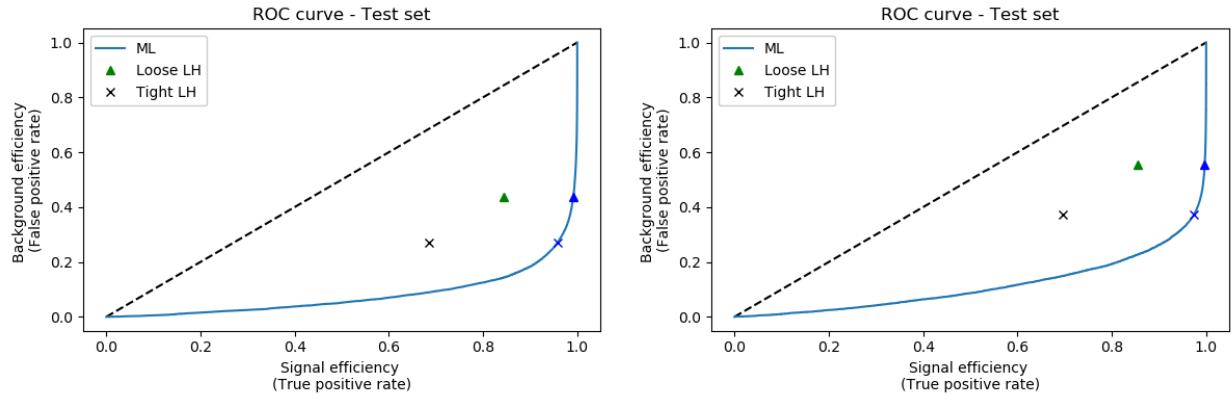
passed the isPhotonEMLoose or isPhotonEMTight cuts - which was used for both models.

Similar to the PID muon models the AUC score is lower for the

Model	Unweighted AUC score	Weighted AUC score
pIso	0.950364	0.920814
pPid	0.929919	0.884324

photon models if the weights are used, see tab 5.13. As can be

Table 5.13: Weighted and Unweighted AUC values for the two photon models.



seen in tab 5.14, the isolation model did achieve an improvement of (17.5%, 39.9%) in the (Loose, Tight) WP, while the identification model achieved similar values of (16.8%, 40.0%) for its (Loose, Tight) WP.

These results are reinforced when shown graphically in fig 5.13, where improvements have been made compared to the results achieved by ATLAS.

Figure 5.13: Roc curves and the FPR and TPR values obtained from ATLAS cuts, for ISO (left) and PID (right) photons.

Model	FPR ATLAS	FPR LGBM	difference	TPR ATLAS	TPR LGBM	difference
pIso Loose	0.4364	0.4364	0	0.8437	0.9915	17.5%
pIso Tight	0.2712	0.2712	0	0.6851	0.9584	39.9%
pPid Loose	0.5555	0.5555	0	0.8538	0.9969	16.8%
pPid Tight	0.3737	0.3737	0	0.6956	0.9742	40.0%

Table 5.14: The TPR and FPR for the ATLAS cuts and the improvements achieved by the photon models.

5.2 Lepton and photon pairs

In the previous sections only one particle was selected at a time, for this section multiple particles are combined into pairs. To do this it will be required to find a smart way to select the primary and secondary particle. The method used here will differ slightly depending on whether it is the leptonic pair or the photon pair.

5.2.1 Data selection for lepton pairs

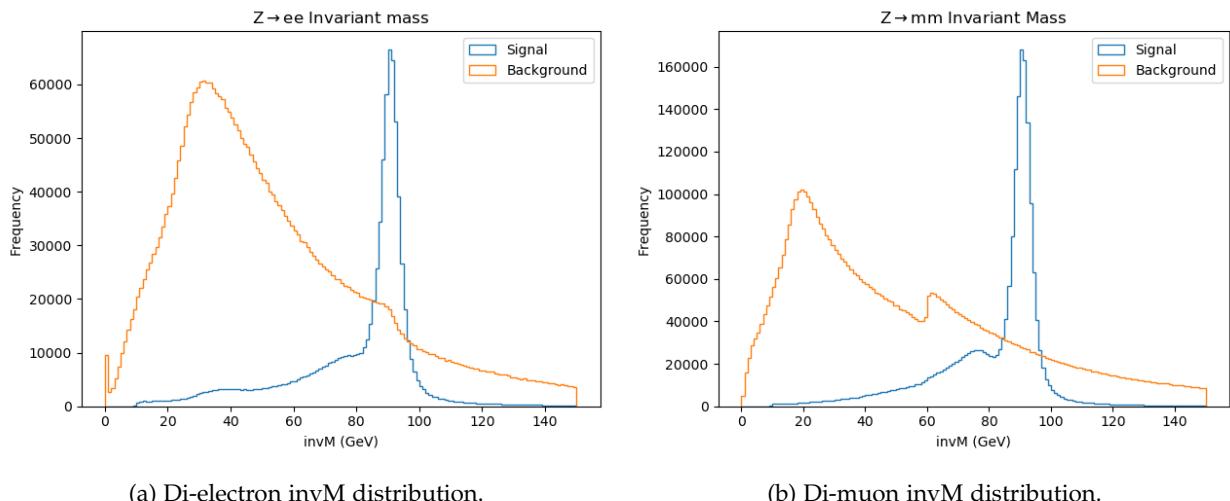
For the leptonic pair selections, candidates with a ele_et (muo_pt) value are taken into account. As mentioned above the tag and probe method is used, so all the lepton candidates which are triggers⁵, are classified as tag leptons, while the rest are probe leptons.

⁵ There are also some other restrictions imposed by ATLAS, but for this thesis it is only important to know whether it is a trigger lepton or not

It is required that events have at least two same flavour leptons - with transverse energy (momentum) over 4.5GeV, with one of them being a trigger lepton, otherwise the events are discarded. If there are multiple tag leptons in an event, they are added to the probe leptons - the first tag lepton would be paired with all other lepton after which the second tag lepton would be paired with all but the first tag (they have already been paired), this should give the most possible combinations.

After creating these pairs, it will be required to determine whether they are signal or background pairs.

ELECTRON SELECTION: Both electrons in the pair are required to origin from a Z boson, have a PdgId value of 11 - they have to be electrons, and they are required to have opposite charge - for charge conservation. Furthermore, they are required to be truthmatched. If all of these requirements are fulfilled, the pair is categorized as signal, if they both origin from a Z boson, but if they are missing one of the other requirements, the pair is discarded. If only one or none of the electrons origin from a Z boson the pair is considered background. The MC files used for this model, are the files with electron pairs, $ee\gamma$, $H \rightarrow Z(ee)\gamma$ and lastly $W \rightarrow ev$. The distribution achieved with this selection is displayed in fig 5.14. If the files with $ee\gamma$ had been omitted from the selection, the signal would loose a lot of the lower energy pairs - this would result in only a peak around the Z mass. The files are kept, so the model may recognize lower energy electron pairs better - which will make a better reconstruction model later on.



MUON SELECTION: Both muons in the pair are required to origin from a Z boson, have a PdgId of 13 and have opposite charge. However, there were no "truthmacthed" variable for the muons, unlike the electrons. If all the requirements are fulfilled the pair is considered a signal pair. If they both origin from a Z boson but are missing one

Figure 5.14: Distribution of invariant mass for the di-leptons.

of the other requirements the pair is discarded. If one or neither of the muon candidates origin from a Z boson the pair is categorized as background.

Similarly to the electron pairs, the MC files included here are chosen so they include the following interactions: muon pairs, $\mu\mu\gamma$, $W \rightarrow \mu\nu$ and $H \rightarrow Z(\mu\mu)\gamma$.

It is once again the files with $\mu\mu\gamma$ which causes a small signal peak below 80GeV, in fig 5.14. The reason for the two peaks in the background at around 20 and 60GeV are due to the distributions in the different file.

5.2.2 Photon pair selection

Where the lepton pairs used the tag and probe method, another method would be required for the photon pairs. The T&P method can not be used, since there are no tag photon - there is no trigger variable for photons.

The photon candidates are sorted, so the candidate with the highest transverse energy is first. After sorting are all the photon combined to pairs - the highest energy photon candidate is combined with all the lower energy photons, the second highest candidate is combined with all the candidates with lower energy etc. It was here required that all of the photon candidates should have more than 9.5GeV of transverse energy.

If both of the candidates origins from a Higgs boson, and they both have a PdgId value of 22, the pair is considered signal. If they have the right origin but are not both photons, then the pair is discarded. If either one or none of the photons origins from a Higgs boson, they are categorized as background.

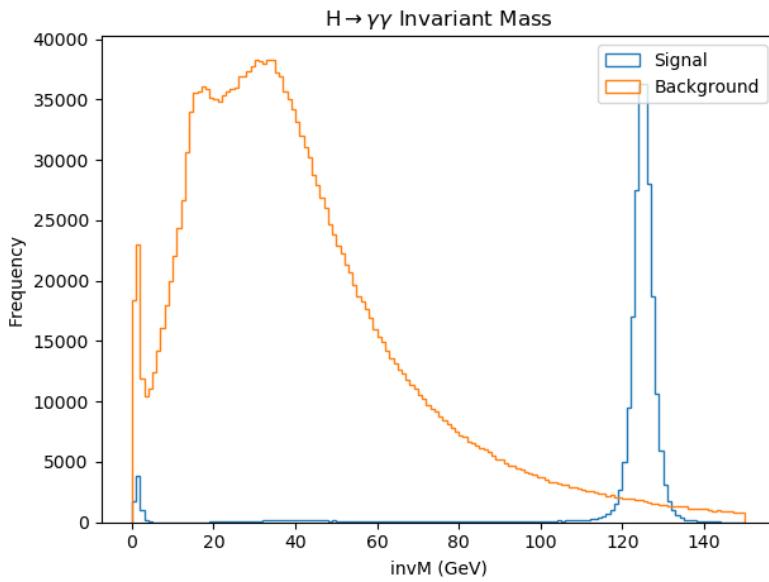


Figure 5.15: The invariant mass of the photon pairs. With the Higgs peak at 125GeV, most of the background has a low mass, the small peak at ≈ 0 GeV arises due to conversions.

The MC files used for this selection are the files with the $H \rightarrow \gamma^{(*)}\gamma$ interaction, which includes Higgs created through Vector Boson fusion and gluon fusion. The distribution for these files are given in fig 5.15. The invariant mass distribution is completely dominated by background at low energies, but has almost no background at the energies, where there is signal.

Ideally additional background with higher energy would have been used, unfortunately were no files with this combination found while working on this thesis. The initial idea was to include some files with jet+jet to achieve a background at higher energies⁶. The jet files in question had some technical difficulties, and were thus omitted from the thesis.

5.2.3 Reweighting

Similar to the single particle models, the low and high energy pairs will be removed before reweighting the data. For these pairs are the pairs reweighed in pseudorapidity η , pileup $\langle\mu\rangle$, Invariant mass of the pair and either the transverse energy or momentum (the muon). The lower limit of invariant mass for the leptonic(photon) pairs is set

Z→ee sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
All Zee	868.170	1.0	3.781.750	1.0
$ \mathcal{M} > 50\text{GeV}$	782.551	0.901	1.839.894	0.487
$ \mathcal{M} < 150\text{GeV}$	767.007	0.980	1.561.703	0.849
Z→μμ sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
All Z→μμ	2.037.652	1.0	6.510.000	1.0
$ \mathcal{M} > 50\text{GeV}$	1.900.180	0.933	3.311.101	0.509
$ \mathcal{M} < 150\text{GeV}$	1.884.761	0.992	2.584.254	0.780
H → γγ sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
All H → γγ	225.753	1.0	2.135.389	1.0
$ \mathcal{M} > 80\text{GeV}$	212.962	0.943	233.172	0.109
$ \mathcal{M} < 150\text{GeV}$	212.924	1.0	201.877	0.866

to 50(80)GeV, while the upper limit of invariant mass is set to 150GeV. The results from applying this cut is shown in tab 5.15. Applying these cuts to the pairs resulted in a large reduction in the amount of background, compared to signal which was lost. The electron dataset was the dataset which lost the most signal, with a loss of 12%, compared to its reduction in background of 59%. The dataset which lost the most background was $H \rightarrow \gamma\gamma$, where only 9% of the background is kept, while 94% of the signal is kept.

The reweighted distributions are shown in fig 5.16.

it was found that the weights found using 10 estimators did worse than the other weights for the electron pair. With the reverse weights using 100 estimators also ran into some issue at the edges of the distribution. Of the remaining three weights (regular 40, regular 100, reverse 40) the regular reweighting scheme with 40 estimators will be used. This selection is based on Occam's razor, followed by picking

⁶ If you look in the appendix A.1, you can find that a jet file is included. This gave only low energy background with energies below 60GeV. Thus most of this background would be removed in the reweighing process.

Table 5.15: The number of signal and background particle in the different data set, before and after using a minimum invariant mass cut of 50GeV (80GeV) for leptons (photons), and after using a maximum invariant mass cut of 150GeV.

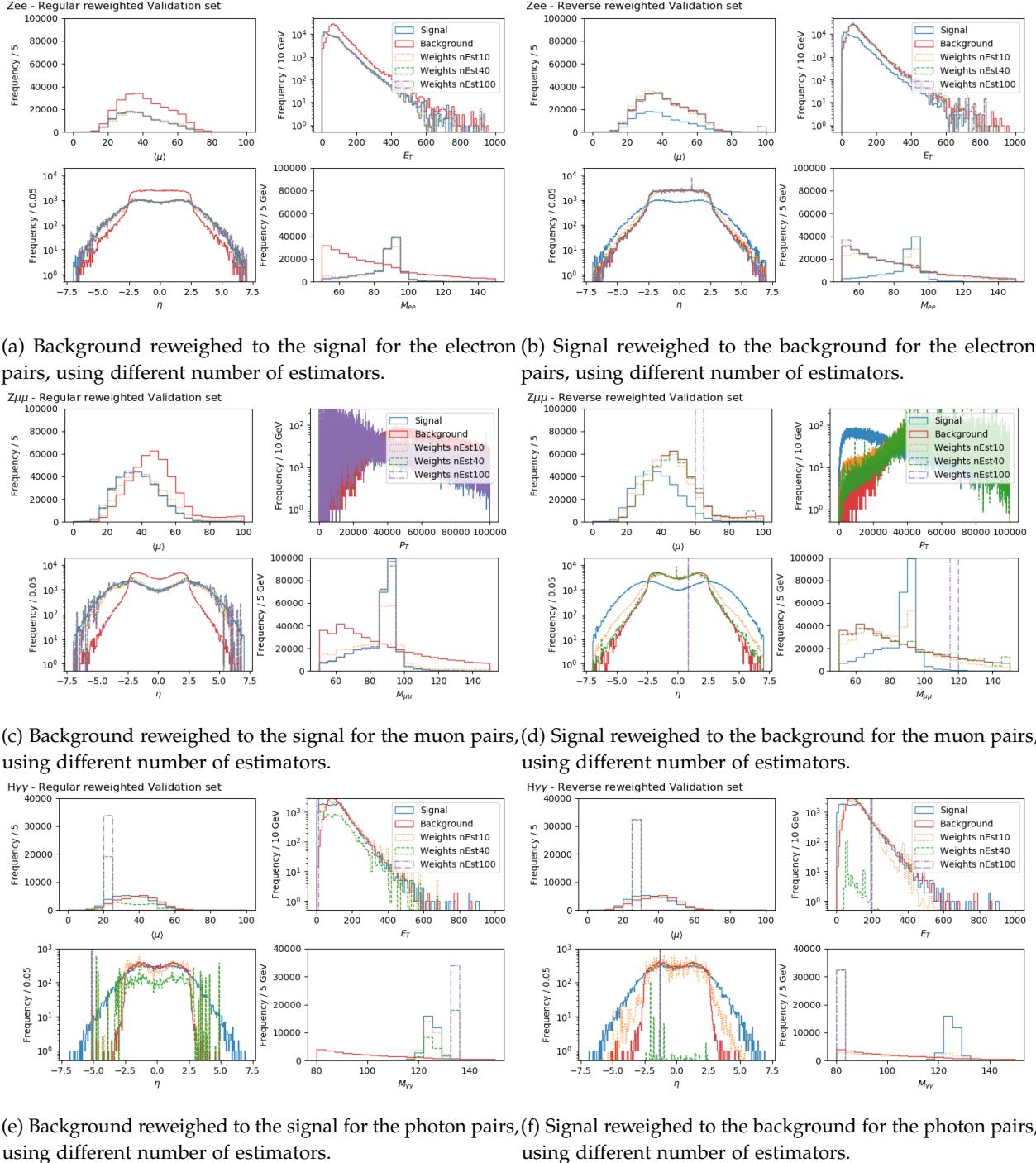


Figure 5.16: Plots of the signal and background distributions for the pair selections, and their reweighted values in $\langle \mu \rangle$, E_T , p_T , μ and $M_{ll}/\gamma\gamma$. Larger plots of this can be seen in the appendix in section A.4.

either of the models using 40 estimators, since they seemed to be doing equally well.

The muon pairs had similarly to the electron pair poor performing weights if 10 estimators had been used. Using the reverse weights with 100 estimators also resulted in poor results, using 40 estimators did slightly better than the 100 but it had still too many outliers. This leaves the regular weights with 40 and 100 estimators, which seems to be doing equally good. Thus referring to Occam's razor will the training on muon pairs be performed using the regular weights with 40 estimators.

The weights for the photon pairs with 40 estimators did very poorly, with the weights with 100 estimators not doing any better. The best choice left is the regular weights with 10 estimators.

5.2.4 Training and evaluation

Now that the models for single leptons and photons have been trained, selections of the data for pairs of leptons and pairs of photons has been made - and reweighed in sensitive variables. It is now time for training a model for these pairs. Many of the variables used in the single particle models, does also make sense to use here, these have been included by using the results from the single particle model. The variables used for the pair models are the combination of the results from the single particle models and the combined variables: ΔZ_0 , which is defined as $|Z_{0,1} - Z_{0,2}|$ and its uncertainty $\sigma_{\Delta Z_0} = \sqrt{\sigma_{\Delta Z_0,1}^2 + \sigma_{\Delta Z_0,2}^2}$. The electron model is also taking $\Delta d_0 = |d_{0,1} - d_{0,2}|$ and its uncertainty $\sigma_{\Delta d_0} = \sqrt{\sigma_{\Delta d_0,1}^2 + \sigma_{\Delta d_0,2}^2}$. Lastly the muon pair model also makes use of $\Delta Z_0 \sin(\theta)$ and the Z-vertex. A full list of electron pair variables is found in tab 5.16, similarly for muon pairs in tab 5.17 and photons pairs in tab 5.18.

Name	Feature	Description
NvtxReco		Number of reconstructed vertices
correctedScaledAveragemu	Pileup	
ele_deltad0	Difference in d_0	
ele_deltad0sig	Uncertainty of Δd_0	
ele_deltaZ0	Difference in Z_0	
ele_deltaZ0sig	Uncertainty of ΔZ_0	
ele1_ePid_score		
ele2_ePid_score		
ele1_eIso_score		
ele2_eIso_score		

Table 5.16: Variables used for the electron pairs.

The SHAP values obtained when training these models are displayed in fig 5.17 for electron pairs, fig 5.18 for muon pairs, and fig 5.19 for photon pairs. The most important variable for all of the models are the PID scores of the secondary particle. When comparing two

Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
	correctedScaledAverageMu	Pileup
	muo1_delta_z0	
	muo2_delta_z0	
	muo1_delta_z0_sin_theta	
	muo2_delta_z0_sin_theta	
	muo1_vertex_z	
	muo2_vertex_z	
	muo1_mPid_score	
	muo2_mPid_score	
	muo1_mIso_score	
	muo2_mIso_score	

Table 5.17: variables used for the muon pairs.

Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
	correctedScaledAverageMu	Pileup
	pho_deltaZ0	Difference in Z_0
	pho_deltaZ0sig	Uncertainty of ΔZ_0
	pho1_pPid_score	
	pho2_pPid_score	
	pho1_pIso_score	
	pho2_pIso_score	

Table 5.18: Variables used for the photon pairs.

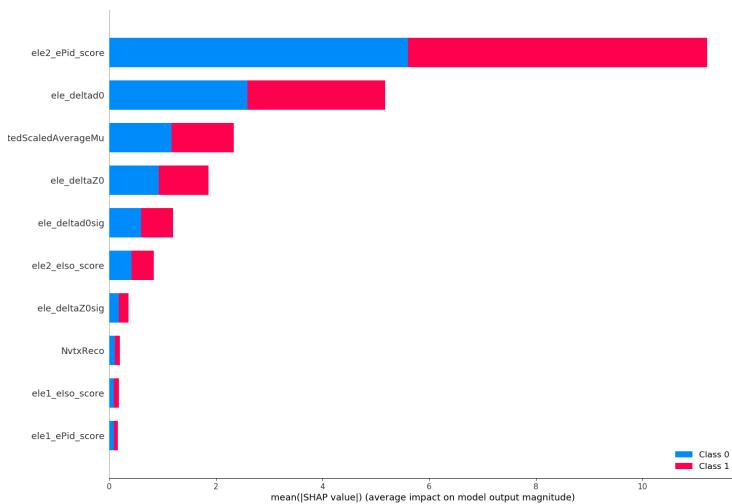


Figure 5.17: Electron pair SHAP values, with the PID score for the secondary electron being the most impactful.

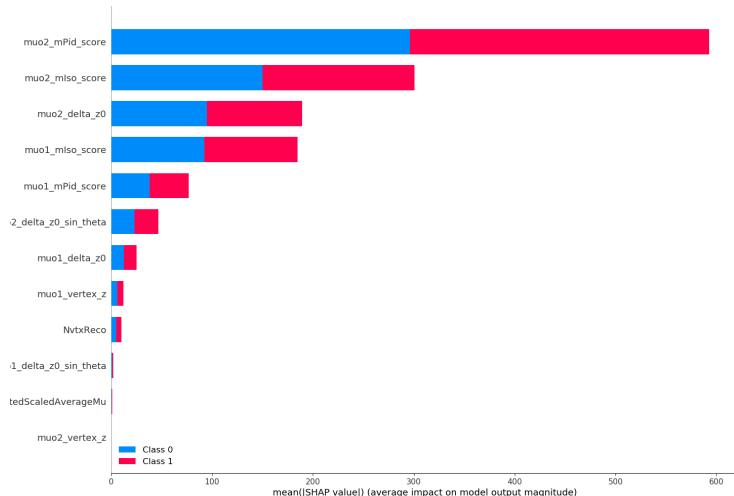


Figure 5.18: Muon pair SHAP values, with the PID score for the secondary muon being the most impactful.

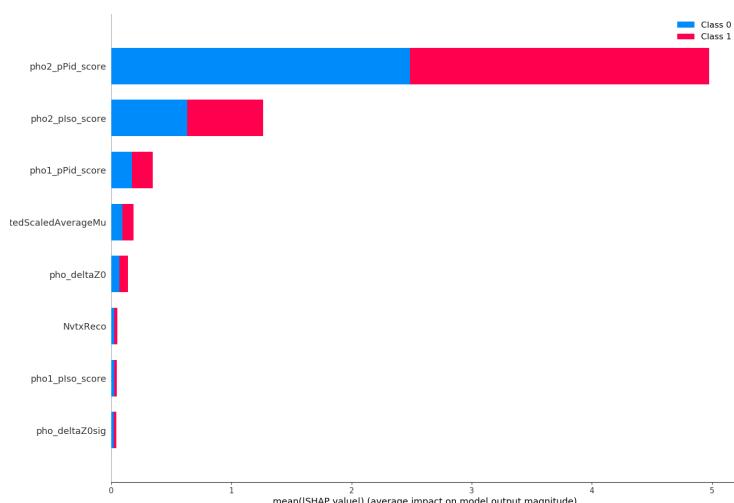


Figure 5.19: Photon pair SHAP values, with the PID score for the secondary photon being the most impactful.

variables, one from primary and the same variable for the secondary, the one from the secondary particle will be more important to the model than the one from the primary particle.

After obtaining training models it is time to compare these models to what ATLAS would have obtained, the exact variables and values varies from particle pair to particle pair. These will be introduced here[1]. It is required that both electrons (muons) have an η value which passes the cuts of not being in the crack, which is in the range of $1.37 < |\eta| < 1.52$ they are furthermore required to be below 2.47 (2.7). The ratio for their d_0 is required to fulfil $d_0/\sigma_{d_0} < 5(3)$, and the absolute value of the longitudinal impact parameter times sin to the polar angle has to be below 0.5mm, so $|\Delta Z_0 \sin(\theta)| < 0.5$. The leptons are required to have higher transverse energy (momentum) than 10GeV, and pass the loose (medium) working point. Last but not least, the leptons are required to have opposite charge - this should hold true per definitions by how the pairs have been selected. One problem rose from these ATLAS cuts. The variable $|\Delta Z_0 \sin(\theta)|$ had not been saved for the electron models - which was realised at this point in the process. To avoid re-running the selection and reweigh for the dataset, the variable was calculated by multiplicating the value of Z_0 and $\sin(\theta)$, where $\theta = 2 \arctan(e^{-\eta})$.

The ATLAS requirements of the photons pairs are[40] that both the primary and secondary photon has to pass the Tight working point, similarly to the leptons, the photon candidates are not allowed to be in the crack, which was in the interval $1.37 < |\eta| < 1.52$ and the absolute value of the pseudorapidity also has to be lower than 2.37. It was also required that the primary (secondary) photon had higher transverse energy than 35(25)GeV, and the ratio of transverse energy to the invariant mass should pass $e_T/\mathcal{M}_{\gamma\gamma} > 0.35(0.25)$. It is also required that both the primary and secondary photon must pass $p_T^{0.2}/e_T < 0.057$.

⁷It is also required that they pass $e_T^{0.2}/e_T < 0.065$, but the et_cone20 variable wasn't in the data file[40]

Model	FPR ATLAS	FPR LGBM	difference	TPR ATLAS	TPR LGBM	difference
ee	0.0250	0.0250	0	0.3909	0.6519	66.8%
$\mu\mu$	0.0035	0.0035	0	0.7230	0.8139	12.6%
$\gamma\gamma$ Tight	0.0320	0.0320	0	0.5690	0.5806	2.0%

Using these at the ATLAS cuts to compare the FPR and TPR values gives the values in tab 5.19. Here, the test sample for the electron pair model achieved an improvement of 66.8%, and test sample for the muon pair model achieved an improvement of 12.6%. Lastly, there is the photon pair model, which obtained an improvement of 2.0%.

The ROC-curves for the photon pair is displayed in fig 5.20, while the ROC-curves for the lepton pairs are displayed in fig 5.21. Similarly to the ROC-curves for the photon pairs, the ROC-curves for the lepton pairs are displayed in fig 5.21, where both the electron pair and muon pair has achieved an improvement compared to the ATLAS selection.

Table 5.19: TPR and FPR of the pair models, and the difference between the values obtained with the models and the ones obtained by ATLAS.

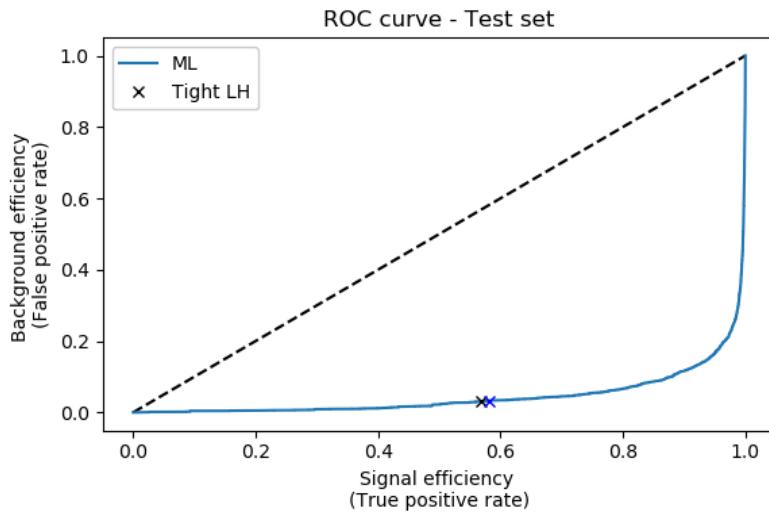
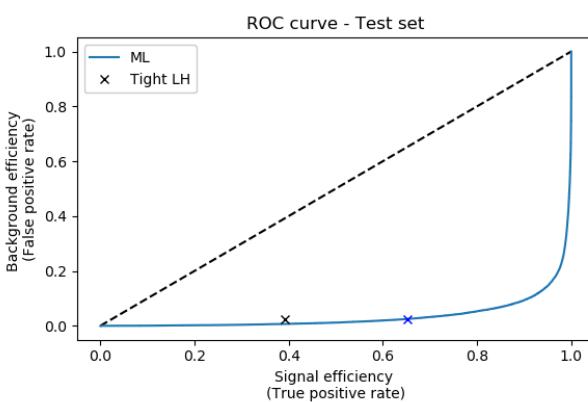
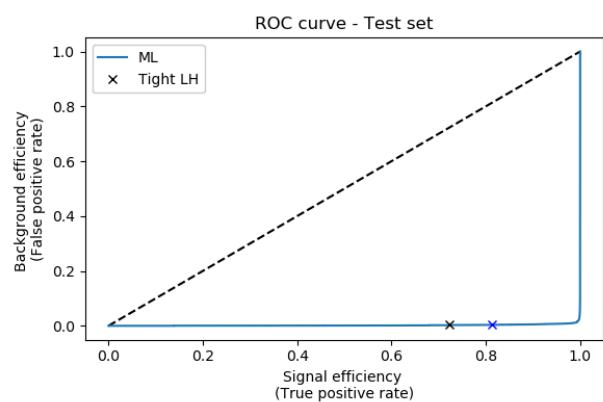


Figure 5.20: Visual presentation of where the ATLAS FPR and TPR values are positioned compared to the ROC-curve generated by the prediction of photon pairs.



(a) ROC-curve and the ATLAS working point for the electron pair.



(b) ROC-curve and the ATLAS working point for the muon pair.

Figure 5.21: ROC-curves for the electron and muon pairs, including the ATLAS working points.

Lastly there is the Area under the curve for the three models, where they all achieved good results of 0.95 or larger, with the muon model doing exceptionally good. Even though this model is this close to being perfect, it still seems like no issue, since there was a lot of background, even after removing pairs with too low or high \mathcal{M} .

5.3 $ll\gamma$ model and data selection

Last of the interactions of interest is the interaction $ll\gamma$. The method used here follows the methods used for the lepton pair in the previous section, with the change that a photon is also required for the selection.

A particle candidate is taken into account, if it has a truthOrigin value for its respective particle type. Both the tag and probe lepton is required to have the correct origin (a Z boson) and PdgId value, the tag lepton is further more required to be a trigger. All photon candidates with more than 4.5GeV of transverse energy are passed forward to the selection process.

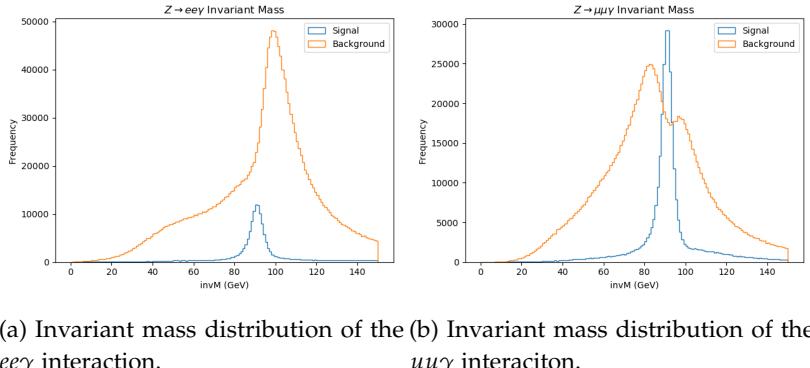
A given event is required to have at least one tag lepton, more than one lepton and at least one photon - there is normally an abundance of photon candidates. If the two leptons in the event have the same charge, they are categorized as background. The photon is required to have a PdgId value of 22 and a truthorigin of 3⁸.

The files used for $ee\gamma$ are the $ee\gamma$ and electron pair files, while only the $\mu\mu\gamma$ files are used for $\mu\mu\gamma$. When testing the selection on $\mu\mu$ files, it was found that there were no muon pairs with random a photon which met the requirements in these files, while an electron pair with a random photon with the correct requirements could be found in the electron pair file.

The invariant mass distributions for both dataset can be seen in fig

Model	Unweighted	Weighted
ee	0.981349	0.964072
$\mu\mu$	0.998631	0.997845
$\gamma\gamma$	0.9665497	0.954505

Table 5.20: The Area under the curve, with and without weights for the pair models.



⁸ This is the value of a single photon. This origin is used due to it being the most fitting of the possibilities[32]

Figure 5.22: The two $ll\gamma$ invariant mass distributions. With signal in blue and background in orange. The huge peak at 100GeV for $ee\gamma$ is due to electron pairs matched with a random photon.

5.22. The two distributions for $ll\gamma$ are similar until the background from the di-electron file is added. The addition of this file gives a lot of background with invariant mass at 100GeV, which is due to two electrons with invariant mass at around 90GeV are combined with a random photon. The majority of the background for these selections

are due to the photons either having the wrong origin - hence not single photons, or having the wrong PdgId⁹.

5.3.1 Reweighting

Similar to the pair models are the range of the reweighting reduced by introducing a lower limit of invariant mass of 50GeV and an upper limit of 150GeV.

Applying these cuts gives the results in tab 5.21, where it is found that 42.6% of the signal is removed when applying the upper limit cut. This seemed weird when compared to what was seen in fig 5.22, so a logarithmic plot was created. In fig 5.23 the range for which the invariant mass is displayed, has increased to to 300GeV. It can here be seen that the signal in each bin stagnates at around 300 per bin¹⁰, while the number of background per bin slowly reduces as the invariant mass increases. The % difference due to cut off can thus be explained by this stagnation in the number of signal per event for invariant masses above 120GeV and the fact that there is almost nine times as much background as there is signal.

The same trend of a larger percentage of the signal being cut when

Z → eeγ sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
All Z → eeγ	250.101	1.0	2.124.039	1.0
M > 50GeV	244.349	0.977	1.989.129	0.936
M < 150GeV	140.155	0.574	1.671.404	0.840
Z → μμγ sel.	Signal #	Signal rel.	Bkg #	Bkg rel.
All Z → μμγ	354.936	1.0	1.297.966	1.0
M > 50GeV	349.105	0.984	1.191.668	0.918
M < 150GeV	299.507	0.858	1.123.453	0.943

applying the maximum invariant mass cut, can be seen for $\mu\mu\gamma$ though to a lesser extend. After applying these cuts the weights are trained, and plotted in fig 5.24.

When selecting what weight to use for $ee\gamma$ it is immediately clear that reverse weights are doing worse than the regular weights. This is especially clear when comparing the transverse energy distributions. The other distributions also shows that the reverse weight with 10 estimators are generally doing poorly, and missing a lot of the features of the background. This leaves the regular weights where the one with 10 estimators is doing worse than the others. For the further training the regular weights with 40 estimators will be used, with the argument for these over the ones with 100 estimators being Occam's razor.

For the $\mu\mu\gamma$ weights it is again found that the reverse weights are underperforming compared to the regular weights. It is further found that the regular weights with 10 estimators are having a hard time getting all of the features correct for the signal. By using the same argument as the $ee\gamma$ weights, the training of the $\mu\mu\gamma$ models will be performed using the regular weights with 40 estimators.

⁹ These distribution aren't that crucial but are plotted in appendix A.5

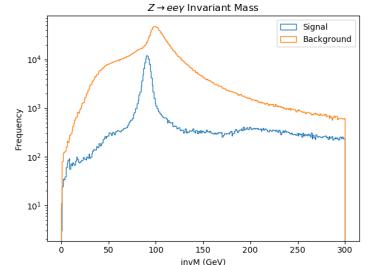
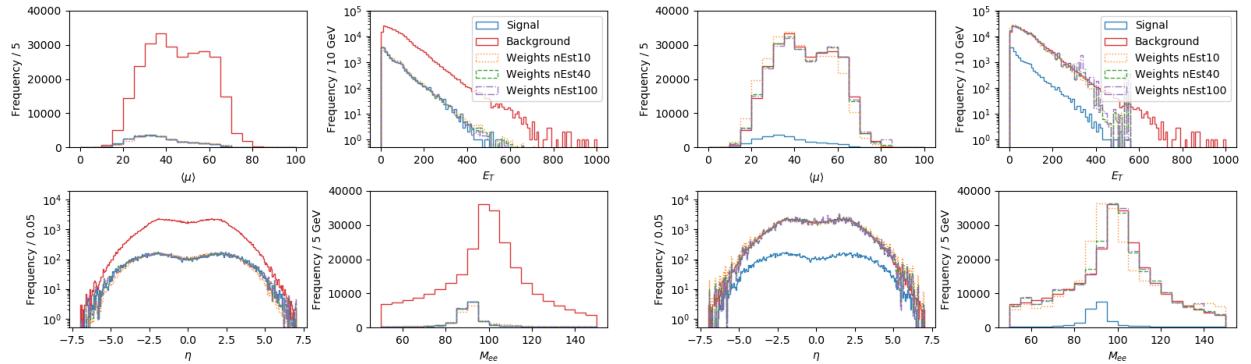


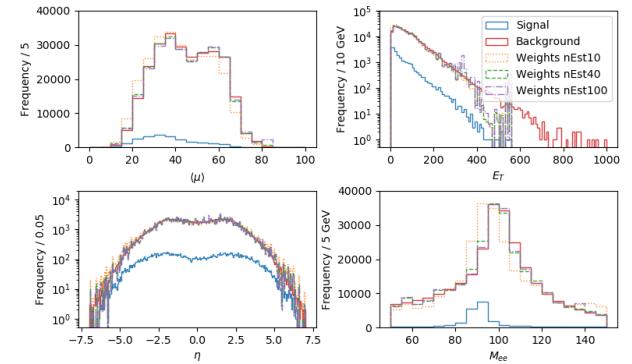
Figure 5.23: Logarithmic plot of the invariant mass of $ee\gamma$ extended to masses to take an extended look at the signal/background distributions.

¹⁰ This does not continue forever, the number of events per bin is merely falling slower than that of the background.

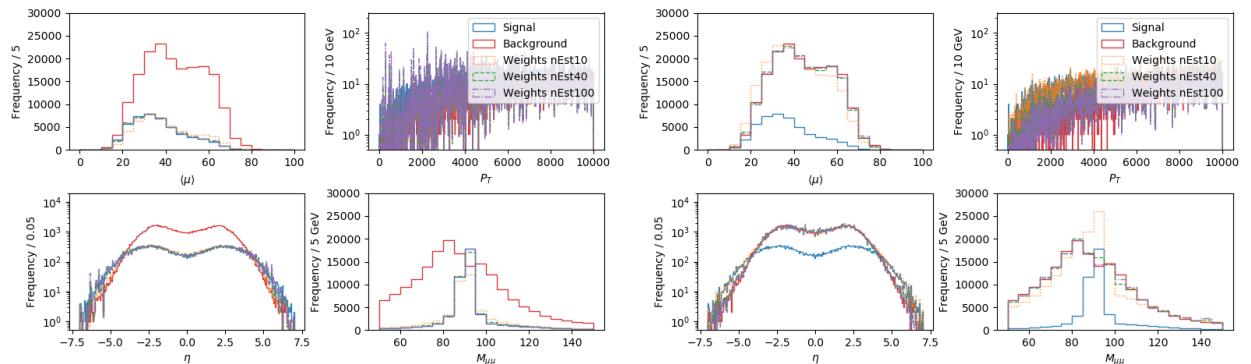
Table 5.21: The number of signal and background events in the $Z \rightarrow ee\gamma$ and $Z \rightarrow \mu\mu\gamma$ data sets, before and after removing the events with an invariant mass less than 50GeV, or higher than 150GeV .



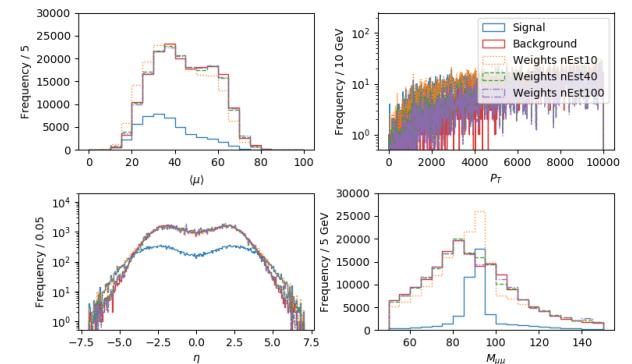
(a) Background reweighted to the signal (regular) for $ee\gamma$, using different number of estimators.



(b) Signal reweighted to the background (reverse) for $ee\gamma$, using different number of estimators.



(c) Background reweighted to the signal (regular) for $\mu\mu\gamma$, using different number of estimators.



(d) Signal reweighted to the background (reverse) for $\mu\mu\gamma$, using different number of estimators.

Figure 5.24: Plots of the signal and background distributions for the isolation selections, and their reweighted values in $\langle\mu\rangle$, E_T/p_T and μ . Larger plots of this can be seen in the appendix in section A.6.

5.3.2 Training

Now having reweighed the $ll\gamma$ data, trained the single particle models and the pair lepton models, the $ll\gamma$ models can be trained. The variables used for these models are a combination of the results from the previous models, since most variables are already handled in those. The list of variables for the $ee\gamma$ model is shown in tab 5.22, while the list for the $\mu\mu\gamma$ model is shown in tab 5.23.

Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
	correctedScaledAveragemu	Pileup
	ee_score	The result from the electron pair model
	pho_pIso_score	The result from the photon isolation model
	pho_pIso_score	The result from the photon identification model

Table 5.22: Variables used for the $ee\gamma$ model.

Name	Feature	Description
	NvtxReco	Number of reconstructed vertices
	correctedScaledAveragemu	Pileup
	mm_score	The result from the muon pair model
	pho_pIso_score	The result from the photon isolation model
	pho_pIso_score	The result from the photon identification model

Table 5.23: variables used for the $\mu\mu\gamma$ model.

It was found that the photon models were the most important variables for both of the models, as can be seen fig 5.25. Surprisingly the pair models had little impact on the final models. This might be due to the data selection having more restrictions on the selection of the photon, than of the lepton pair, since all leptons have the same restriction where as the photon vary from signal to background.

To validate the two models the same ATLAS cuts were used, as in the lepton pair models, with the addition of the photon be required to have the same η range as the $\gamma\gamma$ model, and passing the tight working point. Both the $ee\gamma$ and the $\mu\mu\gamma$ models achieved extremely

Model	FPR ATLAS	FPR LGBM	difference	TPR ATLAS	TPR LGBM	difference
$ee\gamma$	0.0253	0.0253	0	0.1366	0.3802	178.3%
$\mu\mu\gamma$	0.0676	0.0676	0	0.2845	0.8228	189.2%

good results, with the $ee\gamma$ model obtaining an improvement of 178.3% and the $\mu\mu\gamma$ model obtaining an improvement of 189.2%, see tab 5.24.

Table 5.24: FPR and TPR for the two $ll\gamma$ models, and the improvements achieved by the LGBM models compared to the ATLAS cuts.

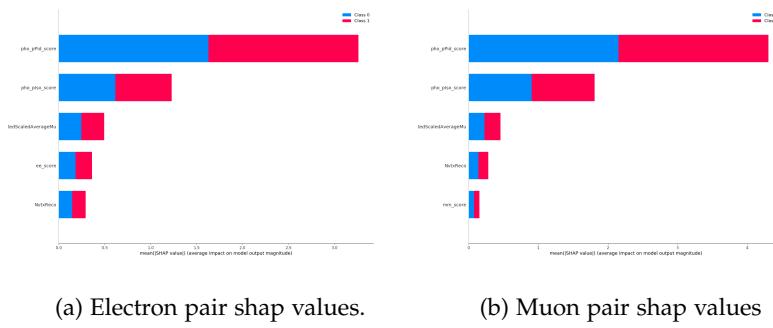
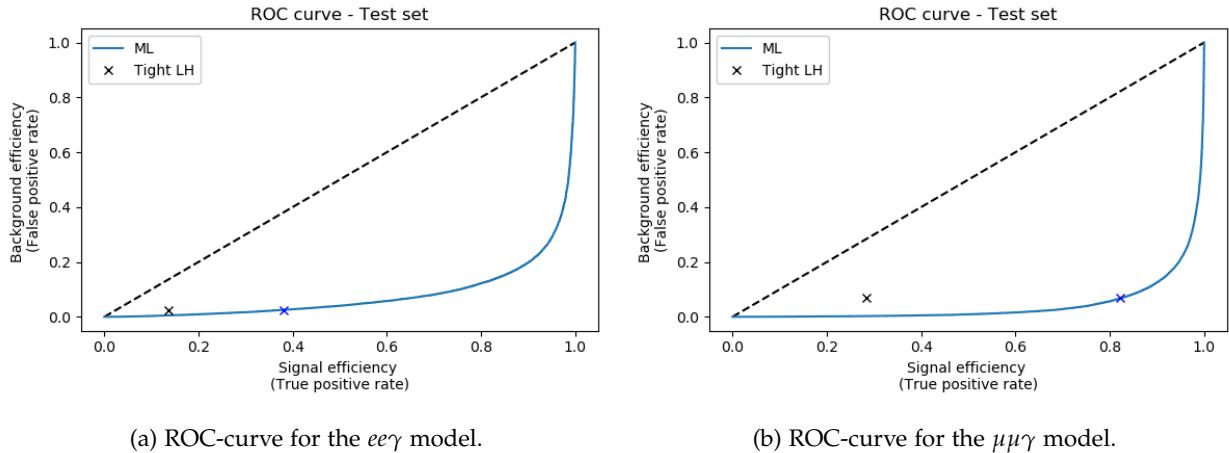


Figure 5.25: SHAP values for the lepton pairs and their corresponding photon using combined variables from lepton pairs and the isolation and identification scores from the single photon models.

The ROC-curve with the ATLAS FPR and TPR are plotted in fig 5.26. It can here be seen that the reason for the improvements are found in good performance from the model and very strict cuts from ATLAS.



Both of the models also achieved good values for their AUC scores, as seen in tab 5.25, with the $ee\gamma$ model getting 0.92 and the $\mu\mu\gamma$ model getting 0.96.

Model	Unweighted AUC score	Weighted AUC score
$ee\gamma$	0.941173	0.920545
$\mu\mu\gamma$	0.957871	0.956012

Figure 5.26: ROC-curve for the two $ll\gamma$ models, $ee\gamma$ left and $\mu\mu\gamma$ right.

Table 5.25: AUC values for the $ee\gamma$ and $\mu\mu\gamma$ models.

6

Reconstruction

Now that the models for the $e\gamma$, $\mu\mu\gamma$ and $\gamma\gamma$ have been trained it is time to compare them to ATLAS. This is done by taking known data - selected using the Truth variables described in the previous chapter. The cuts used by ATLAS will be applied to this data to see how much signal and background passes their cuts. After using these cuts and applying signal efficiencies¹, the models performance can be obtained by requiring the same amount of background in the peak area - with the peak being the invariant mass of the Z- or H-boson $\pm 4\text{GeV}$.

The ATLAS cuts for the two $Z \rightarrow ll\gamma$ interactions are given in ta-

¹ ATLAS removes the top 1% of their $p_{\text{varcone}}^{0.2}$.

ATLAS selections

Electrons (Muons)	$Q_1 \cdot Q_2 < 0$	The two leptons are required to have opposite charge
Trigger		The primary lepton is a trigger lepton
Working point		It is required that both electrons (muons) have passed the Loose (Medium) working point.
$e_T(p_T)$		Both electrons (muons) need at least 10GeV
$M_{ll} < 82\text{GeV}$		$e_T(p_T)$
		The invariant mass of the two leptons has to be lower than 82GeV
$ \eta < 2.47(2.7)$, excluding $1.37 < \eta < 1.52$		Pseudorapidity should be lower than $2.47(2.7)$ for the electrons (muons). The crack is excluded.
$\frac{d_0}{\sigma_{d_0}} < 5(3)$		Significance of d_0 should be smaller than 5 (3) for electrons (muons)
$ \Delta z_0 \sin(\theta) < 0.5\text{mm}$		The leptons in question should be associated with the primary vertex of the interaction.
Track isolation		A cut requiring 99% efficiency in the lepton p_T range, thus removing the top 1% of $p_{\text{varcone}20}$.
Photons	$e_T > 10\text{GeV}$	The transverse energy of the photons needs to be larger than 10GeV
	$ \eta < 2.37$, excluding $1.37 < \eta < 1.52$	Photons in the forward barrel and the crack are excluded
	WP tight	Photons have to be identified with the tight working point

Table 6.1: Selections from [1] used to evaluate ATLAS.

ble (6.1). It is required that the event in question has at least one photon and two opposite charged leptons of same flavour (so either two electrons or two muons), one of which has to be a trigger lepton. The leptons are required to pass either the loose (electron) or medium (muon) working point. It is required that none of the particles have a pseudorapidity in the range $1.37 < |\eta| < 1.52$ and are below (2.37, 2.47, 2.7) depending on the particle (photon, electron, muon).

To ensure the same primary vertex for the lepton candidates, it is required that the longitudinal impact parameter, Δz_0 satisfies $|\Delta z_0 \cdot \sin(\theta)| < 0.5\text{mm}$. Furthermore, it is required that the transverse impact parameter, d_0 , satisfies $|d_0|/\sigma_{d_0} < 5(3)$ for electrons (muons) to suppress the leptons originating from heavy flavour decays.

The invariant mass of the lepton pair is required to be below 82GeV, while the transverse energy of the photons has to be above 10GeV. Furthermore, these photons are required to pass the tight working point. It is lastly required that the event passes track isolation to achieve this isolation, the top 1% $p_T^{0.2}$ values are removed[1].

ATLAS selections

ATLAS selections	
Photons	WP tight
e_T	Photons have to be identified with the tight working point
$ \eta < 2.37$, excluding $1.37 < \eta < 1.52$	The primary photons needs $e_T > 35\text{GeV}$ while the secondary photons needs $e_T > 25\text{GeV}$
$p_T/\mathcal{M}_{\gamma\gamma}$	Photons in the forward barrel and the crack are excluded
Track isolation	The leading and subleading photons are required to have a ratio of 0.35 and 0.25 respectively.
$p_T^{0.2}/e_T < 0.05$	A track is considered isolated if its transverse momentum in a $\Delta R = 0.2$ cone is less than 5% of the transverse energy.
Calorimeter isolation	$e_T^{0.2}/e_T < 0.065$
	A photon candidate has calorimeter based isolation if its transverse energy in a $\Delta R = 0.2$ cone is less than 6.5% of the transverse energy.

The other interaction, $H \rightarrow \gamma\gamma$, requires two photons, and no leptons, see tab (6.2). It starts off requiring that both photon candidates pass the tight working point. In the case of multiple photon candidates, only the two highest e_T candidates will be taken into account - all other candidates are discarded - if included these would result in a lot of low $\mathcal{M}_{\gamma\gamma}$ background. The photons have the same restriction of their η values as the one in the $Z \rightarrow ll\gamma$ interactions. Unlike the $Z \rightarrow ll\gamma$, the photons are required to pass several cuts in transverse energy and momentum. The primary photon is required to have at least 35GeV of transverse energy, while the secondary photon is required to have at least 25GeV of transverse energy. This is

Table 6.2: Selections from [40] used to evaluate ATLAS.

combined with the ratio between the transverse momentum and the invariant mass of the two combined photons $p_T/\mathcal{M}_{\gamma\gamma} > 0.35(0.25)$ for the primary (secondary) photon. Lastly there is calorimeter- and track-based isolation, which is achieved by requiring $e_T^{0.2}/e_T < 0.065$ and $p_T^{0.2}/e_T < 0.05$ respectively[40]. It was discovered too late in the process that there were no information on $e_T^{0.2}$, so a long shot of using topoetcone20 instead of etcone20 for calorimeter isolation was made². However, this was in vain, since only two signal events and no background events would pass the ATLAS cuts on the MC data in question, therefore, this requirement has been omitted.

² This is a long shot since the etcone is used for calorimeter isolation, while topoetcone is used for Topo-cluster E_T -sum

6.1 $Z \rightarrow ll\gamma$ MC models

In the previous section, the steps to acquiring the ATLAS selections were outlined. When using the models, they are separated into a lepton models (where the selection is based on the lepton pair) and photon models. There is also a combined model, which makes use of both the lepton and photon models. Lastly, there is the method that uses the combined model, so instead of combining the lepton pair model with the photon models, the model for $ll\gamma$ will be used.

For all these cases the files with events selected with truth-variables are loaded into dataframes. The ATLAS lepton cuts are applied to the model which only uses my photon model, and ATLAS photon cuts are applied to the model which only uses the lepton model. After this all the scores from the models was added to the applicable dataframe and a requirement of $\mathcal{M}_{ll} < 82$ and $60\text{GeV} < \mathcal{M} < 140\text{GeV}$ was applied. The background for these models are selected, so there is an equal amount of background when comparing the ATLAS model to the other models - this however, is only true in the range $87\text{GeV} < \mathcal{M} < 95\text{GeV}$.

It should be noted that whenever something is referred to as being removed a cut has been applied to the entire dataset. E.g. removing background using the $ee\gamma$ score: A score is selected, so there is the same amount of background in the peak area, as was found in the ATLAS cuts. This score is then applied to all the data, so both signal and background, inside and outside of the peak area are removed, if it doesn't pass this cut.

LEPTON MODEL WITH ATLAS PHOTON CUTS: This case is quite simple, since it only requires one variable, the ll-score gained with the models from the previous chapter. This results in an equal amount of background for this model as the ATLAS cuts, by selecting the correct cut for the ll-pair score.

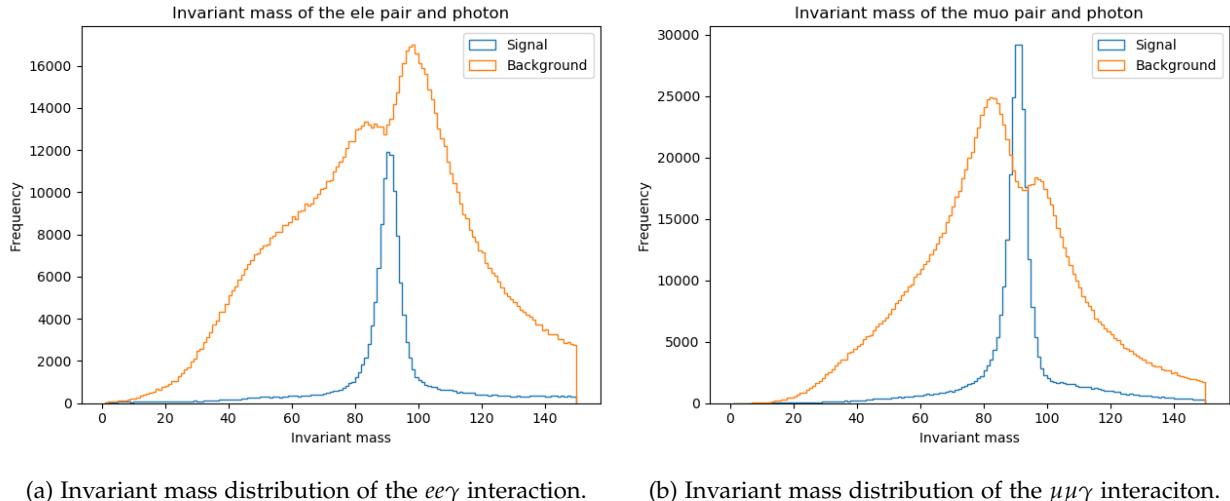
PHOTON MODEL WITH ATLAS LEPTON CUTS: This is slightly more complicated in this case, since the photon models has both photon isolation and identification scores. But it will initially be required to remove 1% of the highest $p_T^{0.2}$ values(for the leptons in question)

for track isolation - it uses the ATLAS lepton cuts. After ensuring track isolation, a 99% signal efficiency in the photon isolation will be required, which removes the lowest 1% photon isolation values, for signal over the entire range. Lastly equal background in the peak energy range will be achieved by setting a fitting photon PID cut.

BOTH LEPTON AND PHOTON MODELS: This is the longest of the models since both the lepton and the photon models are included, resulting in three different variables to tune. Here the lepton pair and photon isolation will be set by signal efficiency. The data will initially be required having a 95% signal efficiency due to the lepton pairs - thus removing the 5% lowest values for the lepton pair scores. After applying this cut is a 99% signal efficiency in photon isolation applied like previously. Like for the photon model, an equal amount of background will be obtained by making cuts in the photon PID scores.

FULL MODEL: This leaves only the full model, where there is only one variable: the $ll\gamma$ score. Here the cut is made by choosing a cut in the $ll\gamma$ score so an equal amount of background is found in the peak area.

The Invariant mass distribution used for these selections are shown



(a) Invariant mass distribution of the $ee\gamma$ interaction.

(b) Invariant mass distribution of the $\mu\mu\gamma$ interaction.

in fig (6.1), which is the same data used for the $ll\gamma$ from the previous chapter - though no background $Z \rightarrow ee$ used in this selection. After applying the electron models and the ATLAS electron cuts to the $Z \rightarrow ee\gamma$ data, the signal which passed these cuts can be seen in fig (6.2).

This can similarly be found for the $Z \rightarrow \mu\mu\gamma$ MC data, which is shown in fig (6.3).

Counting the number of signal events passed by the different models compared to the number of signal events passed by the ATLAS cuts, normalised to the same amount of background. Gives the improve-

Figure 6.1: The two distributions used for the ATLAS cuts and the corresponding selections using the models generated in the previous chapter.

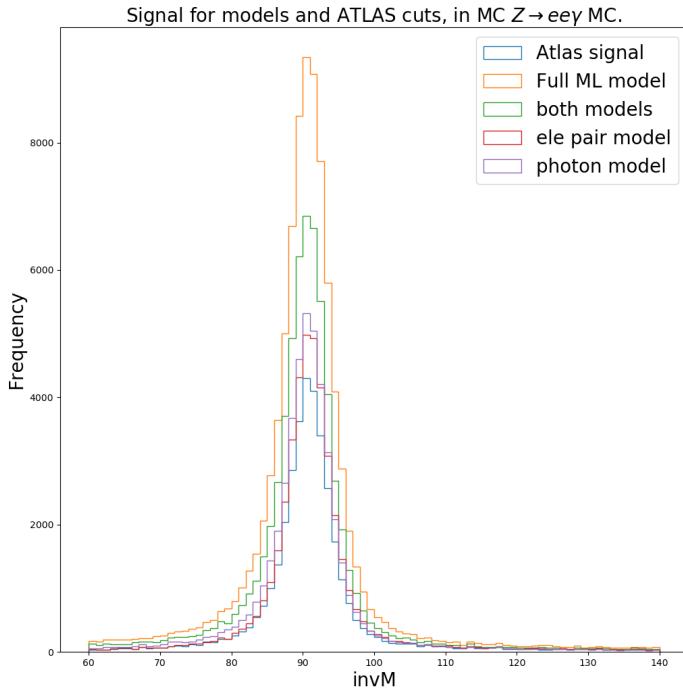


Figure 6.2: The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background for the models and ATLAS.

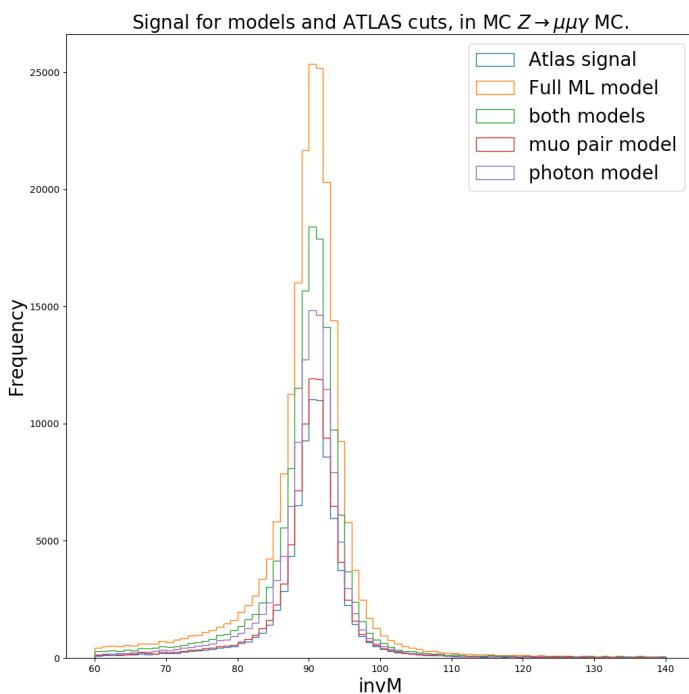


Figure 6.3: The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background for the models and ATLAS.

ments in tab (6.3).

Here it is quickly seen that a model is not a sum of its parts, using photon models with ATLAS cuts for the lepton pair ATLAS gives an improvement of 24.8%(36.1%) for the $ee\gamma(\mu\mu\gamma)$ MC data. If the selections are reversed so the lepton pair model is used with the ATLAS cuts for the photons an improvement of 19.1%(8.8%) is achieved. Using both the lepton pair model and the lepton models gives an improvement of 64.9%(68.0%), while straight up using the $ll\gamma$ models gave an improvement of 128.1%(137.3%).

These are some huge differences in performance, if these are compared to the results achieved in the models, it was there found that the photon PID had an improvement of 40%, the ee ($\mu\mu$) pair had an improvement of 66.8%(12.6%) and the $ee\gamma(\mu\mu\gamma)$ had an improvement of 178.3%(189.2%). Not all of the results from the model training are comparable to the results found here. For both of these it was found that the full models would perform the best, which also holds true for the combination of both the lepton pair model and the photon model. The improvement using only the lepton pair model is comparable to the improvement of the muon pair model, but not in improvement with the electron pair model. .

6.2 $H \rightarrow \gamma\gamma$ MC models

There are similarly to the $Z \rightarrow ll\gamma$ cases some different methods to evaluate the models and their improvements compared to ATLAS. For this thesis three methods has been taken into account. The first of these methods is the full machine learning model, where the $\gamma\gamma$ model is used. The next two are somewhat similar, where they both use the isolation and identification models for the single particles. The cuts selected for the isolation scores are the same (99% signal efficiency), but where they differ is the first method chooses two different values for the identification scores, and the second method chooses the same score.

FULL ML MODEL: The idea of this method is very similar to what was described in the previous section for the full ML method, but now the peak is found at $M_H \pm 4\text{GeV}$. So a cut is selected so the same amount of background is found in the peak energy range.

For the next two models a 99% signal efficiency was applied to both of the photon isolation scores - similar to what was done in the previous section.

DIFFERENT PID SCORES: When selecting two different PID scores some sort of search will be required to find the optimal value. Due to lack of time for a better method, a random search is performed. A

Model	Electrons	Muons
Di-lepton	19.1%	8.8%
Photon	24.8%	36.1%
Both	64.9%	68.0%
Full	128.1%	137.3%

Table 6.3: Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.

grid of the PID scores in the range of $[-10, 10]$ was created for the random search. From this grid, two PID scores can be combined, 500 of these combinations are made each run, if any of these combinations have the same background in the peak area as the ATLAS selection, these values are saved.

This generation of PID pairs continue until either 5000 candidate pairs have been found or 200 runs have been passed.³. After this the candidate with the highest amount of signal is selected, and the scores for this candidate is saved and used.

SAME PID SCORE: This method is a bit simpler, than what was done with the different PID scores. A simple search in the range of the PID scores was done - since this search is only in one dimension. The candidates with the same background as ATLAS are saved. Of these candidates the one with the highest amount of signal is chosen, and the PID of this candidate is saved and used as a cut.

When comparing these three methods with the results from using the

Model	Improvement
di-photon	21.7%
Random search	10.5%
Same cut	9.6%

Table 6.4: Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.

³ This number of runs and pairs of PID scores are done, to counteract the risk of random fluctuation due to missing the minima.

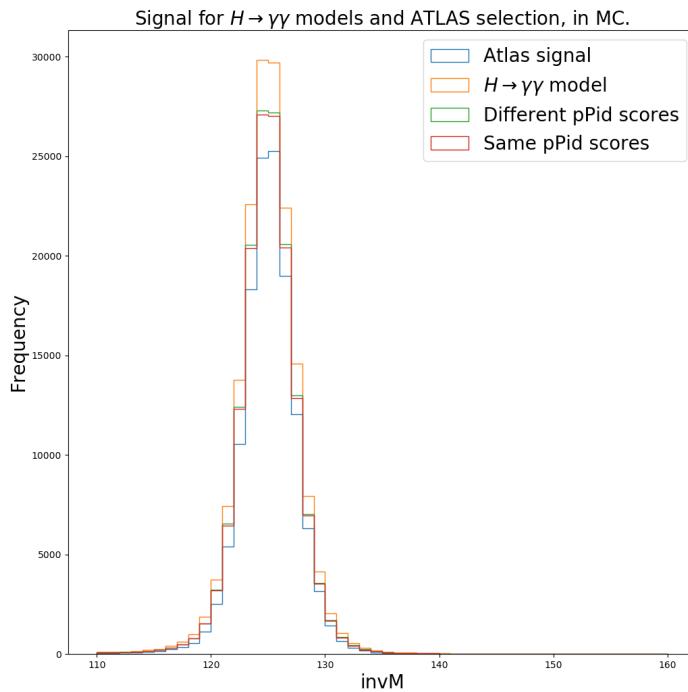


Figure 6.4: The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background in the energy range 121 – 129GeV for the models and ATLAS.

ATLAS cuts, see tab 6.4, it was found that the worst performing was the method where the same cut was selected with an improvement of 9.6%, where a random search for two different values obtained a slightly better result of an improvement of 10.5%. The method which resulted in the best improvement was the one using the photon pair model, which obtained an improvement of 21.7%.

Given additional time, these results could have been improved. Using a finer search for the same PID method, would allow for more chances of better signal efficiency. This would likewise also be the case for different PID scores, since additional guesses of PID scores would result in more score candidates. One of these candidates could have had a better signal efficiency, which was missed. Using another method to search for the candidates could also have proven to be better than the ones used here.

A change in the MC data, where there would have been more background in the peak area, would probably have been an improvement, both for training the models and for checking their performance afterwards. Background from jet+jet could have been used, but this interaction is rather time consuming, since a lot of simulations would be required due to a lot of the events being discarded in the selection process. Interactions which would require less simulation would be a much better alternative, such as gamma+jet and $X \rightarrow \gamma\gamma$ with X being a particle which isn't the Higgs boson. The gamma+jet would be a good alternative, since it requires less simulations due to less events being discarded in the selection process.

6.3 $Z \rightarrow \mu\mu\gamma$ Data models

Lastly there is the $Z \rightarrow \mu\mu\gamma$ for real data. This differs from the others since it is unknown whether the particles in the event are signal or background⁴.

For this reason it will be needed to find another method to make sure, that there is close to an equal amount of background in the ATLAS selection and in the selection using the ML models. But where the background was previously required to be the same in the peak region, this will not be feasible when using data.

A reasonable assumption would be to assume that events with an invariant mass in the energy range of $60\text{GeV} < M < 80\text{GeV}$ or $100\text{GeV} < M < 140\text{GeV}$ are all background. This assumption was made since this should be far enough from the expected peak at 91GeV . The models used here will therefore be the exact same as the ones used in section 6.1, with the only difference being how background is determined and in what energy region they are set to be equal.

By making use of these methods, a comparison to the amount of events in the peak area for the ATLAS cuts could be made, see tab 6.5. It was here found that all the methods did very comparable or slightly worse than the ATLAS cuts. The method which did the best was the one where the cut was based on the $\mu\mu\gamma$ ML value, with an improvement of 2.5%. The second best was the model using the score for μ -pairs, which did 0.7% better than what the ATLAS cuts managed. The model which made use of the photon model performed 0.5% worse than ATLAS. Lastly there was the method

⁴ An initial guess of the same scores, as the ones found in MC was tried. This lead to poor results, since this would result in too high background.

Model	Improvement
Di-muon	0.7%
Photon	-0.5%
Both	-5.0%
Full	2.5%

Table 6.5: Percentage difference between the amount of signal there is in a given model, and the signal in the ATLAS cuts - in the peak area, given the same amount of background in the peak area.

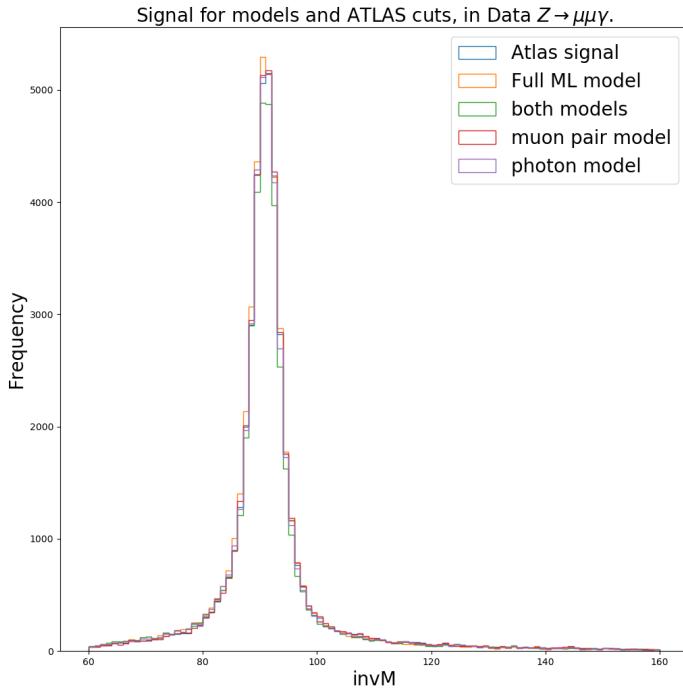


Figure 6.5: The signal after applying ATLAS cuts and requiring signal efficiencies and same amount of background in the energy range $60 - 80\text{GeV}$ and $100 - 140\text{GeV}$ for the models and ATLAS.

where both the muon pair and the photon was taken into account, where it performed 5.0% worse than the ATLAS cuts.

Using that events in the energy range $60\text{GeV} < \mathcal{M} < 80\text{GeV}$ or $100\text{GeV} < \mathcal{M} < 140\text{GeV}$ are all background is most likely a good assumption. This does not require that the amount of background in the peak region is the same, one may assume that they are somewhat equal, but we can not be completely sure.

7

Conclusion

This thesis investigated the $H \rightarrow \gamma\gamma$ and the $H \rightarrow Z(\rightarrow ee/\mu\mu)\gamma$ decays, and made an optimized machine learning models for these decays.

To do this it was originally required to make single particle models: an isolation and identification model for each of the three particles: e, μ and γ . These models were trained using MC data sets.

The isolation models were compared to the isolation WP used by ATLAS. Where the electron isolation model achieved an improvement of 21.6% compared to the workingpoint. The isolation model for the muon managed to obtain an improvement of 15.3%. The last of the isolation models was the one for the photons, which was compared to two different isolation WP, a Loose and a Tight WP, where the model achieved an improvement of 17.5% (39.9%) for the loose (tight) working point.

Besides the isolation models identification models for the three particles were trained. Here the leptonic identification models were compared to the three working points: loose, medium and tight. The electronic identification model managed an improvement of 19.0%, 17.9% and 12.5% compared respectively to the ATLAS loose, medium and tight working points. The muon identification model managed slightly worse improvements compared to the ATLAS working points, with an improvement of 3.6%, 1.1%, 9.1% compared to the loose, medium and tight working points. Last of the identification models were the one for the photon identification model, which was compared to the loose and tight identification WP. Here the photon identification model had an improvement compared to the ATLAS working points of respectively 16.8% and 40.0% compared to the ATLAS loose and tight working points.

After having trained the six single particle models, the di-lepton and di-photon models were trained. The $Z \rightarrow ee$ model was compared against the cuts made by ATLAS, and resulted in an improvement of 66.8%, while the $Z \rightarrow \mu\mu$ obtained a result with an improvement of 12.6%. The di-photon selection was evaluated against the $H \rightarrow \gamma\gamma$ selection and obtained an improvement of 2.0%.

This leaves only two models, the two $Z \rightarrow (ee/\mu\mu)\gamma$ models. When evaluated against the cuts used by ATLAS, the model using electrons

had an improvement of 178.3%, while the model using muons obtained an improvement of 189.2%.

After having trained all of the models, their performance in selection of signal and background were compared to the cuts used by ATLAS. When evaluating $Z \rightarrow ll\gamma$ compared to the cuts made by ATLAS[1] with l being an electron (muon). Using the ATLAS photon cuts and the di-electron(muon) model score to select the events, an improvement of 19.1%(8.8%) was obtained. If the ATLAS cuts for the di-leptons, and the photon model scores to select the events an improvement of 24.8%(36.1%) was obtained. Using a combination of the di-lepton model score and the photon models scores gave an improvement of 64.9%(68.0%) compared to the ATLAS cuts. Lastly, there was the method which used the full machine learning model where the score for $ll\gamma$ was used. An improvement of 128.1%(137.3%) was obtained when comparing the signal efficiency with the signal efficiency obtained by using the ATLAS cuts given the same background efficiency.

When evaluating the $H \rightarrow \gamma\gamma$ in MC data three methods were taken into account, one where the di-photon model was used, and two methods where only the single photon scores was used: One where the PID scores were different and one where they were the same. When using the di-photon model an increase in signal efficiency of 21.7% was obtained. The method where two different PID cuts were used got an improvement of 10.5%, while using the same PID cuts obtained an improvement of 9.6%

Last but not least, was the $Z \rightarrow \mu\mu\gamma$ using real data. Using the MC models on the real data sets gave results which ranged from equally as good to slightly worse than what ATLAS had using their cuts. Using the ATLAS photon cuts and the di-muon model gave an improvement of 0.7%. If the ATLAS muon cuts were used with the photon model, the signal efficiency was 0.5% worse, than using only the ATLAS cuts. This leaves the method where both the di-muon and photon models were used and the one where the combined score of the $\mu\mu\gamma$ event is evaluated, these two obtained respectively 5.0% worse and 2.5% better results than using the ATLAS cuts.

7.1 Future work

The work in this project can be divided into two parts, the work on the models, and evaluation of how well it selects signal events compared to ATLAS.

An improvement of the model training for the $H \rightarrow \gamma\gamma$ would be to increase the amount of background, especially at energies higher than 100GeV. Background for these events could be found in additional $jet + \gamma$ files, and events from $X \rightarrow \gamma\gamma$ decays. An increase of the background for the single particle models would also be an improvement since the majority of the background for all of the particles have energies below 20GeV.

Regarding the evaluation of the models against the ATLAS cuts, using

real data, it will be required to see if any of the suggestions above gives any improvements.

Besides these improvements, it would be wise to explore other ways to make sure there is the same amount of background. This is crucial for further studies, since there is very little resemblance between the increased signal efficiency in MC, compared to the loss of signal efficiency in Data.

It would also be interesting to have an automatic method to select the weights. A mean absolute error or something similar could be used to determine how well the weights are doing.

Bibliography

- [1] G. Aad et al. "A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physics Letters B* 809 (Oct. 2020), p. 135754. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135754. URL: <http://dx.doi.org/10.1016/j.physletb.2020.135754>.
- [2] *About Train, Validation and Test Sets in Machine Learning*. <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>. Accessed: 2021-08-05.
- [3] *An Introduction to Machine Learning*. <https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning>. Accessed: 2021-08-14.
- [4] *ATLAS coordinate systems*. https://www.researchgate.net/figure/The-ATLAS-detector-and-its-subsystems-The-two-figures-on-the-right-visualize-two_fig82_318981598. 16-05-2021.
- [5] *ATLAS: Detector*. <https://atlas-public.web.cern.ch/discover/detector>. Accessed: 2020-09-08.
- [6] *ATLAS EXPERIMENT: The ATLAS Trigger and Data AcQuisition System*. <https://cds.cern.ch/record/2730760/files/ATL-Daq-SLIDE-2020-356.pdf>. Accessed: 2021-07-28.
- [7] *ATLAS TRT plot*. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TRTPublicResults>. Accessed: 2021-07-25.
- [8] Evelina Bouhova-Thacker and Vakhtang Kartvelishvili. "Electron bremsstrahlung recovery in ATLAS tracking using Dynamic Noise Adjustment". In: (Sept. 2021).
- [9] *Calorimeter*. <https://web.archive.org/web/20071220021724/http://rd11.web.cern.ch/RD11/rkb/PH14pp/node19.html>. Accessed: 2020-09-29.
- [10] *Cambridge University: Parton Distribution Functions (a presentation)*. https://indico.cern.ch/event/94815/contributions/1282690/attachments/1106827/1579062/LBLMIT_Stirling_v2.pdf. Accessed: 2021-06-06.
- [11] *Classification: ROC Curve and AUC*. <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>. Accessed: 2021-09-29.

- [12] *Concept of Luminosity.* <https://cds.cern.ch/record/941318/files/p361.pdf>. Accessed: 2021-07-03.
- [13] *DerivationFramework.* <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/DerivationFramework>. Accessed: 2021-08-26.
- [14] *Detector simulation for LHC analyses recasting - Part 2.* https://indico.cern.ch/event/873524/contributions/3697290/attachments/1988392/3315700/Korea_DetectorSimulation2.pdf. Accessed: 2021-07-07.
- [15] *Efficiency Studies using the Tag and Probe Method.* <https://cms-opendata-workshop.github.io/workshop-lesson-tagandprobe/aio/index.html>. Accessed: 2021-09-04.
- [16] *egamma AOD derivations.* <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaxAODDerivations>. Accessed: 2021-08-26.
- [17] *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data.* Tech. rep. Figures available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2016-024>. Geneva: CERN, June 2016. URL: <https://cds.cern.ch/record/2157687>.
- [18] *EM Showers.* https://en.wikipedia.org/wiki/Particle_shower. Accessed: 2020-10-13.
- [19] *Ensemble Learning Algorithm Complexity and Occam's Razor.* <https://machinelearningmastery.com/ensemble-learning-and-occams-razor/>. Accessed: 2021-08-05.
- [20] *Ensemble Methods in Machine Learning: What are They and Why Use Them?* <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f>. Accessed: 2021-08-14.
- [21] G. Aad et.al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Jinst* 3 (2008).
- [22] *Gradient Descent For Machine Learning.* <https://machinelearningmastery.com/gradient-descent-for-machine-learning/>. Accessed: 2021-08-10.
- [23] *Handout 13 : Electroweak Unification and the W and Z Bosons.* https://www.hep.phy.cam.ac.uk/~thomson/lectures/partIIIparticles/Handout13_2009.pdf. Accessed: 2021-08-16.
- [24] Peter Hansen. *Particle detectors and accelerators, Lecture notes.* Third Edition. University of Copenhagen, 2016.
- [25] Benjamin K. Henckel. *Cleaning and Training.*
- [26] *HGamma DxAOD samples.* https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HggDerivationSamples?fbclid=IwAR3KU_uAB2areIKw0fBXiTc6YnXgB5fg0HZT8HdZAZBqHcM70cJACyNaleo. Accessed: 2021-08-28.

- [27] *Higgs couplings in ATLAS at Run2.* <https://arxiv.org/pdf/1907.06297.pdf>. Accessed: 2021-07-15.
- [28] *Higgs search at ATLAS.* <https://cds.cern.ch/record/1546765/files/ATL-PHYS-PROC-2013-114.pdf>. Accessed: 2021-07-15.
- [29] *How CMS weeds out particles that pile up.* <https://cms.cern/news/how-cms-weeds-out-particles-pile>. Accessed: 2021-07-07.
- [30] *Hyperparameters Optimization.* <https://towardsdatascience.com/hyperparameters-optimization-526348bb8e2d>. Accessed: 2021-08-15.
- [31] *Inelastic proton cross section at 13 TeV with ATLAS.* <https://inspirehep.net/files/3ab61a1208a0ef83f49e45e4815b1a36>. Accessed: 2021-08-17.
- [32] Helle K. Leerberg. *Higgs Decay to Z γ A Study of Efficiency.*
- [33] *LightGBM features.* <https://lightgbm.readthedocs.io/en/latest/Features.html>. Accessed: 2021-08-12.
- [34] *Machine learning.* https://en.wikipedia.org/wiki/Machine_learning. Accessed: 2021-08-14.
- [35] *Machine Learning can amplify bias.* <https://www.kdnuggets.com/2019/08/types-bias-machine-learning.html>. Accessed: 2021-08-05.
- [36] B.R. Martin. *Nuclear and Particle Physics: An Introduction.* ISBN: 978-0-470-74274-7.
- [37] *MC truth Classifier definitions.* <https://gitlab.cern.ch/atлас/athena/blob/21.2/PhysicsAnalysis/MCTruthClassifier/MCTruthClassifier/MCTruthClassifierDefs.h>. Accessed: 2021-09-06.
- [38] *MCPDxAOD.* <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPDxAOD>. Accessed: 2021-08-26.
- [39] *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data.* https://www.researchgate.net/figure/Sketch-of-the-lateral-and-longitudinal-segmentation-of-the-ATLAS-electromagnetic_fig1_312094570. Accessed: 2021-07-29.
- [40] *Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \rightarrow \gamma\gamma$ channel using 139 fb $^{-1}$ of $p\ p$ collision data with the ATLAS experiment.* Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-026/>. Geneva: CERN, Aug. 2020. URL: [http://cdsweb.cern.ch/record/2725727/files/ATLAS-CONF-2020-026.pdf](https://cdsweb.cern.ch/record/2725727/files/ATLAS-CONF-2020-026.pdf).
- [41] *Monte Carlo Particle Numbering Scheme.* <https://pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf>. Accessed: 2021-09-06.

- [42] *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s}=13$ TeV.* <https://arxiv.org/abs/2012.00578>. Accessed : 2021-09-24.
- [43] *Overfitting and Underfitting With Machine Learning Algorithms.* <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>. Accessed: 2021-08-08.
- [44] *Overhead (computing).* [https://en.wikipedia.org/wiki/Overhead_\(computing\)](https://en.wikipedia.org/wiki/Overhead_(computing)). Accessed: 2021-11-11.
- [45] Sara E. D. Pinholt. *Improving Z and Higgs Selections in ATLAS using Machine Learning.*
- [46] *Quantum diaries: How a pixel detector works.* <https://www.quantumdiaries.org/2008/07/25/how-a-pixel-detector-works/>. Accessed: 2020-10-14.
- [47] *Quantum Diaries: Piling up.* <https://www.quantumdiaries.org/2011/10/25/piling-up/>. Accessed: 2021-07-07.
- [48] *Run2 Isolation Harmonisation.* <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/%20Run2IsolationHarmonisation>. Accessed : 2021-09-15.
- [49] *The ATLAS Data Acquisition System in LHC Run 2.* <https://inspirehep.net/files/9a5431c196cbe9df0b956a7c01420d23>. Accessed: 2021-08-19.
- [50] *THE ATLAS TRANSITION RADIATION TRACKER.* <https://arxiv.org/pdf/hep-ex/0311058.pdf>. Accessed: 2021-07-25.
- [51] *The Higgs boson.* <https://home.cern/science/physics/higgs-boson>. Accessed: 2021-07-15.
- [52] *The origins of the Brout-Englert-Higgs mechanism.* <https://home.cern/science/physics/origins-brout-englert-higgs-mechanism>. Accessed: 2021-08-16.
- [53] *The Standard Model.* <https://physics.info/standard/>. Accessed: 2020-10-30.
- [54] Mark Thomson. *Modern Particle Physics.* Cambridge University Press, 2019. ISBN: 978-1-107-03426-6.
- [55] *Tree Based Algorithms: A Complete Tutorial from Scratch (in R and Python).* <https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/>. Accessed: 2021-08-15.
- [56] *Understanding Loss Functions in Machine Learning.* <https://www.section.io/engineering-education/understanding-loss-functions-in-machine-learning/>. Accessed: 2021-08-10.
- [57] *W and Z bosons.* https://en.wikipedia.org/wiki/W_and_Z_bosons. Accessed: 2021-07-13.
- [58] *Welcome to the SHAP documentation.* <https://shap.readthedocs.io/en/latest/index.html>. Accessed : 2021-09-28.

- [59] *What is LightGBM, How to implement it? How to fine tune the parameters?* <https://medium.com/@pushkarmandot/https-medium-com-pushkarmandot-what-is-lightgbm-how-to-implement-it-how-to-fine-tune-the-parameters-60347819b7fc>. Accessed: 2021-08-12.
- [60] *What Is Machine Learning?* <https://se.mathworks.com/discovery/machine-learning.html>. Accessed: 2021-08-09.
- [61] *Wikipedia: ATLAS experiment.* https://en.wikipedia.org/wiki/ATLAS_experiment. Accessed: 2021-07-21.
- [62] *Wikipedia: Large Hadron Collider.* https://en.wikipedia.org/wiki/Large_Hadron_Collider. Accessed: 2021-07-21.
- [63] *Wikipedia: Transition Radiation.* https://en.wikipedia.org/wiki/Transition_radiation. Accessed: 2021-07-25.
- [64] *Wikipedia: Mandelstam variables.* https://en.wikipedia.org/wiki/Mandelstam_variables. Accessed: 2021-06-28.
- [65] *Wikipedia: Pseudorapidity.* <https://en.wikipedia.org/wiki/Pseudorapidity>. Accessed: 2021-05-16.
- [66] *Z-Path.* <https://atlas.physicsmasterclasses.org/en/zpath-lhcphysics2.htm>. Accessed: 2021-07-13.

A

Appendix

A.1 Datasets

A.2 Invariant mass and transverse energy/momenta plots

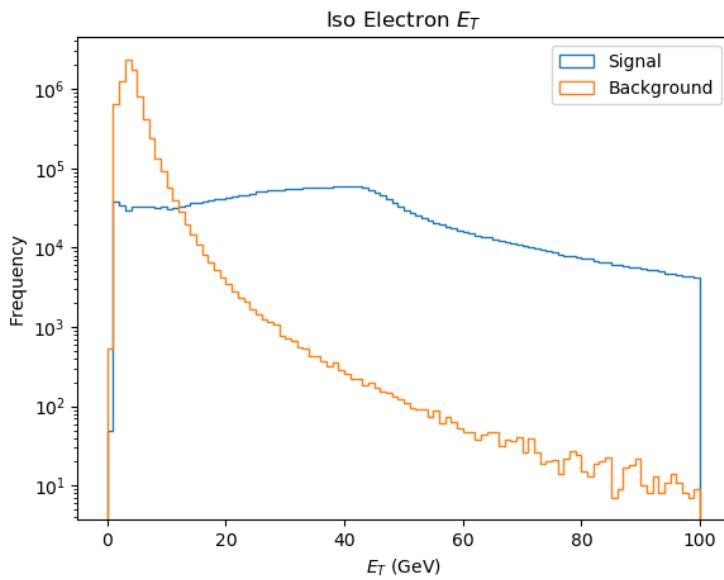


Figure A.1: The transverse energy for the electrons chosen with the Pid method. The signal is displayed in blue, while the background is displayed in orange.

A.3 Reweighted single particle plots

A.4 Reweighted particle pair weights

A.5 Background types for $Z \rightarrow ll\gamma$

The following plots are the distribution of the signal and background, with the background divided between bad photons and leptons - or both. It is bad leptons if the leptons have the same charge, but the photon has the correct pdgId and origin. It is a bad photon if the leptons have opposite charge and the photon has the correct origin and pdgId.

Process	Container	Model
Z \rightarrow ee γ	mc16_13TeV.301535.Sherpa_CT10_eegammaPt10_35.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956 mc16_13TeV.301899.Sherpa_CT10_eegammaPt35_70.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956 mc16_13TeV.301900.Sherpa_CT10_eegammaPt70_140.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956 mc16_13TeV.301901.Sherpa_CT10_eegammaPt140.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956	eIso, ePid, Zee γ , γ Iso, γ Pid, Zee
Z \rightarrow ee	mc16_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.deriv. DAOD_EGAM3.e3601_e5984_s3126_r10201_r10210_p3954	eIso, ePid, Zee, Zee γ
H $\rightarrow \gamma\gamma$	mc16_13TeV.343981.PowhegPythia8EvtGen_NNLOPS_nnlo_30_ggH125_gamgam.deriv.DAOD_HIGG1D1.e5607_e5984_s3126_r10724_r10726_p4180 mc16_13TeV.346214.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_gamgam.deriv.DAOD_HIGG1D1.e6970_e5984_s3126_r10724_r10726_p4180	γ Iso, γ Pid, $\gamma\gamma$
H $\rightarrow \gamma^*\gamma$	mc16_13TeV.345834.PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_gamstargam.deriv.DAOD_HIGG1D2.e7111_e5984_s3126_r10724_r10726_p4114 mc16_13TeV.345961.PowhegPythia8EvtGen_NNLOPS_nnlo_30_ggH125_gamstargam.deriv.DAOD_HIGG1D2.e6740_e5984_s3126_r10724_r10726_p4114	γ Iso, γ Pid, $\gamma\gamma$
Z $\rightarrow \mu\mu\gamma$	data16_13TeV.periodAllYear.physics_Main.PhysCont.DAOD_EGAM4.grp16_v01_p3948 mc16_13TeV.301536.Sherpa_CT10_mumugammaPt10_35.deriv. DAOD_EGAM4.e3952_s3126_r10201_r10210_p3956 mc16_13TeV.301903.Sherpa_CT10_mumugammaPt70_140.deriv. DAOD_EGAM4.e3952_s3126_r10201_r10210_p3956	μ Iso, μ Pid, Z $\mu\mu\gamma$, γ Iso, γ Pid, Z $\mu\mu$
Z $\rightarrow \mu\mu$	mc16_13TeV.361107.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zmumu .deriv.DAOD_MUON1.e3601_e5984_s3126_r10724_r10726_p3629	μ Iso, μ Pid, Z $\mu\mu$
H $\rightarrow Z(\text{ll})\gamma$	mc16_13TeV.344303.PowhegPythia8EvtGen_CT10_AZNLOCTEQ6L1_MPI- OFF_ggH125_Zllgam.deriv.DAOD_HIGG1D2.e5145_s3126_r10724_p4062 mc16_13TeV.345833.PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_Zllgam.deriv.DAOD_HIGG1D2.e7111_s3126_r10724_p4062	γ Iso, γ Pid, eIso, ePid, μ Iso, μ Pid, Zee, Z $\mu\mu$
W $^+ \rightarrow e\nu$	mc16_13TeV.361100.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusenu. deriv.DAOD_EGAM1.e3601_e5984_s3126_r10201_r10210_p3916	eIso, ePid, Zee
W $^- \rightarrow e\nu$	mc16_13TeV.361103.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusenu. deriv.DAOD_EGAM1.e3601_e5984_s3126_r10201_r10210_p3916	eIso, ePid, Zee
W $^+ \rightarrow \mu\nu$	mc16_13TeV.361101.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusmunu. deriv.DAOD_MUON1.e3601_e5984_s3126_s3136_r10724_r10726_p4145	μ Iso, μ Pid, Z $\mu\mu$
W $^- \rightarrow \mu\nu$	mc16_13TeV.361104.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusmunu. deriv.DAOD_MUON1.e3601_e5984_s3126_s3136_r10724_r10726_p4145	μ Iso, μ Pid, Z $\mu\mu$
Jets	mc16_13TeV.423302.Pythia8EvtGen_A14NNPDF23LO_perf_JF35.deriv.DAOD_EGAM7.e3848_e5984_s3126_s3136_r10724_r10726_p4613	eIso, ePid, γ Iso, γ Pid

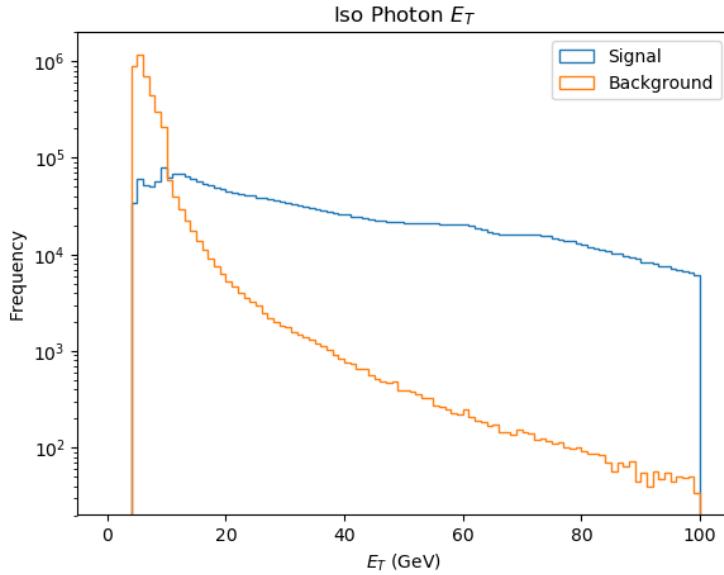


Figure A.2: Transverse energy of isolated photons

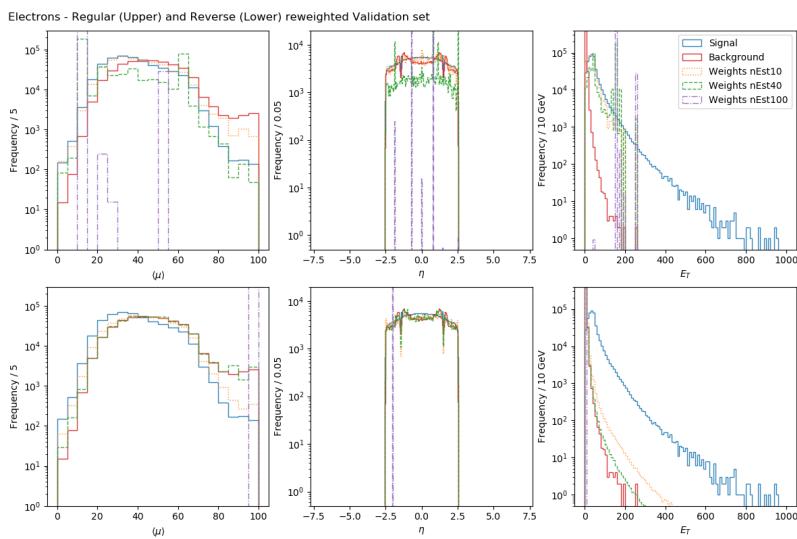


Figure A.3: Signal and background distribution of $\langle \mu \rangle$, e_T and η for isolated electron reweighed using different number of estimators.

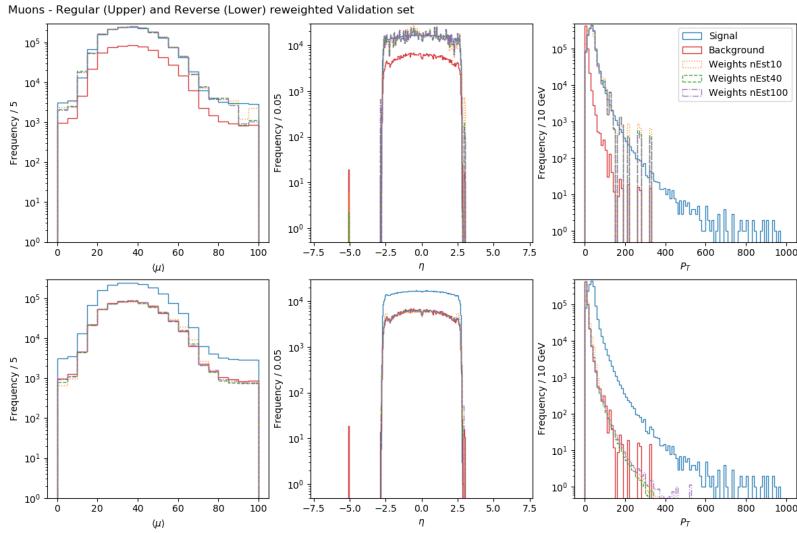


Figure A.4: Signal and background distribution of $\langle \mu \rangle$, p_T and η for isolated muons reweighed using different number of estimators.

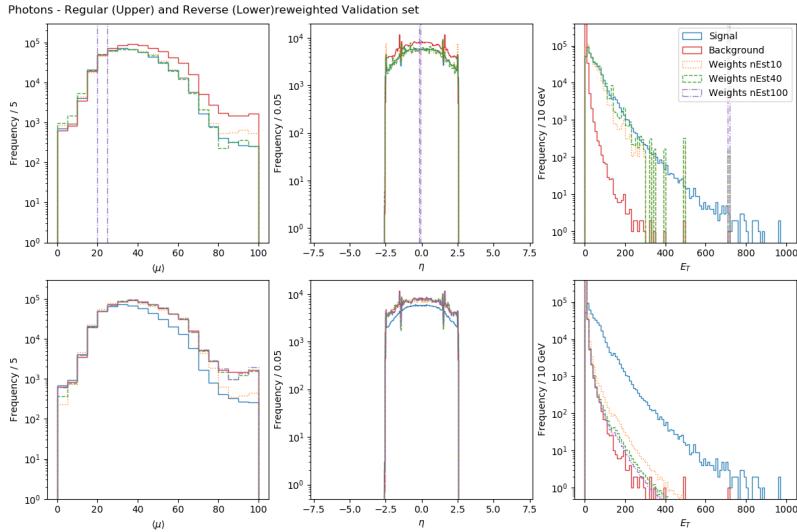


Figure A.5: Signal and background distribution of $\langle \mu \rangle$, e_T and η for isolated photons reweighed using different number of estimators.

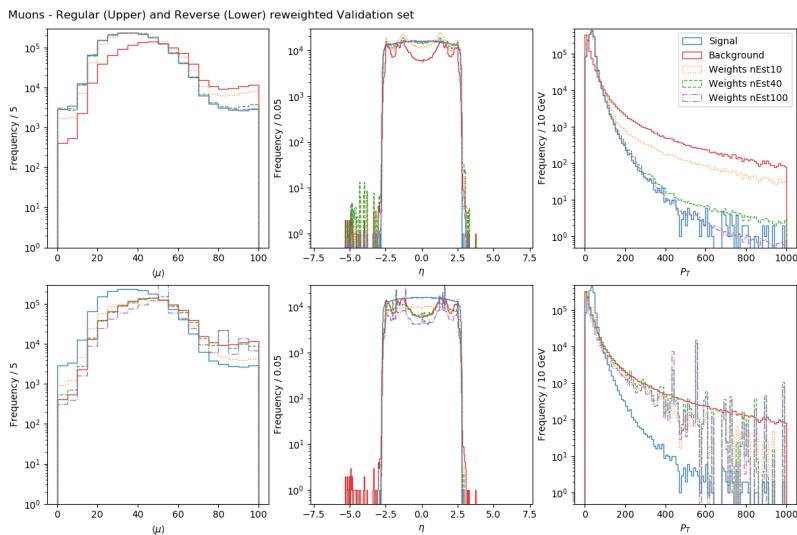


Figure A.6: Signal and background distribution of $\langle \mu \rangle$, p_T and η for identified muons reweighed using different number of estimators.

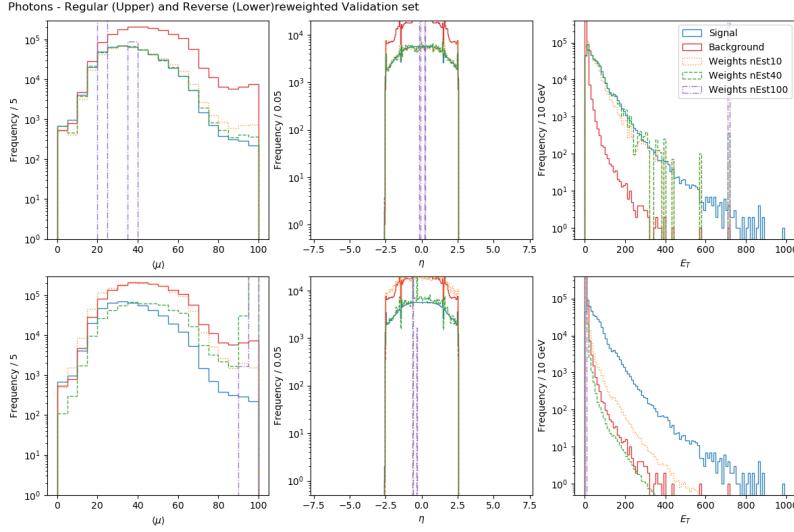


Figure A.7: Signal and background distribution of $\langle \mu \rangle$, e_T and η for identified photons reweighted using different number of estimators.

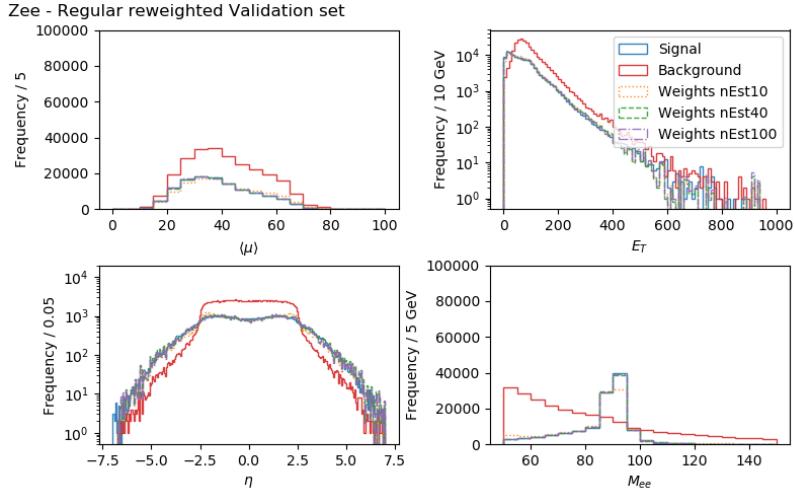


Figure A.8: Background reweighted to the signal for the electron pairs, using different number of estimators

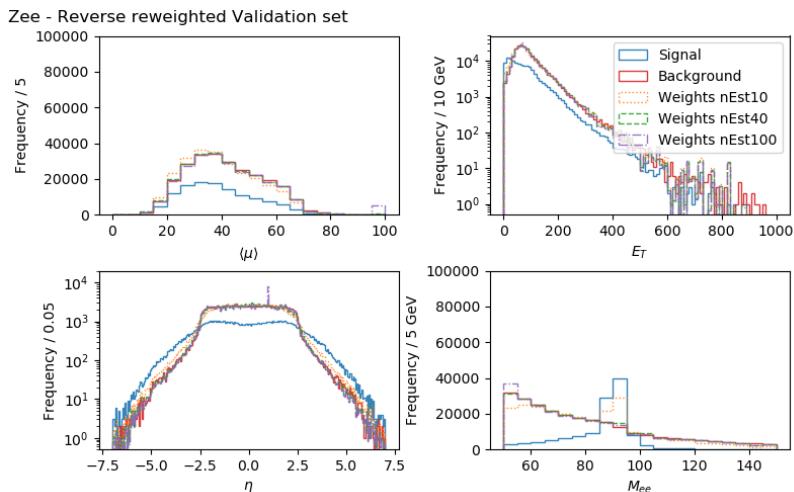


Figure A.9: Signal reweighted to the background for the electron pairs, using different number of estimators

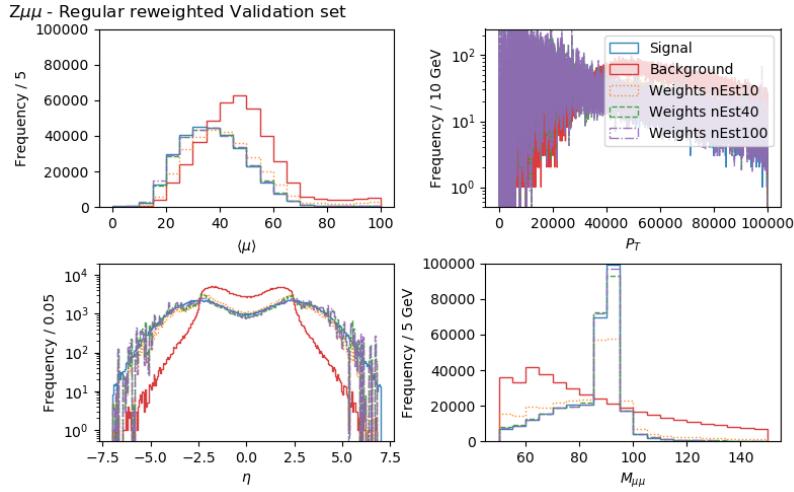


Figure A.10: Background reweighed to the signal for the muon pairs, using different number of estimators

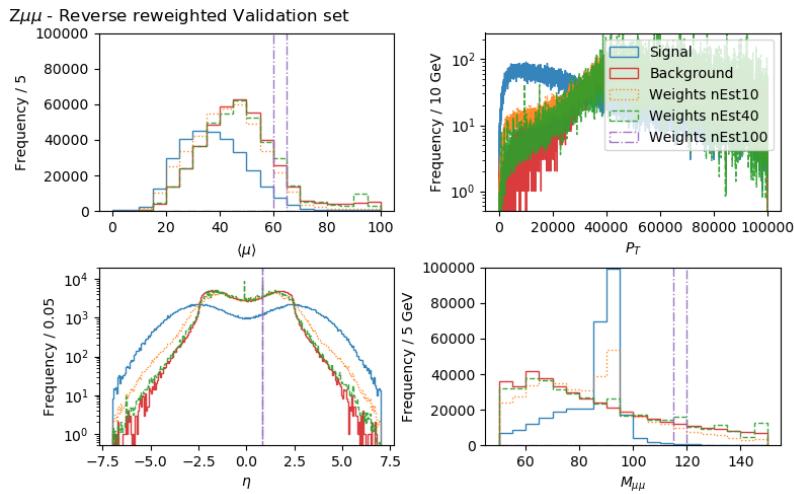


Figure A.11: Signal reweighed to the background for the muon pairs, using different number of estimators

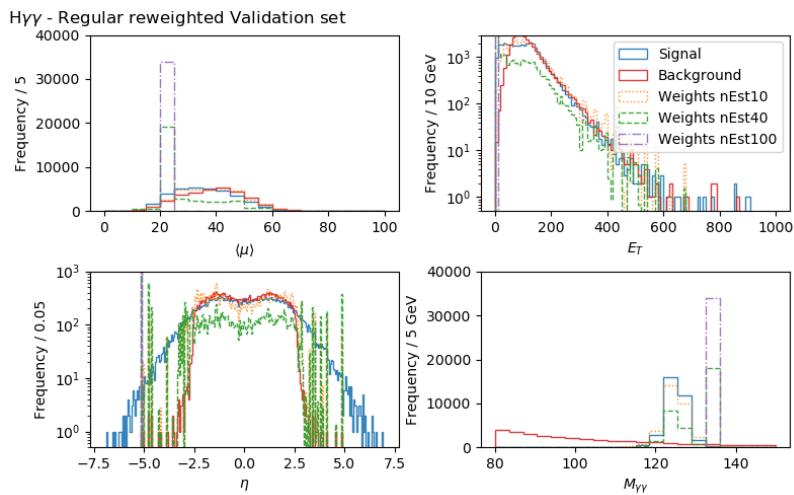


Figure A.12: Background reweighed to the signal for the photon pairs, using different number of estimators

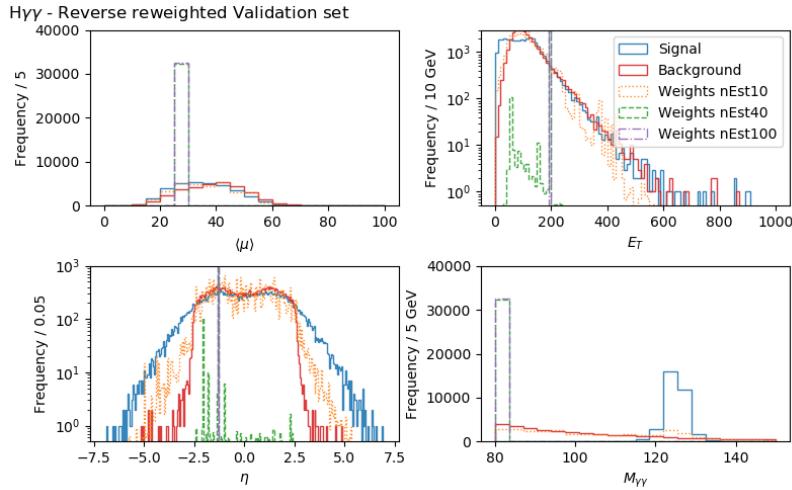


Figure A.13: Signal reweighed to the background for the photon pairs, using different number of estimators

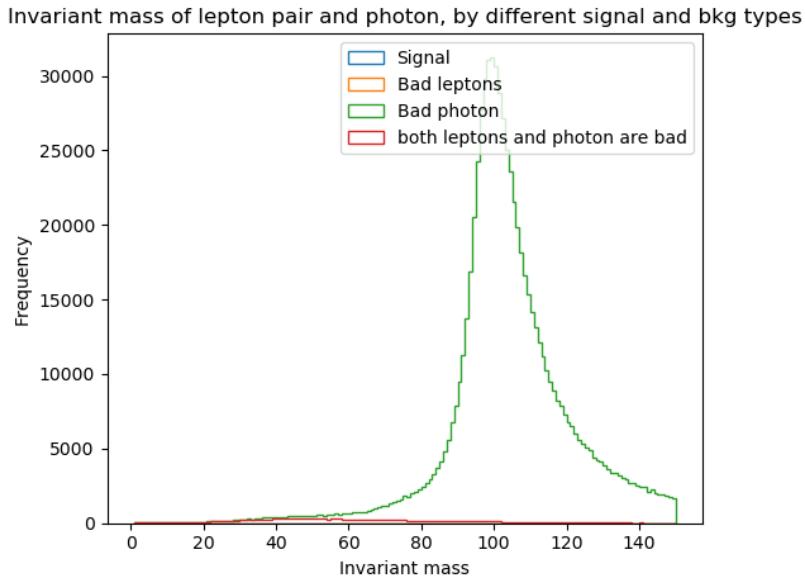


Figure A.14: The different types of background for $Z \rightarrow ee$ in the $Z \rightarrow ee\gamma$ selection

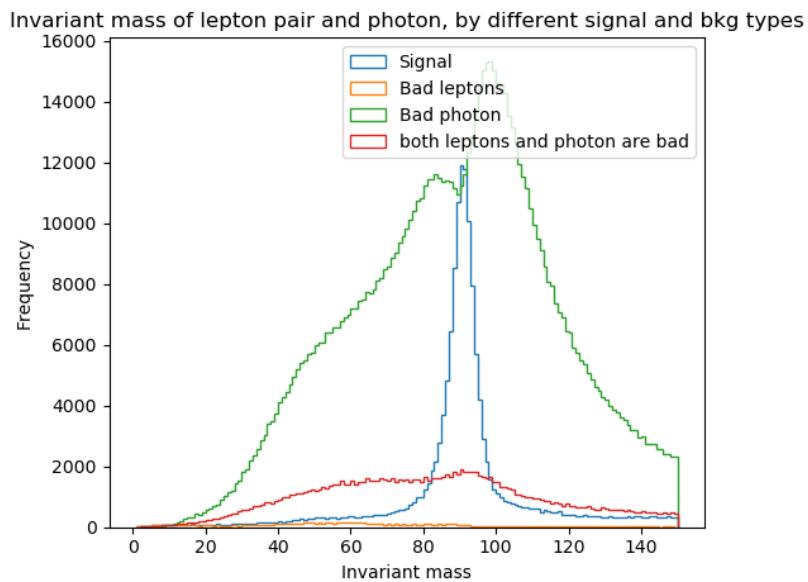


Figure A.15: The different types of background for $Z \rightarrow eeg$ in the $Z \rightarrow ee\gamma$ selection

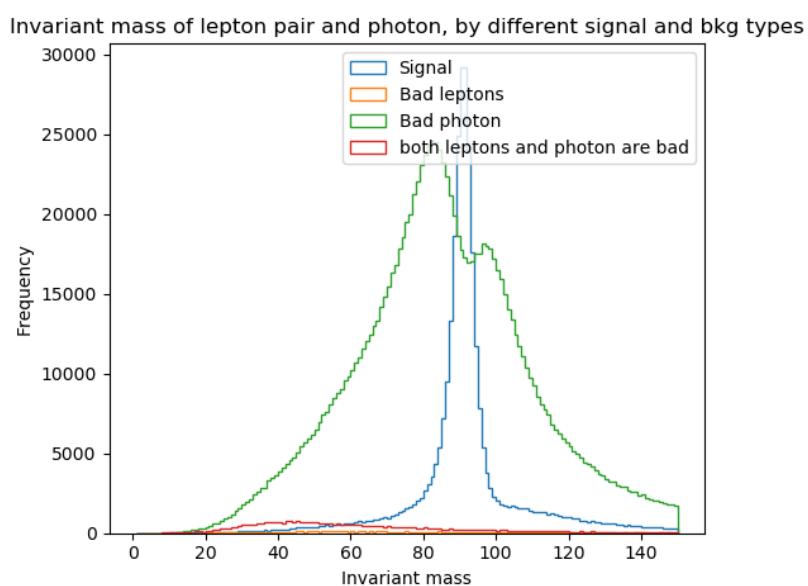


Figure A.16: The different types of background for $Z \rightarrow \mu\mu\gamma$ in the $Z \rightarrow \mu\mu\gamma$ selection

A.6 Reweigh for $Z \rightarrow ll\gamma$

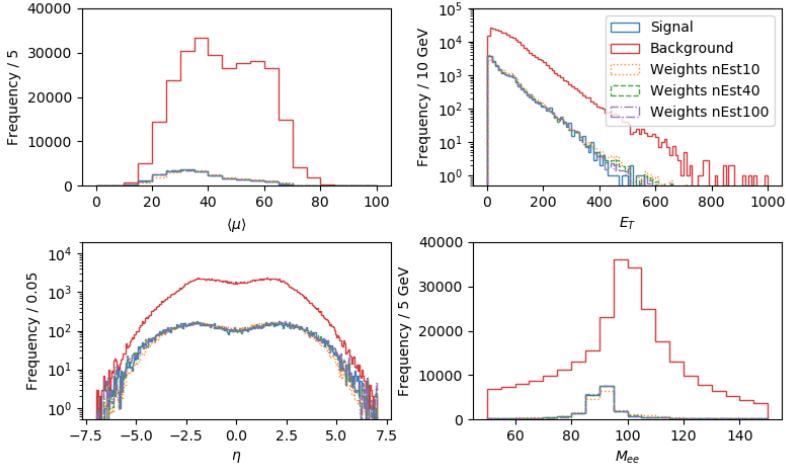


Figure A.17: Background reweighed to the signal for $Z \rightarrow ee\gamma$, using different number of estimators

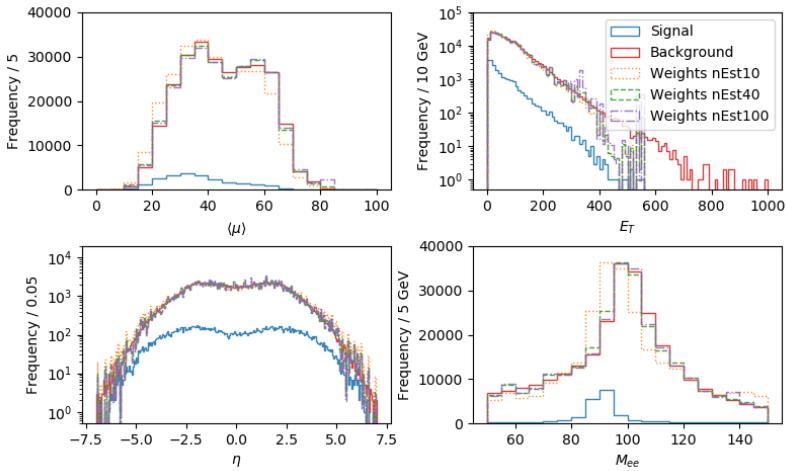


Figure A.18: Signal reweighed to the background for $Z \rightarrow ee\gamma$, using different number of estimators

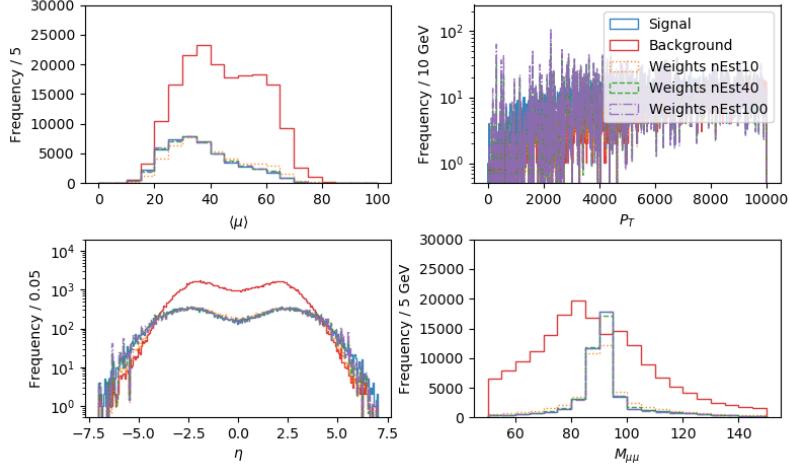


Figure A.19: Background reweighted to the signal for $Z \rightarrow \mu\mu\gamma$, using different number of estimators

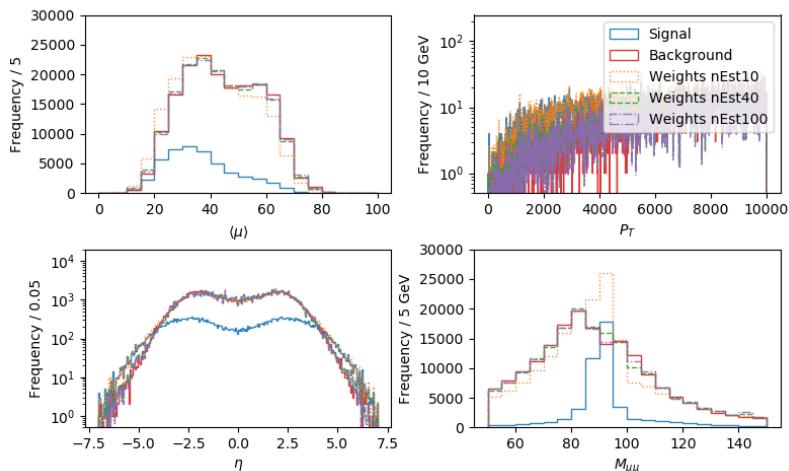


Figure A.20: Signal reweighted to the background for $Z \rightarrow \mu\mu\gamma$, using different number of estimators