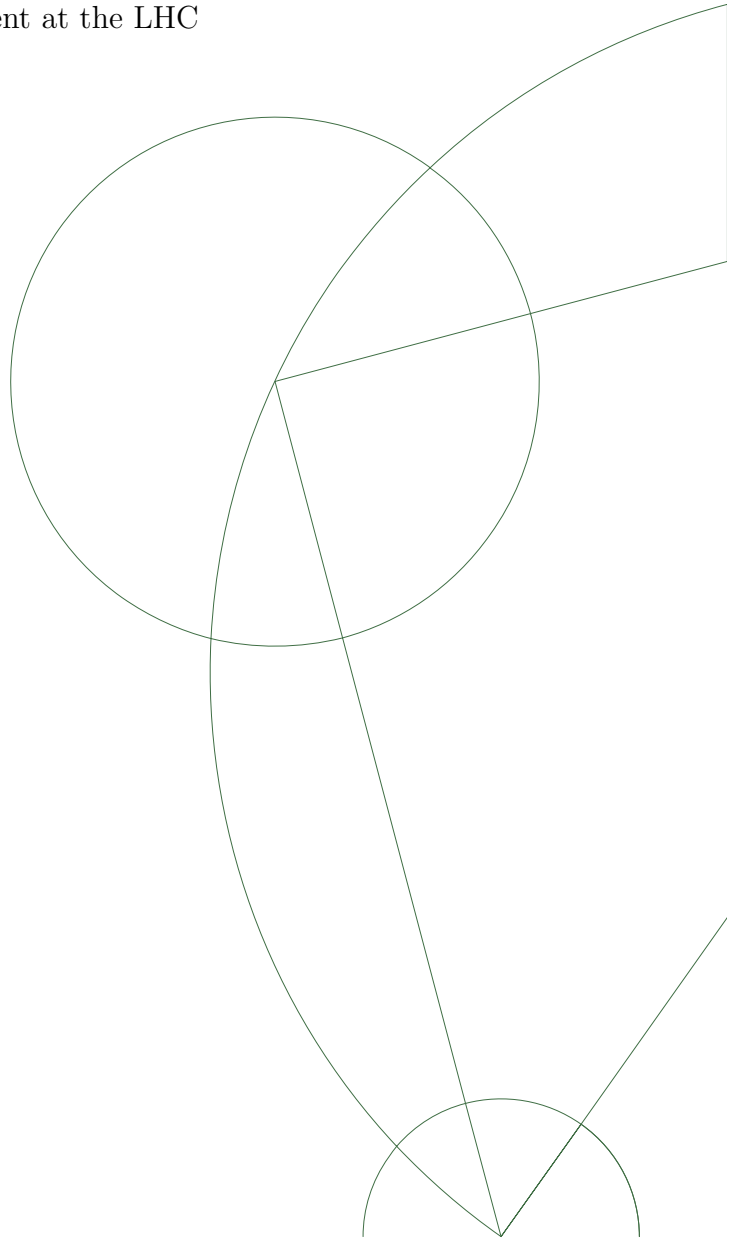
Master thesis

The Search for the Higgs Boson

Optimization of the Higgs sensitivity in the $ZZ^* \rightarrow 4l$ channel using the 2011 datasample, collected by the ATLAS experiment at the LHC

Lars Egholm Pedersen



Academic Adviser:
Troels C. Petersen

Discovery Center
Niels Bohr Institute
University of Copenhagen

August 15, 2012

This thesis is the result of the work I have made on the search for the Higgs particle, at the Niels Bohr institute, from August 2011 to August 2012. It is difficult to imagine a subject more contemporary than the search for the Higgs boson has been the last year, from the first hints observed by the ATLAS and CMS collaborations in December 2011 to the conclusive discovery of a new particle state in July 2012. This being the case, most of the work described was made without the knowledge available at the time of writing. The written work has therefore been kept in an agnostic tone reflecting the situation prior to July 2012. The use of 'Standard Model' is mostly used to describe the theory not including the Higgs boson, even though at this point it would be reasonable to take the boson as part of the model.
The search for the Higgs particle has indeed been a collaborative work relying on the combined effort of thousands of scientists working on everything from experiment design to significance estimation: A single person can only contribute to a peripheral part of the whole picture. For this reason, much of the argumentation used will rely on the work of other people, mostly from the ATLAS collaboration. The most noteworthy of this include the simulation and predicted behavior of the ATLAS detector, the optimized selection of four lepton events derived by the ATLAS Higgs Working Group 2 (HSG2), and predicted cross sections of signal and backgrounds.
The choice of variables in the described multivariate analysis and validity tests of the resulting separation has been made in correspondence with the MVA subgroup of HSG2. The method for including the multivariate analysis into the signal significance estimation has been inspired by Troels C. Petersen.
Images and figures that I have not worked out will be labeled accordingly.
The tools used for this work include ROOT - A Data Analysis Framework, for statistical analysis and graphical representation of results and the ROOT tool TMVA for multivariate analysis. Pythia8 and the PowHeg-Box for proton collision simulations. The thesis itself is written using LaTeX.

Copenhagen, August 15, 2012

Lars Egholm Pedersen

Abstract

The challenges of the $H \rightarrow ZZ^{(*)} \rightarrow 4l$ search channel are dominated by low statistics and the irreducible Standard Model $ZZ^{(*)} \rightarrow 4l$ production without an intermediate Higgs boson.
In an effort to heighten the search efficiency it is examined if the ATLAS transitional radiation tracker can improve the electron identification, loosening other requirements in the process, while the electron acceptance is tried raised by allowing maximally one to be purely observed in the electromagnetic calorimeter.
The main focus will be on the Standard Model production of a $ZZ^{(*)}$ diboson, normally considered an irreducible background to the potential production of $H \rightarrow ZZ^{(*)}$. Considering the 4 lepton final state I have in my work investigated if a separation between the two could be obtained using distributions uncorrelated to the lepton system invariant mass.
While the Higgs boson is produced mainly from gluon fusion and has zero spin, the Standard Model $ZZ^{(*)}$ diboson system is for the most part produced from quark exchange, and will in general not have zero spin, resulting in differences in their respective momenta, rapidities and angular distributions.
The result of a multivariate analysis (MVA) trained on Monte Carlo simulations will be presented, before examining its theoretical and systematic uncertainties.
The multivariate analysis is added to the signal significance estimation, usually based only on invariant mass measurements.
From simulated pseudo experiments it is expected that the additional multivariate information will increase the Higgs sensitivity by:

$$\frac{S_{m_H - MVA}}{S_{m_H}} \sim 115\%$$

Corresponding to a statistics increase of 30%.
The method is applied on the 4.8 fb^{-1} ATLAS data sample taken in 2011 at a center of mass energy $\sqrt{s} = 7$ TeV. An excess of events is observed at $m_H = 124$ GeV with local significance:

$$p_0 = 0.0004 \qquad S = 3.3\sigma$$

Taking the look elsewhere effect into account, estimated by simulated background-only pseudo experiments, the signal significance is found to be:

$$p_0 = 0.004 \qquad S = 2.6\sigma$$

Contents

List of Figures

List of Tables

1 Introduction

1.1 The Higgs search and discovery

The search for the Higgs boson has been an ongoing research project for the last forty years. Starting with the prediction of its existence in the sixties and possibly ending with the Large Hadron Collider project at CERN in 2012. [32], [23], [31].
During the time this analysis has been written, excessive amounts of data was collected by both the ATLAS and CMS detector, increasing the size of the data set tenfold. It was expected that either an exclusion or a discovery was to be established before the end of 2012. It has therefore been of highest interest to extend the understanding of the data as far as possible, extracting as much information as was achievable.

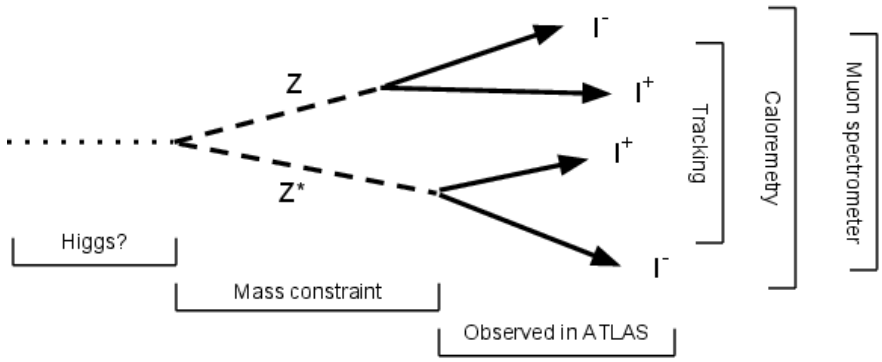


Figure 1: Graphical representation of a Higgs boson decaying in the four lepton channel. The four leptons are observed in the detector. The acceptance of the inner detector tracking, the calorimeter and the muon spectrometer differs. After reconstruction, a mass constraint is imposed on the Z candidates to reduce backgrounds.

The analysis focused on one of the most interesting decay channels of the Higgs search, the decay to two heavy neutral current bosons, which again decays to the two lightest leptons. Figure 1:

$$H \to ZZ^{(*)} \to llll \qquad l = e, \mu$$

This is the golden channel of the Higgs search, as there are extremely few things that will result in a four lepton final state, or will be able to fake it. In the mass interval the search in this analysis was performed, the number of signal and background events are roughly equivalent. This is unlike any other search channel. The probability to create a signal event on the other hand is also very low, making the search sensitive to any decision made in the selection of events.
The analysis can be divided into two parts, one tries to extends the acceptance and efficiency as far as possible while the second improves the information on the found Higgs candidates. The emphasis will by far be on the latter.
The improvement of efficiency is motivated by a signal efficiency of $\sim$20%. This reduction

mainly reflects two effects. One, the acceptance of the ATLAS detector is such, that even though it is likely that a lepton will hit the detector where it is receptacle to electrons or muons, the combined four lepton acceptance is given by the single lepton acceptance to the power of four [1]! As the ATLAS muon spectrometer extends further than the inner detector tracking, where it is required that electrons passes through, it was tried to extend the electron acceptance.
Furthermore it is by far more likely that something will appear to be an electron than a muon in the detector and some methods for improving the differentiation between electrons and electron faking objects were tried implemented.
During these examinations the ATLAS Higgs analysis was being worked on by Hundreds of scientist many of whom in relation to the four lepton channel. As such, some equivalent approaches to the ones mentioned above has been implemented.
The second part of the analysis examines the possibility that a given Higgs candidate in the four lepton channel contains more information than its invariant mass. The majority of the background to this channel is the production of four leptons through two Z bosons without an intermediate Higgs boson. To distinguish these, the production mechanisms and the systems angular momenta were examined and slight differences where found in several aspects 18. These were accumulated in a multivariate analysis, where it was found that a boosted decision tree was the optimal method for the problem at hand.

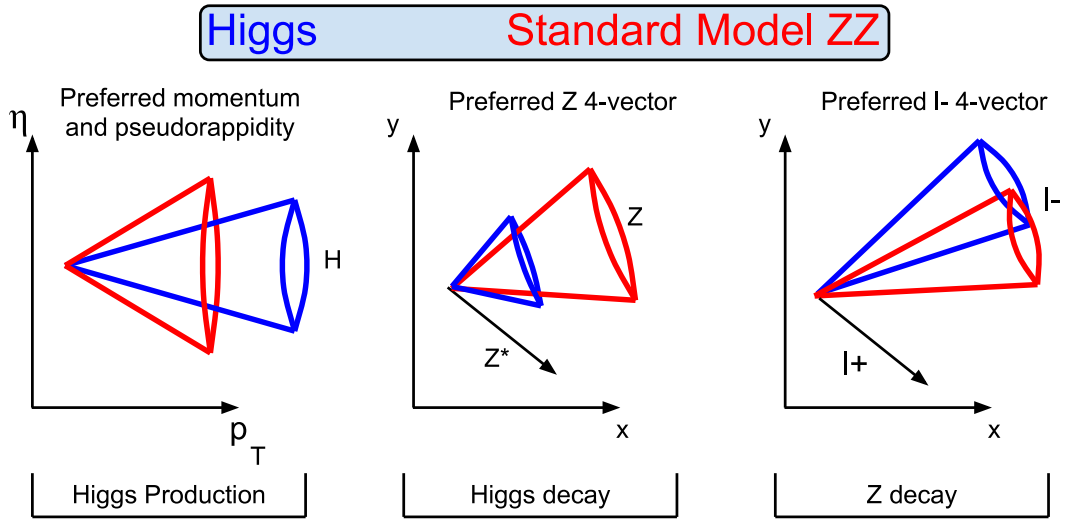


Figure 2: Graphical representation of variables used in the multivariate analysis. In the left diagram, the Higgs boson will tend to be produced with higher transverse momentum but narrower pseudorapidity range than the Standard Model diboson system. Middle diagram. Z bosons from a Higgs decay will tend to have smaller masses and have their angles correlated to a lesser degree with the Higgs candidate direction of flight. Right diagram: Spin correlations will give the final state leptons from a Higgs Z boson and a Z boson from the Standard Model diboson production different directions of flight.

[1] Ignoring lepton pseudorapidity correlations

The normal approach of particle searches, including that of the Higgs four lepton channel, is to measure the system invariant mass, and calculate the probability that the background only hypothesis would result in the found spectrum.
After examining the behavior of other backgrounds than the neutral current diboson production, the multivariate analysis information was combined with the measured mass, to give a stronger claim about the background only hypothesis.
This is will be the main result of the project.
On the 4th of July 2012 it was announced that the ATLAS and CMS collaborations combined had made a discovery of a new particle state with properties similar to those expected by a Standard Model Higgs boson. This was by all standards both an accomplishment and a great success for the joint efforts of thousands of scientist working on all aspects of the experiment. A chapter in the history of particle physics had thus ended, but much still lies ahead.
Even though the new particle discovery resembles the predicted Higgs boson, it is far from certain that it indeed is the one predicted by the Standard Model. Thus far, the strong indications of its existence has only been established in the bosonic decay channels and much more data is needed to measure that its coupling to all other particles is verily proportional to their masses.
To claim that the observed particle is the Standard Model Higgs boson, it is also necessary to establish that it is a spin-zero particle. The photon decay channel gives indications that the found state is either a spin-zero or a spin-two particle, simplifying the measurement considerably. This measurement is of high relevance to the analysis. Since the separation between the Higgs and diboson production uses the angular momentum of the system, it is thought to be geared to change focus from Higgs discovery to spin-0/spin-2 separation. Only further examination can establish this.
Finally, and this is not certain will be settled in the LHC era, it is possible that the discovered particle, sharing all characteristics with the predicted Standard Model Higgs boson has other properties as well. Theories of physics beyond the Standard Model offers scenarios of many or composite Higgses.

1.2 Thesis content

Following the Introduction, the thesis is divided into 3 parts.
The theoretical aspects of the search for the Higgs boson will be discussed in the first part. This will start by an overview of the particles describing matter and interaction found in the Standard Model, together with some comments on the mathematical framework in which everything is formulated. Following this a more detailed discussion will be given of the prediction and properties of the Standard model Higgs boson, together with some notes about theories beyond the Standard Model, also predicting Higgs-like mechanisms.
The theory part is concluded by a discussion of the more technical aspects of the work. The only possible way to make valid predictions concerning proton collisions and the following response of the detector is through simulations. The numerical Monte Carlo integration method will be described together with its implementations into physics simulations. Subsequent the multivariate analysis method 'The Boosted Decision Tree' will be described followed by a de-

scription of how the signal significance in the 2011 dataset is calculated.
The experiment part describes the Large Hadron Collider and the associated experiments, where most emphasis will be given on the ATLAS detector. The ATLAS detector consists of several subdetectors forming coaxial cylinders around the beam pipe and proton interaction point. Each of these are specialized in specific types of energy or momentum measurements, and particle identification. Analyzing both electrons and muons, a great variety of these will be put into use, and some technicalities of the relevant parts will be described.
The experiment part also discuss the data-collecting scheme. The number of proton interactions per unit time, delivered by the LHC, by far exceeds the amount of information that is possible to store. To compensate for this, complex trigger mechanisms are imposed on the ATLAS detector, signaling when interesting events have happened.
The analysis section will describe in detail the majority of the work this thesis revolves around. The topics are as described above, acceptance and efficiency improvement and creating a separation between Higgs and Standard Model neutral current dibosons. A detailed description of the systematic uncertainties of the multivariate analysis will be given before the results of the 2011 data set are presented.
The Thesis is concluded a discussion of the possible improvements of the analysis, before an outlook of the multivariate analysis method for examining properties of the newly discovered particle state.

1.3 Units

Throughout the thesis natural units will be used:

$$c = 1 \qquad \hbar = 1 \tag{1.1}$$

This choice of units means that masses and energy will be measured in units of energy:

$$[\text{Mass}] = [\text{Energy}] = \text{eV} \ , \tag{1.2}$$

where eV is the kinetic energy an electron receives through acceleration over a charge potential of one Volt. Most energies discussed are in the GeV $= 10^9$ eV area or above. In this scheme length and time are also given in powers of energy:

$$[\text{Length}] = \text{eV}^{-1} \qquad [\text{Time}] = \text{eV}^{-1} \tag{1.3}$$

When discussion macroscopic scales, e.g. the geometry of the ATLAS detector, normal SI units of course serves better, and are used accordingly.
In the calculation of a number of expected events, for a given probability and a number of trials:

$$N_{Exp} = N_{Trials} \cdot P(\text{event}) \tag{1.4}$$

It is in particle physics often more useful to convert the probability into an area; the cross section (σ). Classically this can be interpreted as the overlapping area of two colliding objects, but this explanation falls short of the full quantum mechanical complexity. This way, instead of a number of trials, an inverse area is used; the integrated luminosity (L) of the process:

$$N_{Exp} = L \cdot \sigma_{\text{event}} \tag{1.5}$$

Usually cross sections are given in barn, and hence integrated luminosities in inverse barns:

$$[\sigma] = \text{b} = 10^{-28}\ \text{m}^2 \qquad [L] = \text{b}^{-1} = [10^{28}\ \text{m}^{-2}] \ , \tag{1.6}$$

where cross sections relevant for the analysis are of the order fb $= 10^{-15}$ b, while the total integrated luminosity the ATLAS detector has collected since 2010 is around 12 fb^{-1}.

Theory

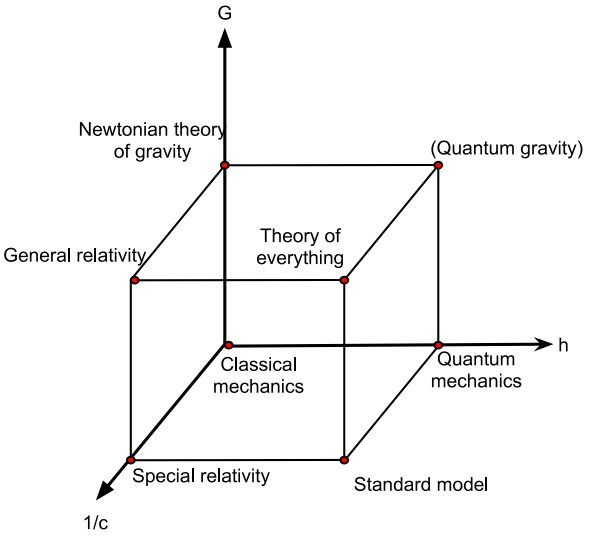


Figure 3: The Unification of theories. Starting with classical mechanics, and changing the physical constants to be non-zero, we arrive at the Newtonian theory of gravity, special relativity and quantum mechanics. It is the grand aim to combine these.

The progress the subject of physics has made through the last 400 years can be illustrated as a cube (figure 3), where one corner is centered at the origin of a coordinate system. Each of the corners of the cube will represent a theory, describing a different domain of physics. Along the three coordinate axis, we can imagine the fundamental constants; the strength of gravity (G), the speed of light (c) and Planck's constant (h). In the origin of the coordinate system we find the classical limit corresponding to a world where all constants are 0 [2]. This is the Galilean-Newtonian world of cue-balls and pulleys. Taking one physical constant to be non-zero at a time, we arrive at different theories. The Newtonian theory of gravity can be considered a world where G$\neq$ 0, describing both the movement of the heavenly bodies and earthly apples. Following the other axis $\frac{1}{c} \neq 0$ gives special relativity describing bodies moving fast compared to the speed of light, and $h \neq 0$ gives quantum mechanics describing microscopic objects. The grand goal of physics is to combine these three areas.
Two combinations have successfully been made. One, general relativity describes the combination of gravity and special relativity. And two, quantum field theory, of which the Standard Model this thesis is a test of, is the combination of special relativity and quantum mechanics. The third combination of two theories, quantum mechanics and gravity has not been described successfully and it is the opinion of most scientist that a theory containing these, would nec-

[2] Parametrize this in $\frac{1}{c}$ instead of c

essarily also contain special relativity. There are not many candidates for theories able to describe all three areas, of which string theory, for some, seems to be the most promising. If one tries to include gravity in the Standard Model, as formulated now, the theory breaks down. Most people feel however that the way of progress is to first gain better understanding of the Standard Model and its flaws, and hopefully then, the new knowledge would show the way to the theory of everything.
In this section, a short theoretical discussion will be given of why the Higgs mechanism is paramount for the Standard Model, or why the failure to measure any such would make physics beyond the Standard Model necessary, and possibly give hints of where to look.
The theory section is divided into three subsections. First a very short description the Standard Model will be given. This is included to give an overview of how nature and experiments progressively have been described using the most fundamental physical theories for the last ∼80 years. Here the particles the Standard Model contains and their fundamental properties will also be described in some detail. This will be followed by some words about the theoretical framework, in which we describe particle physics, and concluded by the theory's limitations.
Next electroweak unification will be discussed together with some properties of the Standard Model Higgs boson.
In the next section it will be described how the measurement of e.g. the W boson mass can lead to indirect predictions of the Higgs mass.
Finally some discussion of Higgs mechanisms other than that offered by the Standard Model will be presented.

2 The Standard Model

The discussion in this section will be on what we think of as fundamental particles, their properties and their interactions.

2.1 Particle Zoo

Figure 4 shows a simplified version of the particles found in the Standard Model. These are thought to be elementary particles such that they do not have any constituents[3]. A particle is characterized by a set of different properties, which include mass, charge and spin. The set of known elementary particles can be divided into matter constituents (fermions) which include leptons and quarks, and force carriers known as gauge bosons. It should be noted that all matter constituent particles also have an anti-partner. The anti particle has the same mass as the particle but opposite sign on conserved quantum numbers[4].

2.1.1 Gauge bosons

The gauge bosons describe the way particles interact. In other words they transmit the 3 forces that the Standard Model describes, the weak, the strong and the electromagnetic.

[3]Excitations of what is thought to be elementary particles, which would imply inner structure, is studied very intensely. Limits on e.g the mass of excited electrons and muons are at the current status ∼ 2 TeV [13]

[4]There are some exceptions to this when considering both left and right handed antiparticles.

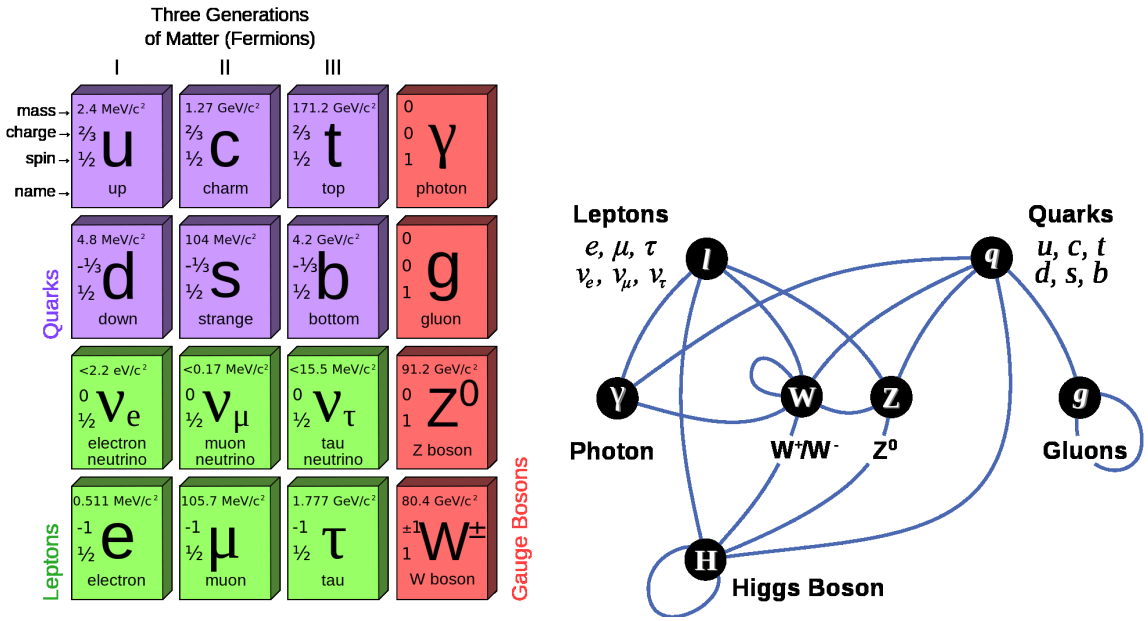


Figure 4: Left: Elementary particles in the Standard Model. Note this picture is somewhat simplified as only the left-handed fermions who are found in doublets are shown here. The right figure shows which of the Standard Model particles are able to interact [39]

The electromagnetic force is carried by the photon (γ) and couples to all particles that have electric charge. The photon is a massless particle which gives the electromagnetic force infinite range [3, p.18]. Like the other gauge bosons, it is a spin 1 particle, but due to being massless its spin has only two possible projections instead of three.

The weak force is transmitted by the two weak charged current ($W^{\pm}$) and the weak neutral current (Z^0), these couple to everything that has weak isospin. In figure 4, the top row of each type of fermion has isospin -1/2 while the second has 1/2.
The massiveness of the weak force bosons limits them to very small range scales. The W bosons has, as an example a mass around 80 GeV, which gives them a characteristic range of $\frac{1}{M_W} \sim 10^{-18}$ meters. It is for this reason that it is not seen in many everyday interactions, but at interactions with energies around the weak scale (~100 GeV), it unifies with the electromagnetic force giving them similar characteristics. The force conserves isospin. Examples of processes obeying this, can e.g. be seen in the beta-minus decay of atoms, where a down quark effectively is converted to an up quark, changing the containing hadron from a neutron to a proton.

The boson responsible for the strong force is the gluon[5]. The strong force differs vastly from the other two in that has three distinct charges, red green and blue, and it is confined. This means that there cannot exist free particles with a colour charge different from zero. If

[5]There are actually eight gluons corresponding to the allowed superpositions of color charges

Name	Symbol	Mass [MeV]	EM Charge	Weak Isospin	Color charge
leptons (spin 1/2)					
Electron	e	0.5485	-1	-1/2	
Muon	μ	113.4	-1	-1/2	
Tauon	τ	1776	-1	-1/2	
Electron Neutrino	ν_e	$< 2{\cdot}10^{-6}$		1/2	
Muon Neutrino	ν_μ	< 0.19		1/2	
Tauon Neutrino	ν_τ	< 18.2		1/2	
Quarks (spin 1/2)					
Up	u	~ 2.3	2/3	1/2	R/G/B
Down	d	~ 4.8	-1/3	-1/2	R/G/B
Charm	c	1275	2/3	1/2	R/G/B
Strange	s	~ 95	-1/3	-1/2	R/G/B
Top	t	$173{\cdot}10^3$	2/3	1/2	R/G/B
Bottom	b	4180	-1/3	-1/2	R/G/B
Bosons (spin 1)					
Photon	γ	$< 1{\cdot}10^{-18}$			
Weak charged current	$W^\pm$	$80.34{\cdot}10^3$	± 1	± 1	
Weak neutral current	Z^0	$91.18{\cdot}10^3$		-	
Gluon	g	~ 0			Octet
Higgs Boson	H^0	$\sim 126.5 \cdot 10^3$	0	0	0

Table 1: Particles and their respective properties of the Standard Model. Mass values are taken from [22], where uncertainties have been omitted. The Higgs Boson mass: [12], The gluon mass is a theoretical prediction.[22].

one tries to detach two particles who combined does not have colour charge, but individually do, the force potential grows as the distance between them grows, until enough energy is stored in the tension that a new set of particles can be created that combines with the first two, to give two new neutral particles. The strong force is responsible for binding quarks together in hadrons of which the the proton and the neutron are examples.

2.1.2 Leptons

The (left handed) leptons are made of three generations, each containing two particles. The first consist of the electron (e^-) and the electron neutrino (ν_e). The following generations contains the muon (μ^-) and tauon (τ^-) and their respective neutrinos (ν_μ, ν_τ). The electron, muon and tauon have electric charges of -1, spin 1/2 and hypercharge of -1/2 meaning they both couple to the electromagnetic and weak force. The neutrinos have hypercharge 1/2 and couple to the weak force only. The difference between the electron, muon and tauon is their masses, as can be seen in table 1 and their lepton number. The lepton number is a conserved quantity, preventing e.g. muons to decay to electrons without the emission of neutrinos. It is believed that the neutrinos also have masses, which could explain some astronomical observations, where there is a delay from the approach of light to the approach

of neutrinos from supernovas. Only an upper limit on their masses and a mass difference between them have been established however. It should be noted that the lepton number states of neutrinos is not the same as the mass states of neutrinos. Seen from one aspect there exists three different neutrinos with separate masses, and from a different aspect, there exists three neutrinos with different lepton numbers. These are not the same, which is what allow neutrinos to oscillate, i.e. change lepton number. Calculations show that the oscillations only depend on the squared difference in the neutrino masses and can therefore not be used as a direct measure [34].

2.1.3 Quarks

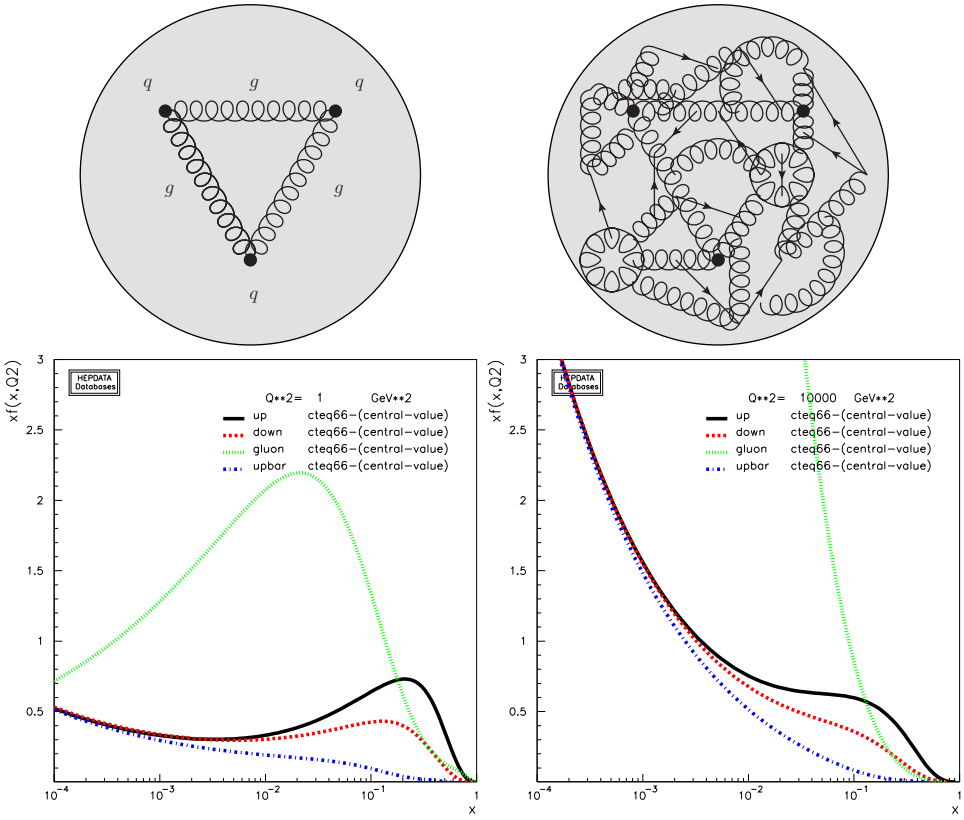


Figure 5: Left column: A model of a proton probed at low energy, containing three quarks tied together by gluons and the corresponding parton density function at $Q^2 = 1$ GeV^2. Partons are represented as: Black:up, red:down, blue:anti-up and green:gluon. Notice that the valence quark (u,d) tend to carry 1/3 of the momentum. Right column: A model of a proton probed at the weak energy scale ($Q^2 = 10^4$ GeV^2). At values of x $\lesssim 0.2$ gluons totally dominates the PDFs. Parton density functions are generated by [52].

The quarks together with the leptons form all known forms of matter. They differ from the leptons in several ways. They do not have lepton numbers. Peculiarly they have charge 2/3 and -1/3, but are also fermions with spin 1/2. The (left handed) quarks again consists of

three families, containing the up (u) and down (d) quark, the charm (c) and strange (s) and the top (t) and bottom (b) quarks, where the first mentioned have charge 2/3 and the second -1/3, figure 4. Due to the colour confinement of the strong force, the quarks tend to cluster into pairs of two (mesons) or three (baryons), while any composition is called a hadron. A great variety of searches for quarks clustering in more than three has been made, but with few exceptions, only mesons and baryons has as of yet been observed[6]. The up and down quarks are the constituents of the proton (uud) and neutron (udd), which together with the electron forms all stable matter observed in everyday life.
A complicating factor, when colliding protons, as is the case with the LHC, is that it is far from sufficient to describe the proton as three quarks. A sea of gluons bind the quarks together and confines them in the proton. Gluons moreover not only couple to quarks, but also to themselves, 4 (right). This means there is a certain probability for a gluon to not only split into a quark-antiquark pair but also into a gluon-gluon pair due to quantum fluctuations. The same actually holds true for photons splitting into e.g. lepton-lepton pairs, but since the electromagnetic coupling at these scales are ∼100 times smaller than the strong, this is a comparably small effect. Figure 5 reflects the complexity of proton collisions. In collider scenarios, it is customary to parametrize the density of each parton[7] in the relative momentum that the parton carries to the proton in the beam direction:

$$x = \frac{p_{L,parton}}{P_{L,proton}}$$

The density is also shown in figure 5.

2.2 Theoretical framework

In the following, a short overview of the mathematical framework in which the Standard Model is formulated, will be given.
Starting with quantum mechanics, we have a theory, be it the Schrodinger or Heisenberg formalism, which enables one to calculate the probability of a given outcome (Φ_O) given the set of initial conditions (Φ_I), and the Hamiltonian ($\hat{H}$) describing the system. This is normally expressed as an amplitude whose square is interpreted as a probability:

$$U(t) = \langle \Psi_O | \hat{H} | \Psi_I \rangle \tag{2.1}$$

An obvious question to ask is how a free particle propagates throughout space. This will indicate an area where quantum mechanics as first formulated fails and thereby motivates the necessity of the field viewpoint. [38, p.14].

$$U(t) = \langle \vec{x} | e^{-iHt} | \vec{x_0} \rangle = \left(\frac{m}{2\pi i t} \right)^{3/2} e^{im(\vec{x}-\vec{x_0})^2/2t} \tag{2.2}$$

This is nonzero for all values of $\vec{x}$ which means the theory predicts that a particle should be able to propagate faster that the speed of light (c) contradicting special relativity [8]

[6] The end products of the X(3872) suggests that it does not fit into this model[21, p.2]
[7] Constituent of the proton.
[8] To be stringent, what contradicts special relativity is that there is a particle type that can both travel slower and faster than c.

Without going into detail, it turns out that a solution is to exchange the position $\vec{x}$ and momenta $\vec{p}$ with fields ($\phi(\vec{x})$, $\pi(\vec{y})$), i.e. they take a value a every point in space instead of having a coordinate, still demanding that these obey the appropriate commutation relations given by quantum mechanics ($[\phi(\vec{x}), \pi(\vec{y})] = i\delta^3(\vec{x} - \vec{y})$). Described in this way, the fields can be constructed such that they preserve Lorentz invariance, reflecting their special relativistic behavior. Instead of describing particles as point like objects, they are understood as quanta of these fields. Calculations done in this manner however becomes exceedingly difficult. If the Dirac equation [38, p.43], describing quantum electro dynamics for example is considered. It is almost impossible to solve for all but the simplest of scenarios.
Luckily a different approach is available, the Lagrangian formalism which is able to express theories, in a way such that calculations can be done in a perturbative way for a wide range of scenarios. It turns out that all known physical processes can be described by maximizing an action S of the form:

$$S = \int d^4x \mathcal{L}(\phi, \partial_\mu \phi) \tag{2.3}$$

In general all of the Standard Model can be expressed as a Lagrangian density $\mathcal{L}$, which take into account electromagnetism, the weak and strong force. The expression in its full glory however fills more than would be reasonable to put in here.
Since this is also impossible to calculate directly for most scenarios, the normal approach is to restrict this to collider scenarios, where a set of particles come from infinitely long away, interacts and a new set of particles propagates to infinity as the final state.
What is great about this approach is that there exist a recipe, first realized by Richard Feynman, which converters the Lagrangian density, to a simple set of rules, [38, ch.4]. The idea is that a number of diagrams can be constructed for each level of accuracy that a given process should be calculated to. Figure 6 shows two examples of such Feynman diagrams for a made up process. If the probability that the incoming particles, represented by the left lines in each diagrams, interacts such that the result is the particles represented by the lines to the right in each diagram is to be calculated, all that is needed is to construct all independent diagrams. The level of accuracy is described by the number of vertices in each diagram, such that the left is the main contributor (leading order), and the right is part of the next order. For each line and vertex there exists a mathematical expression, which can be combined in a intuitive way, once the diagrams are constructed. The so called matrix element $\mathcal{M}$ is then the sum of these expressions. The amplitude which is the differential probability for a given input, i.e. kinematical and spin configurations, is given by the length of the matrix element. The total probability is finally the integral over all relevant configurations.

Another important feature of the Lagrangian formalism is a theorem derived by the German mathematician Emmy Noether. Loosely speaking it states that for every symmetry that a given action is invariant under, there exists a conserved current [38, p.308-310]. The well known examples is invariance under translations and time which translate to conservation of momentum and energy respectively. This principle is often a motivation when stating new theories.

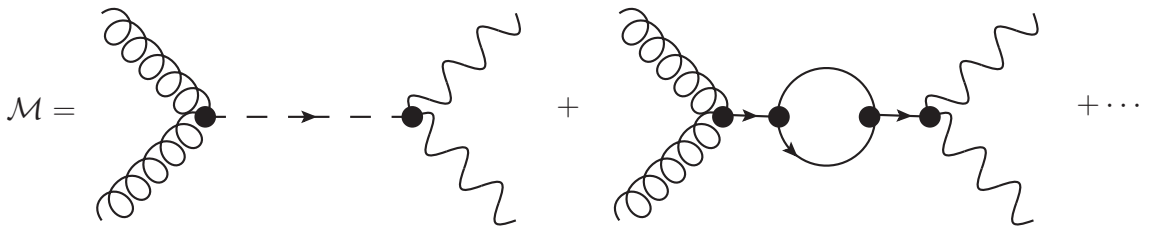


Figure 6: A thought up process where both the leading order, and one of the next to leading order diagrams are shown.

2.3 Limitations

The Standard Model has several limitations. If the new particle state observed at CERN is indeed the Standard Model Higgs boson, an explanation of particle masses would be available. But the existence of a scalar particle raises new questions, where the Standard Model predicts unreasonable behavior.

Assuming that the observed particle state is the Standard Model Higgs boson the theory still faces the Hierarchy problem. The problem arises by considering that the Planck scale, which is the natural energy scale set by gravity, is of the order order 10^{19} GeV while the weak scale is of the order 10^2 GeV. If particles can exist with masses up to the Planck scale, and thereby couples to the Higgs boson, there is nothing to protect the Higgs mass against extremely large quantum corrections compared to its mass [37, p.3-4][9]. There is no explanation of this in the Standard Model, and a very delicate fine tuning of parameters will have to be invoked.

Another more evident problem with the Standard Model is the existence of huge amounts of dark matter in the universe. Cosmological observations indicate that the visible amount of matter is only able to account for around 20% of the mass in the universe. This is seen in the angular rotation of galaxies, which would have been much smaller, if only the gravity of the visible matter pulled it together. Furthermore the so called bullet cluster has been observed [36], where galaxy clusters collide on an astronomical scale. It was observed that the visible parts interacted with each other and were slowed down, while gravitational lensing indicated that the majority of the mass passed right through leaving a discrepancy between the visible center of mass and the gravitational center of mass. No known particle of the Standard Model has sufficient mass to explain this while it also is stable and does not radiate light.

Another more obvious problem with the Standard Model is that it is not able to account for gravity. It is generally assumed that a theory of quantum gravity will include a particle dubbed the graviton. It is necessary that the graviton should be a spin two particle to be accountable for the spacetime metric $G_{\mu\nu}$. It can in fact be shown, that any spin two particle would give rise to a field indistinguishable from gravity[49, p.43-44]. If however a spin two particle $G_{\mu\nu}$ would be incorporated into the framework of the Standard Model, it can be shown that the calculations break down completely. String theory offers a solution to this, but any predictions made are so far from what is technologically possible to test, that this is, as of yet, not interesting from an experimentalists point of view.

[9] The large correction is not refereed to as the Hierarchy problem in this text

2.4 Electroweak unification

Just as was the case with Maxwell's unification of magnetism and electricity into electromagnetism. It is possible to unify electromagnetism and the weak interaction. One obvious question however is, why are the two interactions so different in the scale at which they are normally observed. In order to account for this, one has to introduce a so called spontaneous symmetry breaking, such that the forces are unified at a high energy scale and separated at lower. Another feature of this unification is that it predicts the Higgs mechanism, or at least a Higgs-like mechanism.

2.4.1 Spontaneous symmetry breaking

A way spontaneous symmetry breaking is often visualized is comparing it to a ball in the bottom of a wine bottle. The wine bottle has a cylindrical symmetric dome in the bottom, such that it looks the same no matter how you rotate it. If the ball is placed at the exact center, the system as a whole will also share this property. This however is an unstable equilibrium, and soon the ball will roll down, to a lower energy, breaking the symmetry of the system.
As a simple example, take the potential of the two dimensional field ϕ, from Goldstone [26, 4].

$$V(\phi) = -10 \cdot |\phi|^2 + |\phi|^4 \qquad \phi = r \cdot e^{i\theta} \tag{2.4}$$

It is seen that for r=0 the potential is symmetric in any direction of r and θ, but this is an unstable equilibrium. If however the field leaves the center and fall to the global minimum of the potential, $\phi = \sqrt{5}e^{i\theta}$, the symmetry in the r direction is broken, while changes in θ still keeps the potential invariant, indicating a new type of conservation.

2.4.2 A model of leptons

As originally formulated, the Standard Model is not able to explain particle masses, which of course contradicts observation. To account for this flaw another field has to be introduced to the Lagrangian density.
What was originally proposed was to introduce an exact symmetry of the Lagrangian between the electromagnetic and the weak interactions and let it be broken by the vacuum [53]. What this mean in lay mans term is that the observed fields we see under normal conditions, i.e. photons from a candle or beta-decay of certain unstable nuclei, can be described by the same set of fields but are seemingly different because of the low energy scale in which we observe them.
These fields are manifested by introducing to the Lagrangian the gauge fields $\vec{A}_\mu$ and B_μ, and a spin-zero doublet ϕ to break the electronic isospin and hypercharge symmetries and give the electron its mass. It can be derived that under an appropriate choice of gauge some very desirable features will emerge: The electron mass is proportional to the vacuum expectation value of the field ϕ: λ.
There emerges three different bosonic fields: a massless field A_μ corresponding to the photon and Z_μ and W_μ corresponding to the neutral and charged weak current. Moreover it is seen

that $M_W = M_Z \cos\theta_W$ where $\cos\theta_W$ can be measured independently [15] [10]. In conclusion it is possible, by the introduction of fields $\vec{A}_\mu$, B_μ and ϕ and using a smart choice of gauge, to reproduce the electromagnetic and weak interaction fields. But for this to work it is necessary to observe a spin-zero particle whose couplings to other particles are proportional to their masses, this is what has become to be known as the Higgs boson.

2.4.3 The Higgs boson

The existence of the Higgs boson has thus been predicted and would be able to fix some of the problems the Standard Model encounters. The theory reveals some of its properties, but unfortunately these make it exceedingly hard to observe the boson directly. The problem encountered is that the Higgs boson would couple to all particles in proportion to their mass. Since accelerators only collide light particles, due to the heavier being unstable, the probability of creating a Higgs boson is very small. Moreover since it is able to decay to everything having a mass, it is not feasible to just look at one decay channel, many of whom are totally overwhelmed by backgrounds, and wait until enough statistics has been gathered.

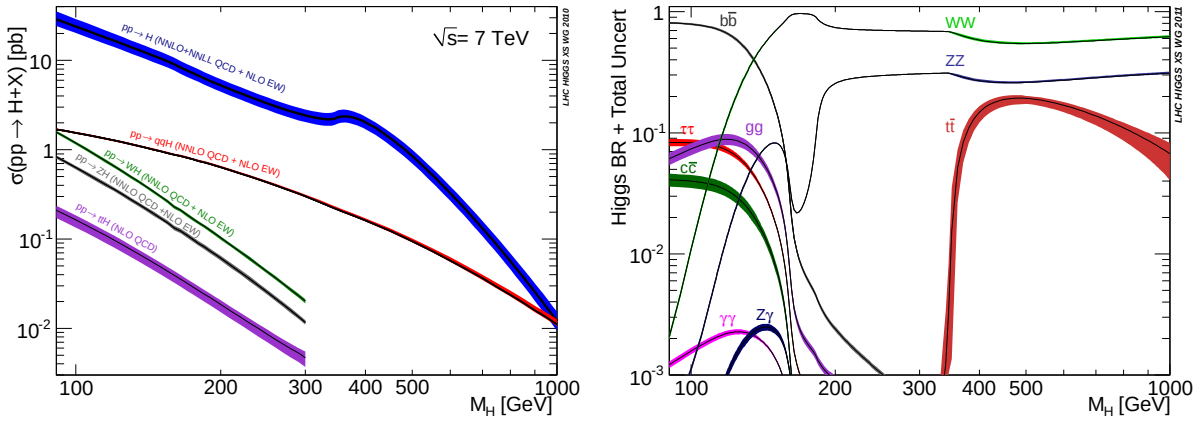


Figure 7: Left: Predicted cross sections for a Standard Model Higgs Boson[28, p.117] for 7 TeV. Right: Branching ratios of the Higgs boson [29, p.8]. Note that the dominant decay channels in the low mass region (∼125 GeV) are $b\bar{b}$, gg, $\tau\tau$ which are exceedingly difficult to detect in a hadron collider and WW whose invariant mass cannot be measured directly in the leptonic channel.

Yet another problem, which can be seen in figure 7 emerges if the search is focused on a light Higgs boson (110 GeV - 160 GeV). The dominant decay channels are here $b\bar{b}$, and gg, which have in common that they are produced in excessive amounts without having decayed from a Higgs boson. Using a hadron collider the irreducible backgrounds of these are thus so many times the size of the signal, that a discovery is virtually impossible. The $\tau\tau$ channel on the other hand involves emission of several neutrinos.
The strongest and most interesting channels are, in the low mass region, the $\gamma\gamma$, the W^+W^{-*}

[10] The notation of the weak mixing angle θ_W was introduced as a parameter later

and the Z^0Z^{0*}, where the analysis of this project will focus on the latter. The W^+W^{-*} has the advantage that it, by far, has the highest branching ratio, but has the disadvantage that the branching ratio for both W's to decay to either an electron or a muon is $\sim 4\%$[22]. The leptonic decay is accompanied by the emission of a neutrino, which makes it impossible to measure the potential Higgs boson mass directly and consequently the search problematic.
The $\gamma\gamma$ analysis has a lower branching ratio than the Z^0Z^{0*}, but since the most interesting case for the Z^0Z^{0*} channel is when both bosons decay to either an electron or a muon this channel is suppressed by:$\Gamma(Z^0 \to e, \mu)^2 \sim 0.005$ [22]. It is very unlikely that something would fake four leptons, such that the dominant background of this search is the production of Z^0Z^{0*} without it being initialized by a Higgs. There are also searches in other Z^0Z^{0*} channels, but since they suffer from either having neutrinos or being hadronic, they tend not to be as significant in the low mass region. In conclusion the $\gamma\gamma$ channel has more candidates, but also a higher background, while the $Z^0Z^{0*} \to 4l$ channel has low statistics but is very clean.
The expected cross sections and branching ratios for the four lepton channel can be found in table 2.

M_{Higgs} [GeV]	σ [pb]	$\Gamma_{H\to 4l}$	$\sigma \cdot \Gamma_{H\to 4l}$ [fb]	$N_{Exp}/5$ fb^{-1}
7 TeV				
110	23 ± 3	$2.2 \cdot 10^{-5}$	0.51 ± 0.08	2.6 ± 0.8
125	18 ± 2	$1.3 \cdot 10^{-4}$	2.3 ± 0.4	12 ± 2
150	12 ± 1	$3.8 \cdot 10^{-4}$	4.6 ± 0.9	23 ± 5

Table 2: Some relevant cross sections and branching ratios for Higgs production [28]. Here only electrons and muons are considered leptons. Cross sections includes the gluon fusion, vector boson fusion and W/Z associated production channels. For 8 TeV and 14 TeV all cross sections are scaled by aprox. 1.2 and 3 respectively.

2.4.4 Indirect predictions

The derivation of the electroweak unification did not give any indication of the energy scale of which the Higgs boson is expected to be observed. It is however possible from the knowledge of the masses of heavy particles to make indirect predictions.
If e.g. the cross section of WW scattering at high energies is calculated in the no-Higgs scenario, it is seen that it does not preserve unitarity. I.e. at high enough energies, the probability of the process exceeds 1. The introduction of a Higgs mechanism solves this problem, if the boson weighs less that $\sim$1 TeV[35, s. 2.2.1].
Moreover, the W and top mass can be calculated under the assumptions that there is no Higgs-like mechanism and that there is, where the positive assumption leads to a result better in agreement with the observed mass measurements. Turning the argumentation around, an indirect predictions of the Higgs mass can be derived given the knowledge of the W and top mass.

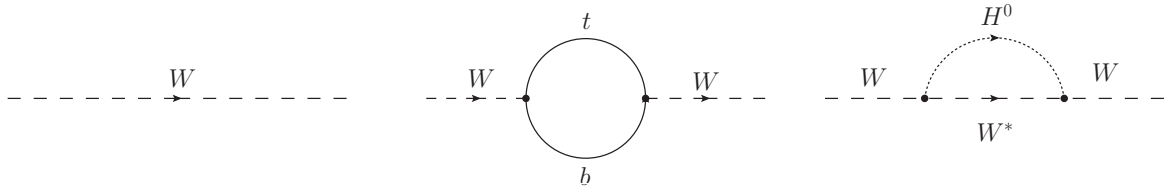


Figure 8: Left: The on shell propagator of a W boson. Middle: Largest single loop correction to the W propagator by top and bottom fermions (note one of these is an anti quark). Right: Loop correction to the W propagator by a Higgs boson.

The following calculations will follow the argumentation described in [27, ch. 4.8]. From QFT it is known that the propagator of the W boson takes the form $\sim \frac{1}{p^2+M^2}$. As illustrated in figure 8, the propagator will have higher order corrections, effectively giving it an operational mass rather than a bare mass. From the Dirac (fermions) and Klein-Gordon (bosons) equations, it can be derived that the propagators are respectively 1/p and $1/p^2$ for large momenta $p >> M$. The mass correction is then given by the correction propagators integrated over all possible momenta:

$$\delta M^2_{\text{corr}} \sim m^2_{\text{fermion}} \qquad \delta M^2_{\text{corr}} \sim \ln m_{\text{boson}} \tag{2.5}$$

As explained in the section 'A model of leptons' the bare W mass is related to the bare Z mass in the following way:

$$M_W = M_Z \cos\theta_W$$

Due to charge conservation, the loop corrections to the Z and W propagator will however be different, such that for Z the loop will contain $t\bar{t}$ while e.g. W^+ will contain $t\bar{b}$, as the largest contributions. The very precise knowledge of the Z mass, the fine structure constant α, the electroweak coupling (G) and the weak mixing angle $\cos\theta_W$ makes it possible to calculate the effective W mass up to the corrections completely dominated by the top and Higgs mass.

$$M^2_W = M^2_Z \cos^2\theta_W(1+\delta)\ , \tag{2.6}$$

where the top and Higgs corrections are:

$$\delta_t \sim \frac{3\alpha_W}{16\pi}\left(\frac{m_t}{M_W}\right)^2 \qquad \delta_H = -\frac{11\alpha_W \tan^2\theta_W}{24\pi}\ln\frac{M_H}{M_W} \tag{2.7}$$

Inverting these equations to obtain the Higgs mass as a function of the W and top mass gives a non-analytical expression, but the results can be seen in figure 9.
It should be noted that the impressive improvement of the W mass measurement (80387±19 MeV) made by the CDF collaboration at Tevatron recently, [14] gave results compatible ($\sim 1\ \sigma$) with a Standard Model Higgs boson in the 125 GeV area.

The above calculations makes the assumption that the Standard Model together with the Higgs boson constitute all particles in existence. If however heavy undetected particles exists they could also contribute to the correction loop diagrams, making the predicted Higgs mass biased.

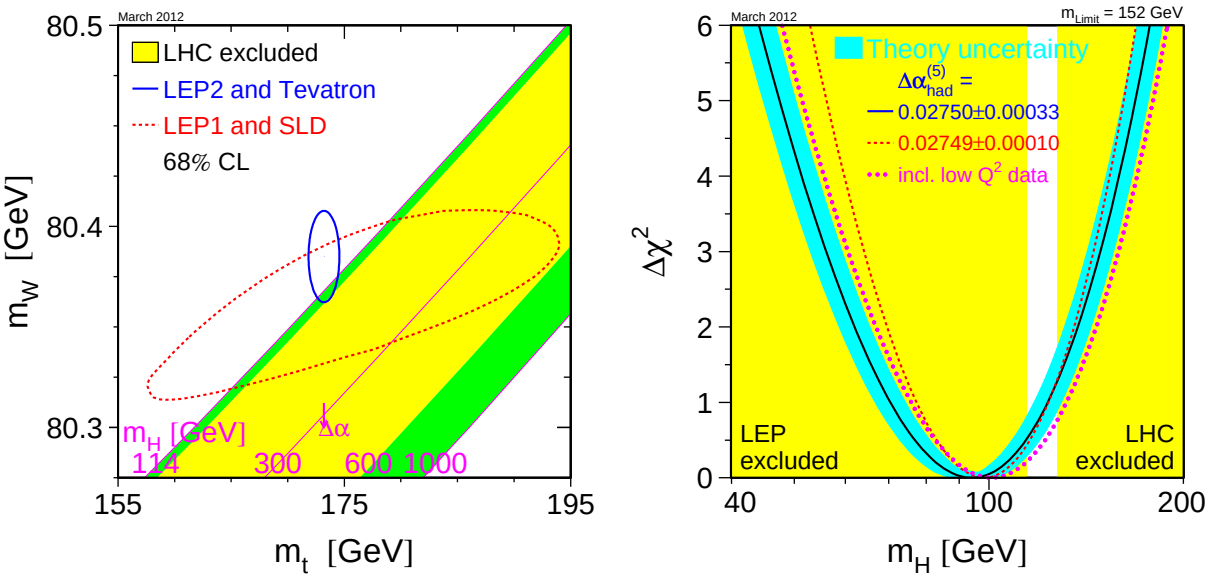


Figure 9: Left: Indirect Higgs mass prediction as function of the acknowledged W and top mass. Right: Indirect Higgs mass prediction from combined precision measurements together with LEP, Tevatron and LHC. exclusions. Figures shows the status in March 2012. [30]

2.5 Beyond the Standard Model

It is possible to construct theories which have the Standard Model as their low energy behavior but with a different solutions to the problem of explaining the elementary particle masses and unitarity violations. There have been proposed several methods to include a Higgs mechanism other than that of the Standard Model. I will here only mention some of the more popular.
Technicolor proposes the introduction of a new gauge interaction similar to the one of QCD. It could be imagined then that a Higgs-like mechanism emerges from a spin zero, meson-like particle, i.e. it is actually a composition of two unknown particle that are confined under the new type of interaction. If that would be the case, a single Higgs boson would not exist, but a series, corresponding to the exited states of the technimeson particle[19, p.5].
The theory that most people have found interesting is the introduction of Supersymmetry or SUSY. This introduces a symmetry between fermions and bosons, such that each fermion found in the Standard Model has a Supersymmetric boson partner and vice versa. In this scenario, a Higgs mechanism would still need to exist, but the Hierarchy problem would vanish. This is due to the fact that fermions and bosons carry different signs in their quantum corrections to the Higgs mass. If a symmetry existed between fermions and bosons, each contribution from a given particle would be canceled exactly by its Supersymmetric partner.
If Super Symmetry is to be compatible with experiments, it is necessary to introduce another conserved quantum number, R-parity. Without it the theory would predict too unstable protons in comparison to what is observed [46]. The proton lifetime is measured to be greater

than $2.1 \cdot 10^{29}$ years at a 90% confidence level [22].
The conservation of R-parity would also lead to the conclusion that there exists a stable lightest Super Symmetric particle, which could be a candidate for the large amount of dark matter observed.
One consequence of Super Symmetry is the existence of, not only one, but several Higgs bosons[11]. It should be noted that no signs of physics beyond the Standard Model has been observed yet, and the lower limit on many Supersymmetric particle masses, has climbed to $\sim$1 TeV.

[11] 5 in most realistic models

3 Phenomenology

In order to make valid conclusions about the existence or absence of the Higgs boson in the data, it is necessary to make a simulation of the experiment.
In this section an overview of the most basic principles of particle physics simulations will be given.

3.1 Monte Carlo integration

The favorite of numerical integration methods among particle physicist.
A widely used example, for a simple representation of Monte Carlo (MC), is the numerical estimation of π: Generate a random number set x, y uniformly in the interval $[-1, 1] \otimes [-1, 1]$, count how many falls within a circle of radius $r = \sqrt{x^2 + y2} = 1$ and divide this by the total number of generated sets. Given the area of the square with in the numbers are generated and the area of the circle it intuitively makes sense why this will converge towards π. This type of integration is called the hit and miss or Von Neumann method.
More generally any given function can be numerically integrated in this way by modification of the example above. If the function however is not confined to an interval which more often that not is the case it is not possible to confine the generated number-set to a box and more elaborate methods will have to be used. One method would be to find a suitable transformation of one generated number to an unbound function, whose integral is known. It should be noted that this can be used, not only for numerical integration, but if the numbers that falls within the desired area is kept, these will automatically also be distributed according to the PDF which is integrated.
It is also possible to estimate the uncertainty on the integral. Given that a reasonable amount of random numbers have been generated, the probability of hitting beneath the function can be estimated by $p = \frac{N_{Hit}}{N_{Total}}$, which have a binomial uncertainty of: $\sigma^2 = N_{Total} \cdot p(1-p)$, yielding that the uncertainty on the average goes as:$\sigma_{\int} \frac{1}{\sqrt{N_{Total}}}$ [18, p.44].
You thus need 4 times the computational power to half the error, which is way slower than dividing the area of integration into trapezoids and counting how many contains the integral, where the uncertainty would be inversely proportional to the square of the number of calculated points $1/N^2$. The strength of the MC integration enters when the integration is in a higher number of dimensions (D), since it can be realized that the uncertainty still goes as $\frac{1}{\sqrt{N}}$ for any. It is thus only limited by the ability to generate random numbers. The uncertainty of the box method, and more elaborate variations thereof, on the other hand grows as:$\frac{1}{N^{2/D}}$, making the Monte Carlo method more efficient at dimension higher than 4 [18, p.44].

3.2 Hard process, physics and detector simulation

Combining the knowledge of Feynman rules from the theory section and MC integration, it is possible to estimate distributions and probabilities of physical processes using a computer. Returning to the QFT discussion in the theory section, the Feynman rules derived from the QFT Lagrangian offers a recipe for calculating process amplitudes, and thereby cross sections

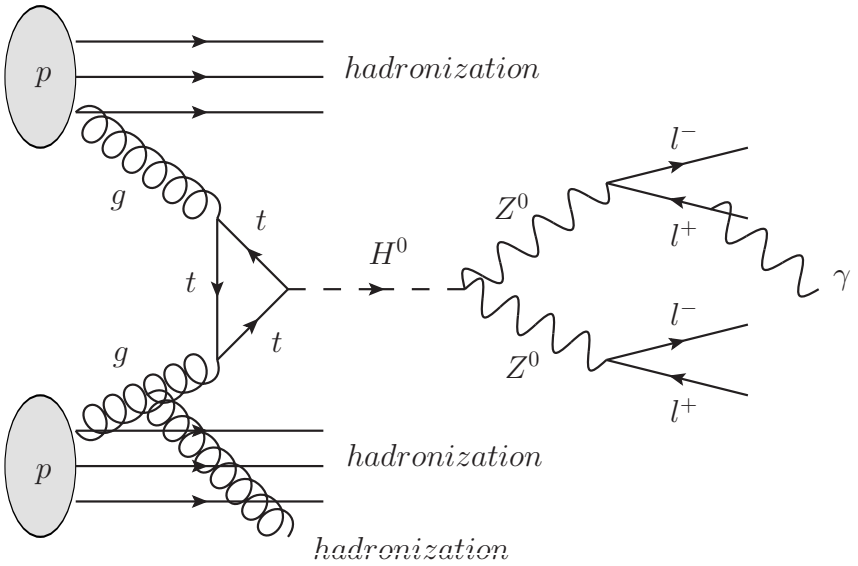


Figure 10: Diagram depicting the simulation of Higgs production by gluon fusion. The full interaction include decays, hadronization and emission of brehmstralung and is of high complexity. This calculation of the full process amplitude, is only possible numerically through MC simulations.

(probabilities). The amplitudes will in general be a function of the kinematics of the system, which is where the MC generation comes into play.
To clarify the idea to some degree, consider the process:

$$gg \to H^0$$

The corresponding Feynman diagram constitutes the center of figure 10, which can be used to calculate an amplitude. It is apparent that, due to there only being one particle in the final state, the kinematics are not very complicated:

$$s \cdot x_1 x_2 = M_H^2 \ ,$$

where x_i is the parton momentum fraction of its corresponding proton and $\sqrt{s}$ is the center of mass energy. In this case the physics simulation would work in a quite simple manner: Choosing x_1 and x_2 from the gluon PDF yields M_H fixed by kinematics, in correspondence with the above equation. Actually for Higgs masses around 125 GeV its width $\sigma_{M_H} \sim 10$ MeV, would effectively restrict (x_1, x_2) to a hyperbola. After a number of generations, the number of hits in each phase space region will be proportional to the probability of landing there, while the total Monte Carlo integration over the amplitude, calculated by the Feynman rules, gives the cross section of the process.
After this, the simulation of the Higgs decay is made, in this case to a pair of Z bosons. The generated Z boson is here described by two features: Its mass and its spin. This means that the whole system is described by the two Z boson 4-momenta and their spin. The Z bosons will then decay again leaving four lepton 4-momenta and corresponding spin. It is

still possible, but complicated, to calculate the amplitude for this process as a function of the resulting kinematics [33]. The process is depicted as the $H^0 \rightarrow ZZ \rightarrow llll$ part of figure 10. Here the strength of the Monte Carlo simulation starts to enter, since the problem now has 16 (momenta)+4 (spin) degrees of freedom, constrained by 4-momentum conservation and conservation of angular momentum.
When doing real experiments however, it is not sufficient to just consider the hard process partons. In one case, one of the initial gluons could emit another gluon before fusing as is the case in figure 10. This would result in several complicating factors. Since the final state is now more complicated, the allowed parton momentum fractions are not restricted by the Higgs mass only, and the Higgs boson is thus allowed to have transverse momentum. The generator now has to pick a random number which describes the probability of emitting a gluon at a given momentum. Several other complicating factors also emerges, such as multiple interaction, where several partons of a proton interacts, and soft scattering of partons. The subject of proton collision simulations is of extreme complexity, and much more could be said.
In the final step before the detector simulation, the simulation will have to handle the hadronization of QCD particles, from the emitted gluons or remnants of the interacting protons, as well as emission of photons. It is also worth noting that it is as of yet not possible to calculate hadronization directly, and effective and numerical models have to be used instead. In the end it is also necessary to simulate how the detector will respond to the generated particles. This is again based on a Monte Carlo approach, where effective models for particle-matter interactions are used. The detector simulation is in terms of computational power extremely expensive compared to the collision simulation.

4 Statistical techniques

This section will start by a description of the multivariate analysis method used, the boosted decision tree (BDT). This will be followed by a description of some of the more technical tools used for implementing the BDT into the analysis.
The section will conclude by a description of how the significance level of a signal is calculated.

4.1 Multi variate analysis

The multivariate analysis is applicable whenever a signal and background behavior has to be differentiated using a number of variables, possibly with complicated and correlated PDFs. It is the meaning to accumulate slight differences in each of their variables to maximize the signal and background separation. There exists a number of ways to do this, each having its own favor, but here only the Boosted Decision Tree (BDT) which is widely used in the analysis will be discussed.
The boosted decision tree is actually a composition of a set of trees, each having a structure as seen in figure 11 (left). The algorithm involved uses a training sample where it is known whether an input is signal or background. From this it derives a set of trees each giving a signal and background separation which is stronger than using a probability based on the

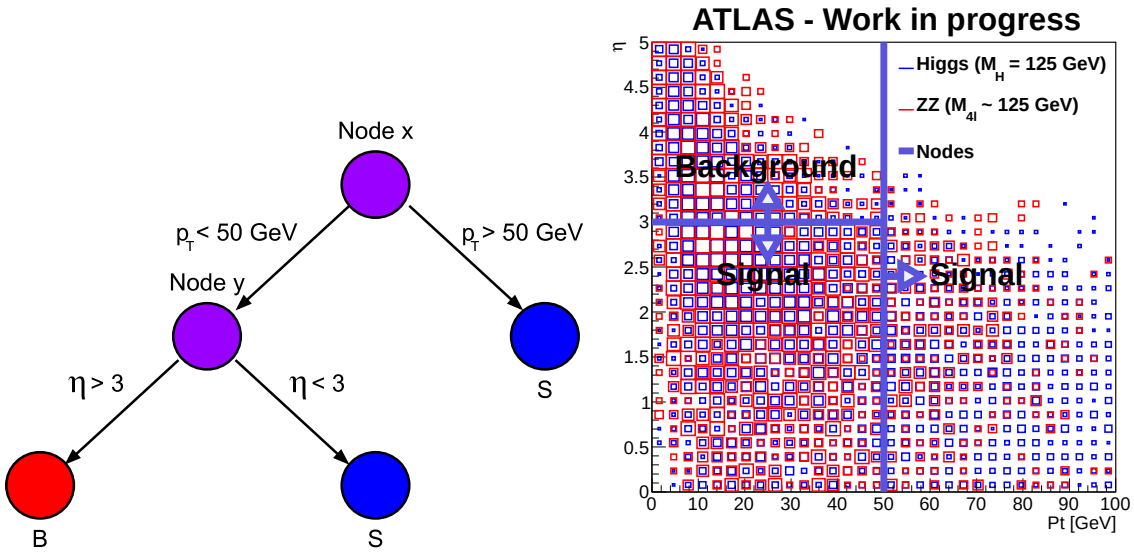


Figure 11: Left: the structure of a Boosted Decision Tree. For each node a cut is evaluated which gives a signal purity. Here variables p_T and η are evaluated. The investigated datapoint can either be sent to another node for evaluation or be considered signal-like or background-like. Right: An example taken from the analysis, with the two variables p_T and η. A cut on a single variable would not take into account correlations.

single variables[12]. To see why this is the case, consider an example from the analysis, figure 11 (right). Here the transverse momentum and pseudorapidity correlations are plotted for a given signal and background. It is seen that these are correlated and the signal tends to have higher transverse momentum. At lower momentum however, the background tend to have higher pseudo rapidity. A cut on a single variable would not take these kinds of effects into account, and separation would be lost.

A single combined parameter is finally produced as the weighed sum of the score of each individual tree.

Another important concept when discussing boosted decision trees and a wide range of multivariate analysis types, is overtraining. The data set on which the boosted decision tree is to be made from, is usually randomly divided into a training and testing sample. The trees are then created using information from the training sample. Afterward, the data from the testing sample is evaluated on the created BDT, and the PDF shapes of the testing and training sample are compared for both signal and background respectively. The worry is that if the BDT is too complex, it will also take statistical fluctuations from the training sample into account. These types of effects will be reflected in the train-test comparison. The complexity of the BDT can be limited by setting upper boundaries on the number of trees, nodes and levels allowed and demanding that each hypercube, classified either as signal-like or background-like should contain a minimal amount of data points. Also, if some variables used has small sepa-

[12]Given that the variables used are correlated in some manner.

ration power, they can in some cases be omitted such that the trees trained on the remanding variables are more protected against overtraining, allowing for more complex trees.
The algorithm used to create the trees and combining these into a BDT is too complex to discuss here. Further description can be found in [1, sect.8.12].

4.2 Probability density function transformations

The PDFs produced by e.g. the boosted decision tree will in general have random means and widths in the unit area. In the analysis of this thesis, it will be of interest to train several BDTs and transform those, such that the background distribution is the same for all trained samples. It should be noted that there is no gain in separation by doing this, but it will make the analysis more presentable.
To understand the transformation, it is necessary to first discuss some general concepts about PDF transformation:
Given a function $f(x)$ describing the probability density, we want to find a function $x \rightarrow y(x)$ such that y is distributed according to the desired PDF $g(y)$.
The probability of a random variable x' is in the interval $[x, x + dx)$, and similarly for the transformed random variable $y(x)'$ to be found in $[y, y + dx)$ is given by:

$$P(x \leq x' < x + x) = f(x)dx \;=\; P(y \leq y(x)' < y + dy) = g(y)dy \tag{4.1}$$

These must necessarily be equal for the transformation to be bijective, a necessary condition for conservation of probability. Conservation of probability also establishes that the transformation must be monotone ($dy/dx \geq 0$ or $dy/dx \leq 0$). Per construction it is normally chosen to let the transformation be monotonously growing, such that if the signal distribution is centered to the right of the background before, it will also be after the transformation.
From the above it is seen that solving the differential equation:

$$\frac{dy}{dx} = \frac{f(x)}{g(y)} \tag{4.2}$$

Will yield the desired transformation function.
A simple example, which is also used in the analysis is the rarity transformation. Here the desired shape of the transformed variable is a uniform distribution for the background on the interval $[0, 1]$:

$$\frac{dy}{dx} = f_{Bkg}(x) \Rightarrow y(x) = \int_{-inf}^{x} f_{bkg}(x')dx' \tag{4.3}$$

Another relevant example is the transformation from the rarity distribution to a Gaussian distribution. The calculation will be restricted to the unit Gaussian with mean 0 and width 1, while the more general case should follow without great complications. Inserting into the above expression:

$$\frac{dy}{dx} = \frac{1}{\frac{1}{\sqrt{2\pi}}e^{-\frac{y(x)^2}{2}}} = \sqrt{2\pi}e^{y(x)^2/2} \tag{4.4}$$

Upon examination it is seen that this differential equation is not easily solvable analytically. But some simple arguments can enlighten the problem a great deal. It is evident that the solution to this problem is a function that satisfies:

$$x \to 0 \Rightarrow y(x) \to -\inf \qquad x = \frac{1}{2} \to y(x) = 0 \qquad x \to 1 \Rightarrow y(x) \to \inf \tag{4.5}$$

There are of course many functions that satisfies this, but being in the Gaussian realm, it seems appropriate to try the inverse error function $\mathrm{Erf(x)}^{-1}$. This function can be seen in figure 12(top right). Together with the function, is depicted a set of uniformly distributed random numbers, x_i, in the [0,1] interval. Filling a histogram with $\mathrm{Erf}(\mathrm{x}_i)^{-1}$ of each generated value, results in the desired Gaussian distribution.

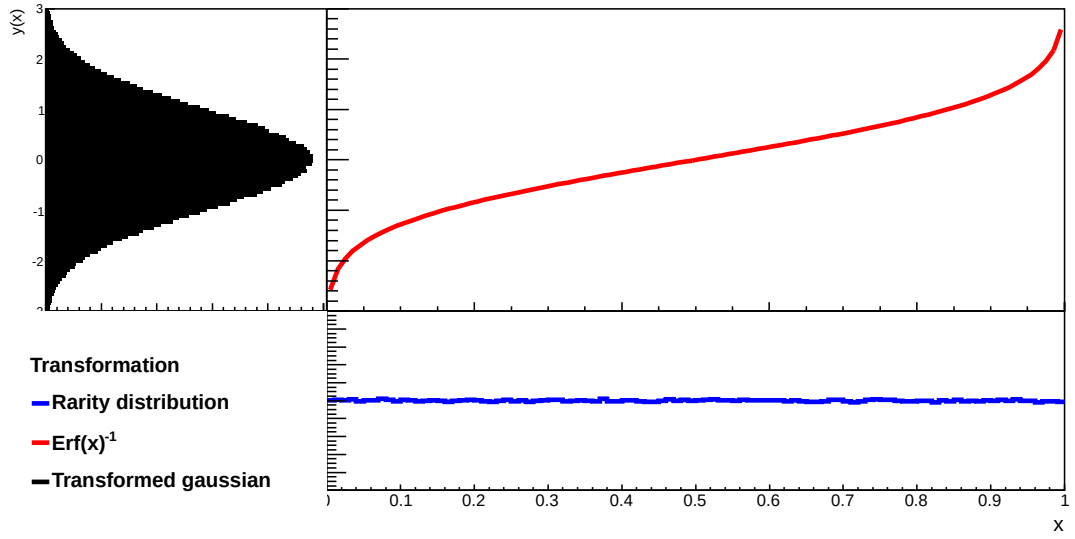


Figure 12: In the bottom right corner the distribution of 100000 uniformly generated numbers in the interval [0,1] is seen. Transforming each with the inverse error function $(\mathrm{Erf(x)}^{-1}$,top right) results in the Gaussian distribution seen in the top left.

The reason these two examples were given is that the TMVA package [1] is able to perform a PDF estimation of the function the BDT returns, which can be fed into the corresponding rarity transformation[13]. If it is desired to transform the background PDF to a unit Gaussian, as is the case in the analysis, it is technically simplest to let TMVA do the PDF estimation, return a rarity distribution, and finally convert this, using the inverse error function, to a Gaussian.
TMVA has two distinct methods of PDF estimation; PDF fitting using spline functions [1, sect.5.1], and kernel density estimators [1, sect.5.2]. The spline functions smears the output BDT data sample, and uses polynomials to fit small intervals, which are then interpolated, to give a $\mathcal{C}^1$ function if polynomials with a degree greater than one are used. It was found that

[13]TMVA does not offer the Gaussian transformation

the use of this method resulted in boundary problems, where the statistics of either the signal or background had vanished, resulting in unnaturally large signal or background estimates. The KDE method proved more stable to these problems. Here the PDF p(x) is estimated using a Gaussian smearing of the data points x_i [1, eq.19]:

$$p(x) = \frac{1}{N\ h} \sum_{i=1}^{N} e^{\frac{x-xi}{h}} \ , \tag{4.6}$$

where N is the number of data points, while h is the smoothing parameter (bandwidth). The bandwidth can either be chosen to be a constant value (non-adaptive) or x dependent (adaptive). In the non-adaptive case it can be shown, that the optimal value is given by[1, sect.5.2]:

$$h_{\mathrm{NA}} = \left(\frac{4}{3}\right)^{1/5} \sigma_x N^{-1/5} \tag{4.7}$$

For distribution RMS σ_x. The adaptive bandwidth, scales the non-adaptive bandwidth using the non-adaptive PDF:

$$h_{\mathrm{A}}(x) = \frac{h_{\mathrm{NA}}}{\sqrt{p_{\mathrm{NA}}(x)}} \tag{4.8}$$

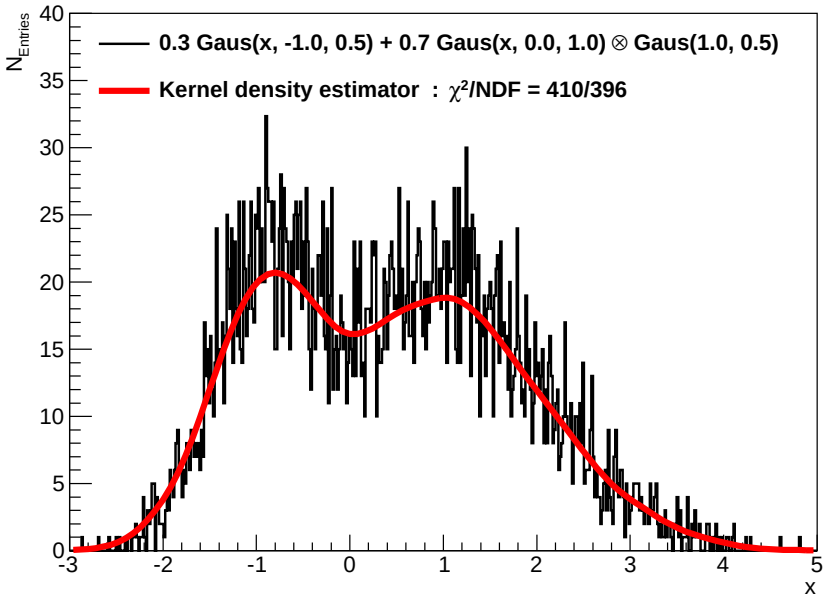


Figure 13: Example of kernel density estimation. Here 5000 random numbers have been generated according to an arbitrary complicated PDF, consisting of the sum of a Gaussian function and a Gaussian function folded with yet another Gaussian. The corresponding non adaptive KDE is shown in red.

Figure 13 shows a toy example of a non-adaptive KDE, where 5000 random numbers have been generated according to an arbitrary PDF. Some important properties of the KDE approach can be learned from this simple example. It is seen that the valley around 0 is probably assigned to great a value. It is a general feature that the approach is sensitive to low statistics, tails and peaks/valleys. The adaptive approach can help giving a better description of the low statistics areas, but can tend to wash out peaks [1, sect.5.2].
This is yet another example of a tool where the success of each implementation has to be evaluated individually.
It is evident that even though it is possible to choose the shape of one of the two output distributions of the BDT, here be it the background, the other will have to be transformed using the same map. In the relevant case the signal distribution turns out Gaussian-like, but not Gaussian.

4.3 Significance estimation

The objective of the analysis is to derive the probability that what is observed in the data is a Higgs signal, or conversely what the probability of the background faking any observed signal-like behavior is.
Usually the difference between how well the combined signal and background and the background only hypothesis describes a data sample is expressed as a significance. The significance (S) can be interpreted as a probability measure in a number of σs and is defined as the lower boundary of the integral:

$$P = \frac{1}{\sqrt{2\pi}} \int_S^\infty dx \; e^{-\frac{x^2}{2}} \; , \tag{4.9}$$

where P is the corresponding probability for the background to be distributed according to what is measured.
Figure 15 (right) shows the conversion from a probability of 0.16, illustrated as the tail integral of the unit Gaussian, to a significance of 1 σ, illustrated as the lower boundary of the integral. The conversion from the first 5 σ levels to probabilities can be seen in table 3.

σ	1	2	3	4	5
P	0.16	0.023	0.0013	$3.2{\cdot}10^{-5}$	$2.9{\cdot}10^{-7}$

Table 3: Conversion from the first 5 σ levels to probabilities.

Consider the case where a possible signal is observed in the measured mass spectrum, figure 14. Here a number of events resembling a mass spectrum containing both signal and background has been generated using a toy Monte Carlo generator. From left to right, a total number of 20, 30 and 70 events have been generated respectively while the number of signal to background events have been kept constant at 2/3, i.e figure 14a contains 8 signal and 12 background events.

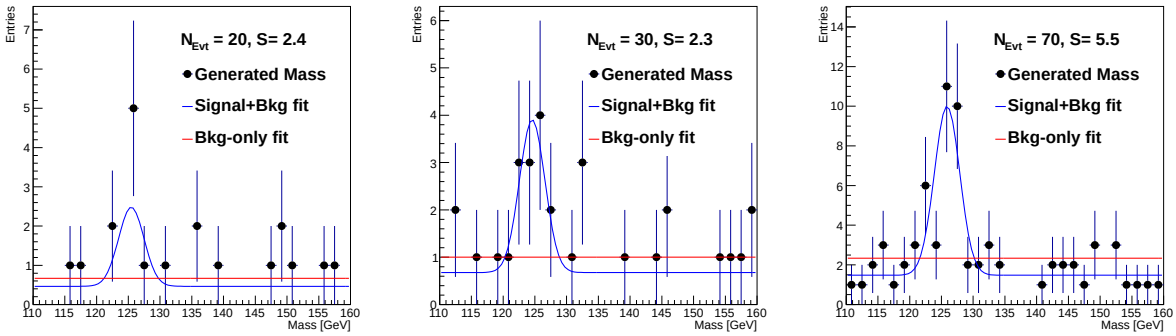


Figure 14: Signal plus background and background only fit for mass distributions generated by a toy Monte Carlo. The Significance of the signal hypothesis is given for the three cases in the top right. The ratio of generated signal and background events is 2/3 for all three cases. The samples have, for simplicity, been fitted using the *binned* likelihood method. Though the actual method applied is unbinned.

Two models are used to describe the samples; one containing signal and background and one containing background only. Theoretical considerations could e.g. predict that the PDFs (p) describing these are of the form:

$$p_{Sig}(m; m_H) = \frac{1}{2\pi\sigma} e^{-\frac{(m-m_H)^2}{2\sigma}} \qquad p_{Bkg}(m) = \frac{1}{m_{\max} - m_{\min}} \ , \tag{4.10}$$

where m_H is the signal mass hypothesis and $m_{\min}$ and $m_{\max}$ are the lower and upper boundaries of the mass search.
The following description will follow [45, sect.1], modified to fit the example in question.
For a given data sample, containing the list of measured masses m_i, the likelihood of a m, f hypothesis can be calculated:

$$\mathcal{L}_{sig+bkg}(f, m) = \prod_i \left[f \cdot p_{Sig}(m_i; m) + (1 - f) \cdot p_{Bgk}(m_i) \right] \ , \tag{4.11}$$

where the corresponding likelihood for the background only hypothesis can be constructed by fixing $f = 0$:

$$\mathcal{L}_{bkg-only}(f = 0) = \prod_i p_{Bgk}(m_i) \tag{4.12}$$

From this the profile likelihood test is given by:

$$\lambda(f_H, m_H) = \frac{\mathcal{L}_{bkg-only}(f = 0)}{\mathcal{L}(f_H, m_H)} \qquad \{(f_H, m_H) | \mathcal{L}(f_H, m_H) = \max\{\mathcal{L}(f, m)\}\} \ , \tag{4.13}$$

where the set f_H, m_H is the mass hypothesis and fraction of signal events which maximizes the likelihood function. Since there are only two free parameters in this example, as will also be the case in the analysis, a simple scan can be performed. This calculates the likelihood for masses and fractions in small intervals over the total area of the search.

The use of the profile likelihood test enters, due to Wilks theorem [45, p.3], which states that the log-likelihood ratio:

$$t = -2 \ln \lambda \ , \tag{4.14}$$

will be distributed according to a χ^2_{NDF} distribution, with number of degrees of freedom (NDF) equal to the difference in degrees of freedom between the two hypothesis, here NDF=2. Wilks theorem is strictly speaking only true for large samples, but it will be assumed that its use is appropriate in this context.
The introduction of the logarithm, as a side benefit, also eliminates possible numerical complications that can enter by multiplying many small numbers.
The test statistic t gives a measure of how good the background only hypothesis fits the data as compared to the alternative model. The probability for a sample only containing backgrounds to fluctuate over a level c is thus:

$$P(t > c) = P(\chi^2_{NDF=2} > c) \tag{4.15}$$

The expression for the cumulative distribution for the $\chi^2_{NDF=2}$ probability density function is given by:

$$\int_0^x dx' \ \chi^2_{NDF=2}(x') = 1 - e^{-x/2} \ \Rightarrow \ P(t > c) = e^{-x/2} \ , \tag{4.16}$$

where the tail probability is one minus the cumulative distribution function.
This can now be converted into a significance, measured in a number of σs, by use of the inverse error function.

$$\sigma_c = \frac{1}{2} \left(1 - \mathrm{Erf}^{-1} \left(P(\chi^2_{NDF=2} > c)\right)\right) \tag{4.17}$$

The process is depicted in figure 15, where a t value of 3.7 is converted to a significance of 1.0 σ.

It is customary in particle physics that the significance level should be greater than 5 σ to claim that a particle has been discovered.
This approach does not take into account the look elsewhere effect (LEE). In the scenario where the correct model is the background only hypothesis, when performing the search over a wide enough mass window, it is likely that a statistical fluctuation will result in a high significance at some point. If a significance is calculated for a given mass, it is necessary to correct for the probability for finding a significance of the same size or greater due to fluctuations of the background.
The simplest way to estimate this would be to use pseudo experiments with background simulations to estimate the probability of fluctuations above the found significance. It turns out in the relevant analysis that this approach requires much more simulation statistics than is available. The analysis will make use of an approach proposed by [20]:
If the search is performed over a mass window, t, can be expressed as a function of mass, where it is optimized for all other nuisance parameters, $t(m)$. It is quite intuitive, that the probability for a $t(m)$ produced only by backgrounds, to fluctuate over the significance level c is equal to the expected number of up crossings of the function $t(m)$ over the level c, denoted

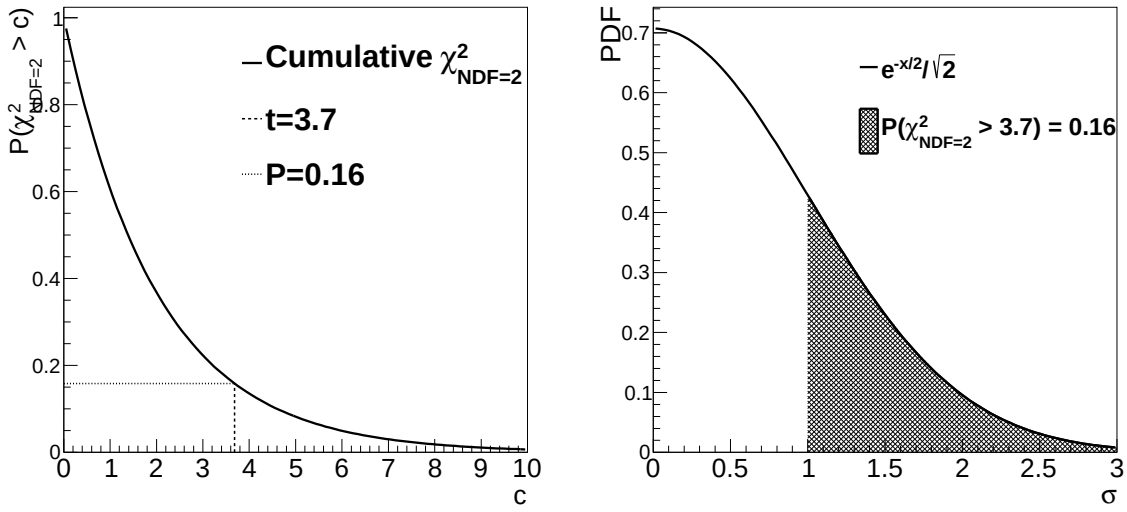


Figure 15: Example of a conversion from a log likelihood test statistic of t =3.7 to first a probability using the cumulative $\chi^2_{NDF=2}$ distribution. This probability can then be converted to a significance level using the Gaussian integral.

by $\langle N(c)\rangle$, figure 16. The expression for the tail probability now becomes [20, eq.1]:

$$P(t > c) \leq P(\chi^2_{NDF} > c) + \langle N(c)\rangle \tag{4.18}$$

The direct estimation of $\langle N(c)\rangle$ runs into the same problems with the lack of background statistics. $\langle N(c)\rangle$ can however be estimated by examining $\langle N(c_0)\rangle$ at some lower significance level c_0, [20, e.q.3].

$$\langle N(c)\rangle \sim \langle N(c_0)\rangle \left(\frac{c}{c_0}\right)^{\frac{NDF-1}{2}} e^{-\frac{c-c_0}{2}} \tag{4.19}$$

The approximation becomes more adequate for higher c, while the optimal choice of c_0 is given by NDF-1.

As is evident, this method offers a very simple approach for estimating the significance of a signal corrected for the LEE:

Examining the data, a log likelihood ratio, t, is calculated that maximizes the difference between the background only and the alternative hypothesis. The corresponding look elsewhere effect is then estimated by counting the number of up crossings $t(m)$ over 1 (NDF=2) for simulations only containing backgrounds. The expected number of up crossings at the level c is then given by:

$$\langle N(c)\rangle \sim \langle N(1)\rangle \sqrt{c} e^{-\frac{c-1}{2}} \tag{4.20}$$

An example where the look elsewhere effect for a significance level of 3σ is estimated using four pseudo experiments is seen in figure 16. Note that the LEE estimations in the analysis is derived from 200 pseudo experiments. The tail probability and corresponding significance

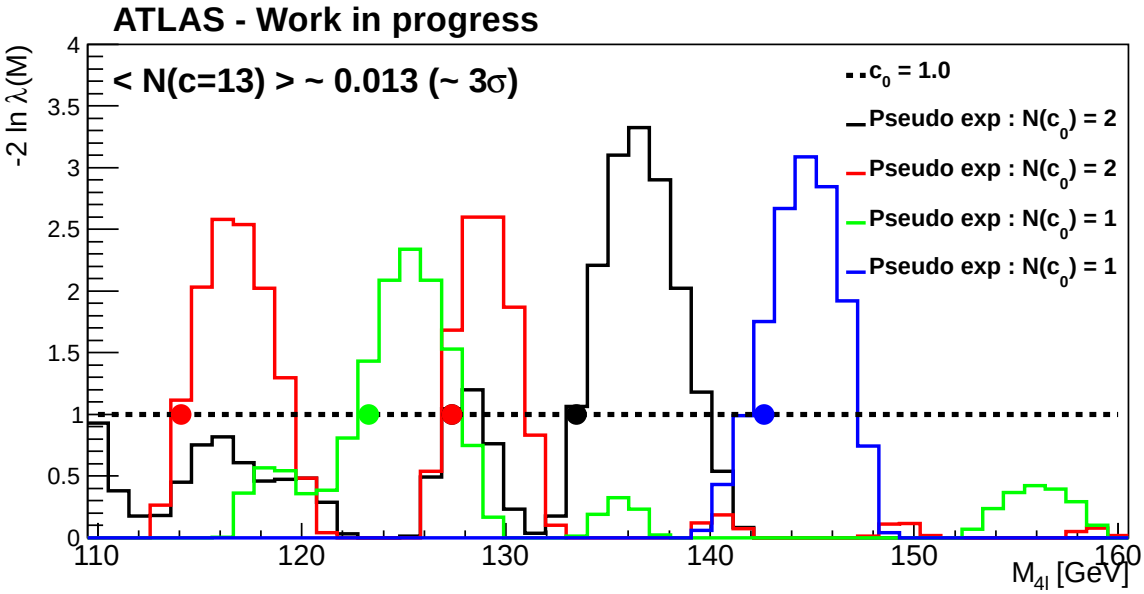


Figure 16: Estimation of the look elsewhere effect by use of background simulations. The log likelihood ratio test statistic has been calculated for four pseudo experiments as function of mass. The expected number of up crossings at a given significance level can be extrapolated from the number of up crossings at some lower level, here $c_0 = 1$.

can then be calculated as above.

In conclusion, performing the experiment, a series of mass measurements will be made. The distribution of these can be described by two different models, one containing signal and background and one containing solely background.
The difference in how well these models performs is expressed in the ratio of their respective likelihoods λ, which can be converted into a probability that the distribution of the measured masses can be described by the background only hypothesis.
A series of different mass hypothesis is tested. This raises the probability that the background will fluctuate upwards in the search window. To take this into account, the look elsewhere effect is subtracted from the local probability calculated above.

Experiment

In this section the main features of the experiment will be described with focus on the parts relevant for the analysis. The section will start with a short overview of the Large Hadron Collider and the experiments connected to it. After this, the design of the individual parts of the ATLAS detector will be discussed in detail. The information used in this section will mainly rely on [7].

5 The Large Hadron Collider

The Large Hadron Collider (LHC) is the worlds biggest particle accelerator, situated at CERN (Organisation europeenne pour la recherche nucleaire) in Switzerland. It is 27 kilometers in circumference, and is designed to be able to accelerate protons from an energy of 450 GeV to 7 TeV [14]. It is also designed to accelerate lead nuclei to an energy of 287 TeV (1.38 TeV pr. nucleon). For several reasons it has not accelerated protons to its maximal energy as of yet, but in 2010 ran with a center of mass energy of 900 GeV, 2011 7 TeV and 2012 8 TeV. Before protons are injected into the LHC they go through a rather elaborate journey of first being separated from the electron in hydrogen atoms and sent through a series of accelerators, each giving them more energy, figure 17. The protons are, after extraction from hydrogen, injected into the LINAC2, a linear accelerator, followed by the BOOSTER, the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS) and finally the Large Hadron Collider (LHC).

The protons finally collide in four different locations, where the different LHC experiments, ALICE, ATLAS, CMS and LHCb, are situated. While ATLAS and CMS are general purpose detectors, and their designs are somewhat similar, ALICE and LHCb are vastly different. The ALICE experiment focuses on heavy ion physics, and the design is optimized to be able to reconstruct lead ions collisions. Since the number of charged tracks is huge compared to proton collisions, it has to be able to differentiate these. For this reason ALICE has high resolution but can not operate with as high luminosity as the other experiments.
The LHCb design is optimized for b-physics. Where the other detectors are formed as cylinders, with the center at the collision point and the axis along the beam line, LHCb is situated to only one side, very close to the beam pipe, giving it high acceptance to b-hadrons. Theoretical predictions of e.g. CP-violations can be measured and serves as important tests of the Standard Model.

[14] While this is a big increase in energy it actually only corresponds to an increase of $\sim$700 m/s

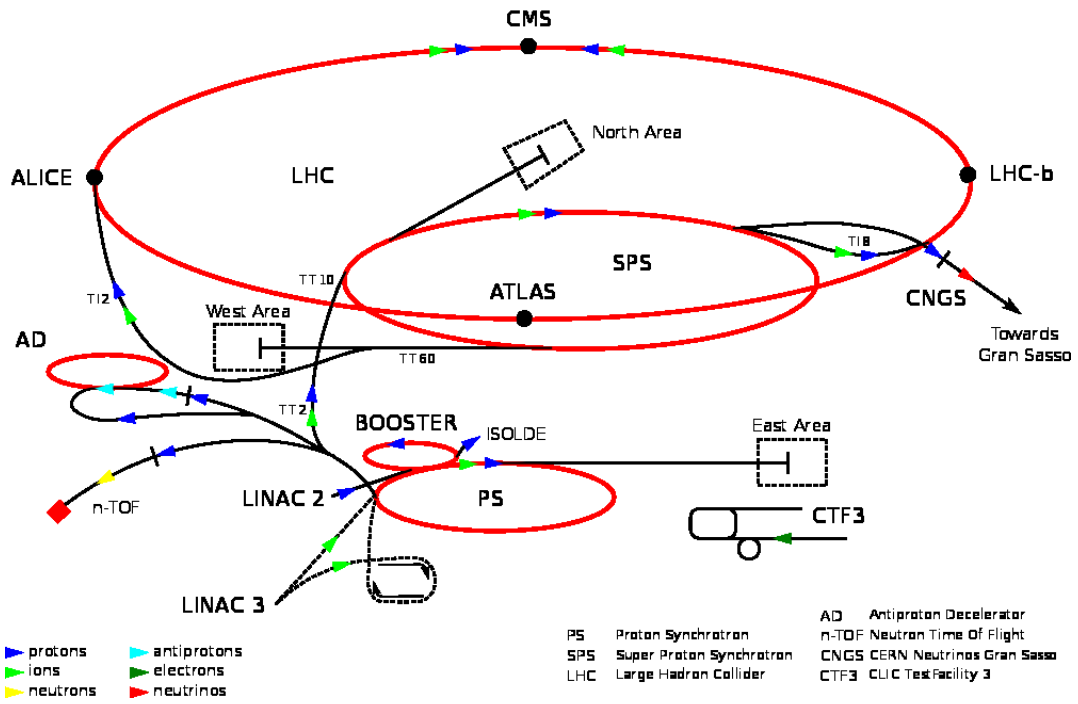


Figure 17: Map of the accelerator and detector complex found at CERN.[25]

5.1 Luminosity

The concept of luminosity is of great importance when discussing particle accelerators. The luminosity is defined such that, the number of times pr. second, a given process with cross section σ will happen, is given by:

$$N = \mathcal{L} \cdot \sigma \qquad [s^{-1}, s^{-1}m^{-2}, m^2]$$

In a collider, the luminosity can be estimated from the number of particle bunches (k) in the accelerator, their circular frequency f, the number of particles in each bunch N_i and the vertical and horizontal bunch RMS a_h, a_v [24, p.101]:

$$\mathcal{L} \simeq \frac{fkN_1N_2}{4\pi a_h a_v} \tag{5.1}$$

What is normally interesting is how many times a given particle has been produced after some time of data collection. The expression above is integrated over time resulting in the integrated luminosity:

$$L = \int dt\ \mathcal{L}$$

In order to get more sensible units, these expressions are normally converted to powers of barn: $1\ b = 10^{-28}\ m^2$.
A number of factors go into the size of the integrated luminosity. First a single bunch contains

$\sim 10^{11}$ protons. These have to be compressed as best as possible in order to maximize the probability for interaction and are expected to be have a width of ~ 20 μ when colliding in CMS and ATLAS [42, p.3]. The mean number of interactions for the 2011 and 2012 data set can be seen in figure 18(top right). The next factor that comes into play is how often proton bunches collide, this is determined by how many bunches the ring holds at a time, and by what frequency they make a full circle. So far, the bunch spacing has been 50 ns, but tests are being performed to run at the design spacing of 25 ns. All these factors combine to give a luminosity. The peak value per run can also be seen in figure 18(top left), together with the recording efficiency18(bottom left). It is truly a marvel of technology, that it is possible to measure the complex events of ∼20 protons collisions happening at a time 20 million times per second.
Finally, the length of each run and how many of those are performed determines the total integrated luminosity. A run is typically ~ 10 hours, by then the bunches are depleted due to beam-beam and gas interactions, while the experiments typically run from April to October. The total integrated luminosity from the data collection of 2010, 2011 and 2012 can be found in figure 18(bottom right).

The luminosity measurement is difficult to perform and therefore associated with some uncertainty. The simplest way to perform this measurement would be to plug in the measured number of events and the cross section of a process, in the above equation. But even for processes with high statistics, such as Z production, these numbers are limited by uncertainties of the predicted cross section, and to lesser degree, analysis performance and parton density uncertainties. Another method of luminosity measurements is to measure elastic scattering of protons, where the protons are given a slight kick in the transverse direction, and are picked up by small detectors situated far from the collision point. The relative uncertainty of the luminosity measurement is around:$\delta\mathcal{L}/\mathcal{L} \sim 4\%$ [9].

6 The ATLAS Detector

ATLAS (A Toroidal Lhc ApparatuS) is one of the two general purpose detectors of the LHC experiment.
The detector has a cylindrical shape where the proton beams follow and collide on the longitudinal axis. Moving out from the center, different layers of detector parts are situated, each with a different purpose. The description of the detector will be divided into several parts. Starting from the beam axis, and moving radially outwards, the name and a short description of the purpose of the individual sub detectors is given in table 4.

6.1 The ATLAS coordinate frame

Figure 19 shows how the two coordinate frames used in the ATLAS experiment are defined. The Cartesian, which due to the cylindrical symmetry of the detector is used very rarely, is defined such that x points toward the center of the LHC ring, y points upwards and z points in the direction of the beam such that they define a left handed coordinate system. The

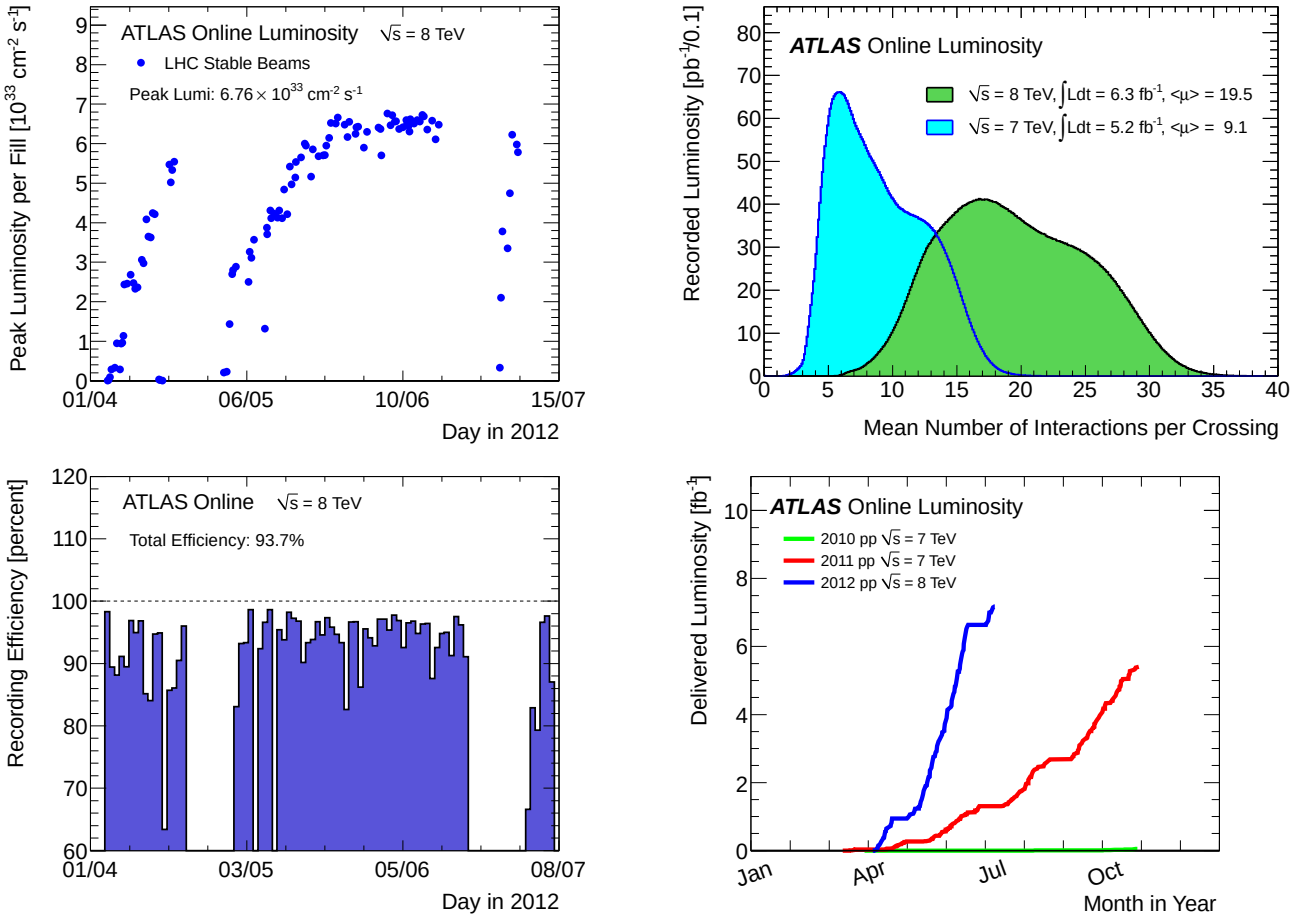


Figure 18: Top left: Peak Luminosity per day in 2012. Top right: Mean number of interactions per bunch crossing. Bottom left: ATLAS recording efficiency as function of day in 2012. Bottom right: Integrated luminosity as function of time for 2010, 2011, and 2012. Notice that the 2012 run has already delivered more data than in 2011. [5, Luminosity plots]

cylindrical coordinate system which better represent the symmetries of the detector can be calculated as follows:

$$r = \sqrt{x^2 + y^2} \qquad \tan(\phi) = \frac{y}{x} \qquad \eta = -\ln\left(\tan\left(\frac{\arctan(r/z)}{2}\right)\right)$$

The rapidity which also will be used in the analysis is defined as:

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

Notice that the pseudorapidity η is the same as the rapidity y in the limit of massless particles.

Type	Purpose
Tracking	
Pixel detector	Pixels measures bending of charged particles and interaction vertices.
Silicon strip detector	Silicon micro strips measures bending of charged particles and seeds tracking.
TRT	Gas filled tubes measures bending particles and can separate electrons from hadrons.
Calorimetry	
Electromagnetic	Liquid Argon stops electrons and photons, giving an energy measurement.
Hadronic	Stops hadrons, giving an energy measurement.
Muon spectrometer	Measures bending of muons by toroidal field in the outermost part of ATLAS.
Trigger	Indicates when a desired event has happened making it possible to record information.

Table 4: Names and purposes of the ATLAS sub detectors and trigger

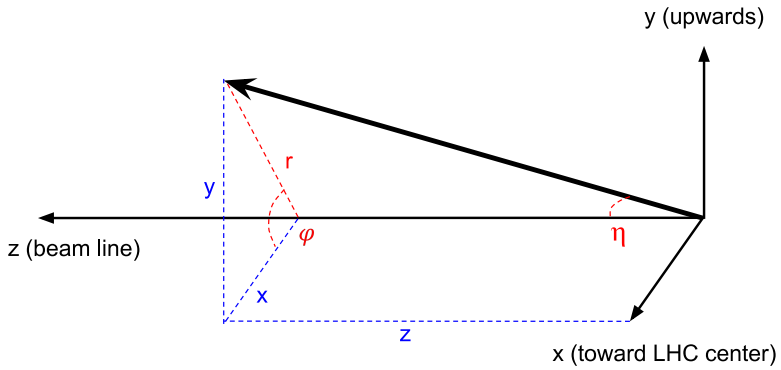


Figure 19: Figure showing the Cartesian and the cylindrical coordinate frame used in the atlas detector.

6.2 Tracking and particle identification

The purpose of the detector design, is to measure momentum and energy and to identify particles. The momentum measurement of a particle relies on the magnetic field permeating the inner detector and the muon spectrometer. The inner magnetic field is solenoidal pointing in the direction of the beam line, meaning that a charged particle traversing the inner detector will bend in the transverse plane. The output of e.g. the silicon detectors, indicates that a charged particle has passed through in some coordinate with corresponding uncertainty. Combining these 'hits' it is possible to reconstruct tracks and calculate a radius of curvature. This can, under some assumptions, be converted to a transverse momentum.
The calorimeter serves to stop particles hitting it. If e.g. an electron hits the electromagnetic

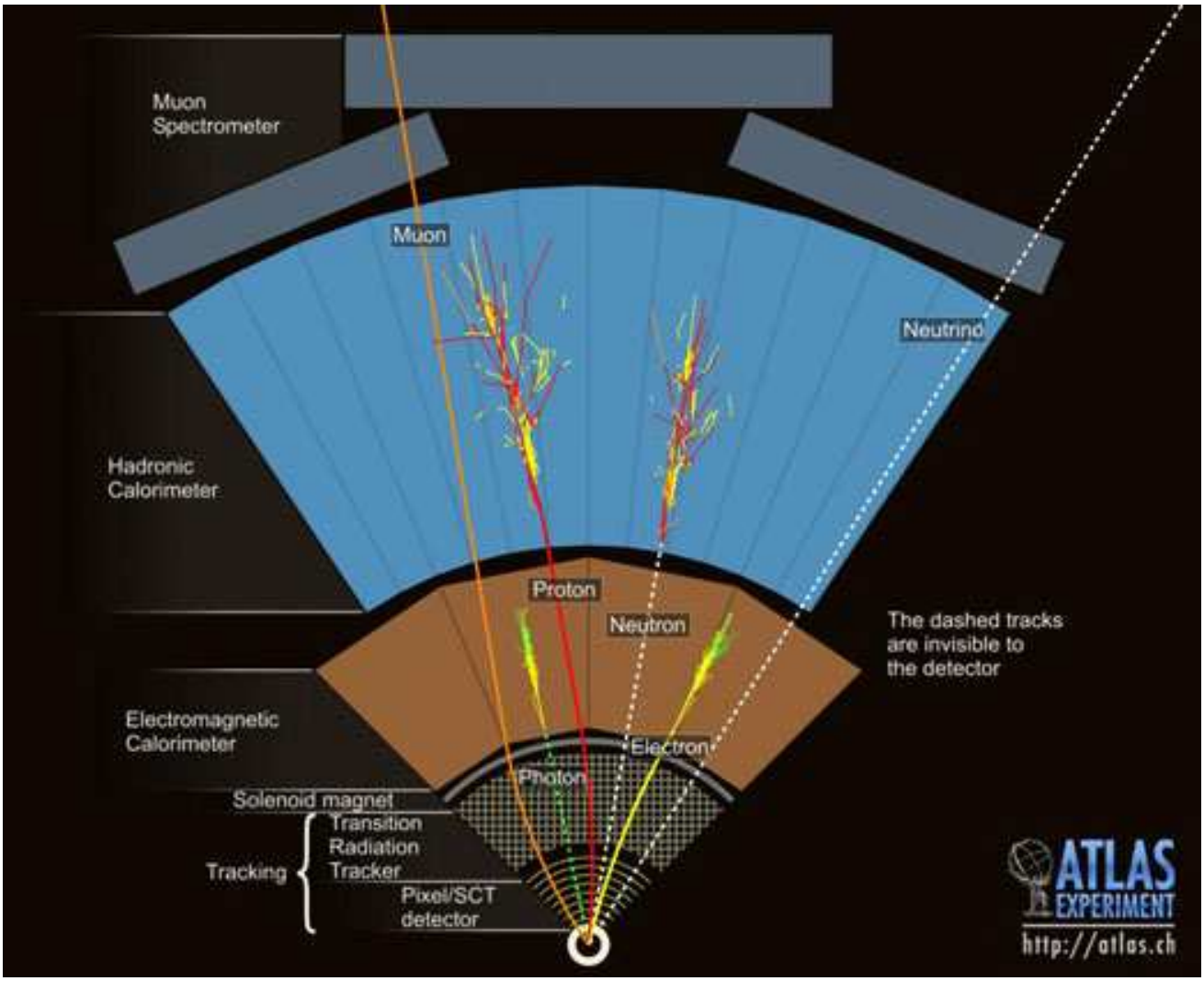


Figure 20: Depiction of how particles are identified using the ATLAS detector.**ATLAS Experiment @2012 CERN**

calorimeter, it will interact with the material, emitting a photon. The photon and the continuing electron will interact again, where the photon converts to an electron positron pair, and the electron emits yet another photon, and soon the energy of the original particle will be divided among so many that they will be stopped completely, this type of process is called a shower. By interchangeably having material with a high interaction rate and detectors able to measure the particle density as a function of coordinate, an estimate of the energy the original particle carried can be calculated. Finally it is worth noting that particles tend to interact differently with materials and magnetic fields. The detector has among other things been designed to optimize this. It is e.g. extremely difficult to tell an electron from a photon using the electromagnetic calorimeter, since the shower is produced by electrons and photons

being created interchangeably. On the other hand a photon does not carry electrical charge, which makes them pass right through the tracking system[15], such that a shower matched with a track is most likely an electron, while one not matched is a photon. Further examples of this is seen in figure 20, where e.g neutrinos are not stopped by the detector and result in missing transverse energy.

6.3 Tracking

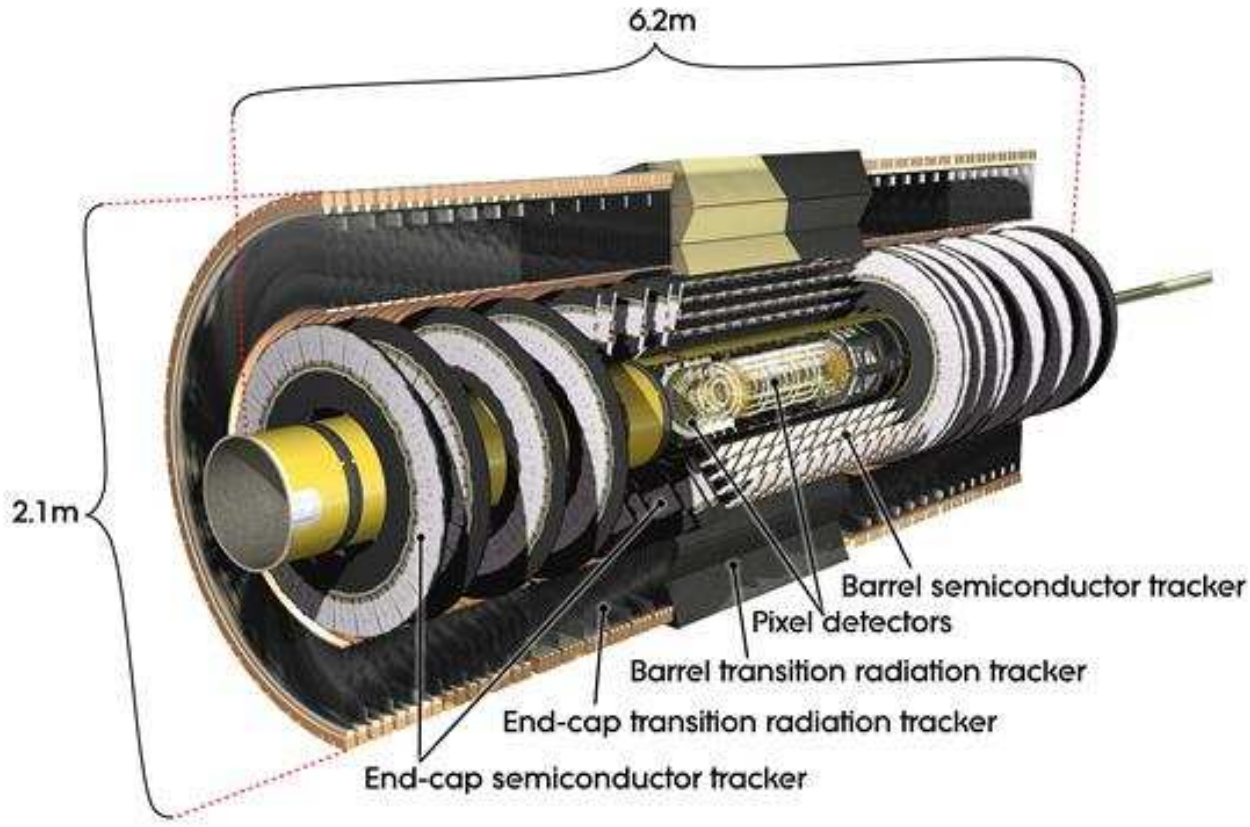


Figure 21: Drawing of the inner detector, with labels showing the main areas.

The tracking of ATLAS can be divided into several parts. The pixel, silicon microstrip detectors and transitional radiation tracker. The overall design can be seen in figure 21, while table 5 summarizes its geometry.

Part	radius [mm]	length [mm]	Aprox. η cov.	R-ϕ acc. [μm]
Pixel				
3 barrel layers	50.5 - 122.5	0 - 400.5	0.0-2.0	10
2 $\times$ 3 end-cap disks	88.0 - 149.6	495-650	2.0-2.5	10
SCT				
4 barrel layers	299 - 514	0 - 749	0.0-1.5	17
2 $\times$ 9 end-cap disks	275 - 560	839 - 2735	1.5-2.5	17
TRT				
3 barrel layers	554 - 1082	0 - 780	0.0-1.0	130
3 $\times$ 3 end-cap disks	617 - 1106	827 - 2744	1.0-2.0	(z-ϕ)130

Table 5: Geometry of the inner detector. [7, table 1.2, 4.3, p.6-7] Listed are the radius, length, approximate η coverage and characteristic accuracy of the individual parts.

6.3.1 Pixel and silicon microstrip detectors

The silicon tracking system consist of two parts: A pixel and a silicon microstrip (SCT), figure 21. These are each divided into a barrel region and end-caps. The barrel region is formed as a cylinder, and has a number of layers at constant radii from the beam axis. It should be noted that each layer of the SCT is actually able to register a particle twice, as it consist of two strips which form a small angle between them. In this way they are able to both give η and ϕ measurements. They are directed in a manner that maximizes the ϕ resolution, which again propagates the best transverse momentum resolution.
The end-caps consist of layers with constant z-coordinates, i.e. disks in to the transverse plane, used to measure higher pseudo rapidity regions.
Each pixel or strip is able to give a signal when a charged particle passes through. Using the knowledge of the detector geometry, elaborate tracking algorithms can combine these hits into tracks, [24, ch.1.7.2]. Since the inner detector is submerged in a solenoidal magnetic field parallel to the beam axis, charged particles will bend in the transverse plane. Given the radius of curvature as calculated from the tracking algorithm, it is possible to calculate the particle transverse momentum using the Lorentz force acting on a charged particle in a magnetic field: [24, p.325],

$$\dot{\vec{p}} = q \cdot \bar{v} \times \mathcal{B} \;\Rightarrow\; p_T = 0.2998 \cdot Br \qquad [\mathrm{GeV, T, m}] \;, \tag{6.1}$$

where r is the radius of curvature. The error on these measurements can be divided into two; interaction with matter (mostly relevant at low momenta) and uncertainty from the macroscopic size of the detector pixels and finite knowledge of the magnetic field strength. The design expectation on the relative error on transverse momentum is [7, table 1.1]:

$$\sigma_{p_T}/p_T = 0.05\% \cdot p_T \otimes 1\% \qquad [\mathrm{GeV}] \;, \tag{6.2}$$

where the constant term to the right hand side stems from interactions and the term proportional to p_T comes from resolution. It should be noted that the analysis is not very sensitive to the tracking uncertainty. The geometrical features of the inner detector can be seen in

[15] If pair production is not considered

table 5. The closeness of the Pixel detector to the interaction point, together with its high resolution enables it to measure interaction vertices very precisely.

6.3.2 Transitional radiation tracker

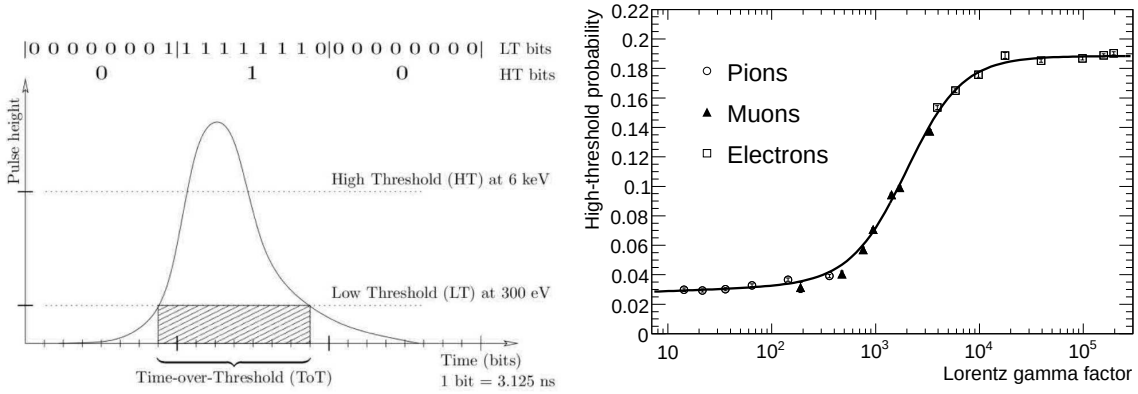


Figure 22: Left: Transition radiation pulse strength for a single TRT straw as a function of time. Notice the two different thresholds, capable of detecting ionization in the straw gas(low) and transition radiation(high) [41]. Left:Probability for getting a high threshold hit as a function of the traversing particles γ-factor. [7, p.311]

The geometry of the Transitional radiation tracker (TRT) is basically the same as that of the silicon detectors, a barrel region and two end-cap regions as can be seen in figure 21. Instead of pixels or strips, it consists of a large number of gas filled tubes with an anode wire in the center, where the tube itself makes the cathode. When a charged particle passes through the wall and enters the tube it will ionize the gas molecules who will start accelerating towards the center, knocking more electrons free on its way and in doing so starting a cascade. While the tubes are somewhat larger than the pixels and strips in the silicon detectors, they contain less material resulting in less particle interaction and they are moreover also cheaper to produce. For this reason it is possible to have a vast number of straws, such that in the barrel region there are three layers with 22, 38 and 52 tubes respectively while the end-caps have three layers with 8 tubes each [7, table 4.3].
There are important differences in tracking and momentum measurements comparing the TRT to the Pixel and SCT detectors. The barrel straws are ~ 70 cm in length [7]. For this reason the TRT is not able to measure the z component of a traversing particle in the barrel region accurately, and likewise for the transverse component using the end-cap discs. Tracking has to be combined with the other parts of the detector. The use of the TRT however still improves resolution, especially at low momenta.
Yet another complicating factor is the drift time for the gas electrons. If a signal is observed in a single straw, it is not directly possible to tell the distance from the wire in the center to where the charged particle has passed through the tube, since the time between the initial ionization of the gas to the time it hits the center is unknown. The output of several straws can be combined to solve this, but it makes the TRT sensitive to pileup.

Part	Primary function	Aprox. η cov.
EM (lAr)		
1 barrel presampler		0-1.52
1 end-cap presampler		1.5-1.8
3 barrel layers		0-1.35
2 barrel layers		1.35-1.475
2 end-cap layers		1.375-1.5
3 end-cap layers		1.5-2.5
2 end-cap layers		2.5-3.2
Hadronic end-cap (LAr)		
4 end-cap layers		1.5-3.2
Forward cal (LAr)		
3 end-cap layers		3.1-4.9
Scintillator tile		
3 barrel layers		0-1
3 barrel layers		0.8-1.7

Table 6: Geometry of the calorimeter systems, [7].

The output strength of each straw tube, a measure of how many electrons hit the anode, is divided into time bins. Two thresholds are defined, for which the output has to be over, for a number of time bins, to be defined as either a hit or a high threshold hit (hT). The relative amount of hT hits can be used to identify electrons, which will be of importance to the analysis. This is because the signal strength depends on the γ-factor of the transversing particle. Given the same momentum an electron has a ~ 200 times higher γ-factor than a pion. The resulting probability for a high threshold hit can be seen in the same figure.

6.4 Calorimetry

The ATLAS calorimeters serves to stop and measure the energy of particles that hit them. The calorimeter systems can be divided into two separate systems: The Liquid Argon (LAr) calorimeter and the Scintillator tile calorimeter. The LAr calorimeter is mainly used as an electromagnetic calorimeter, stopping electrons and photons, while the tile calorimeter is mainly used as to measure energy of hadrons. A drawing of the system can be found in figure 23, while the geometry is summarized in table 6.

6.4.1 Interaction with matter

A whole series of different phenomena has to be taken into account when a particle interacts with a medium. It can be shown that for electrons and positrons, the dominant process of radiating energy is Brehmstralung, whose cross section evolve as $\sim \frac{1}{M_e^2}$ [24, eq.2.42]. There also exists other processes, such as excitation and ionization, but of energies above 100 MeV

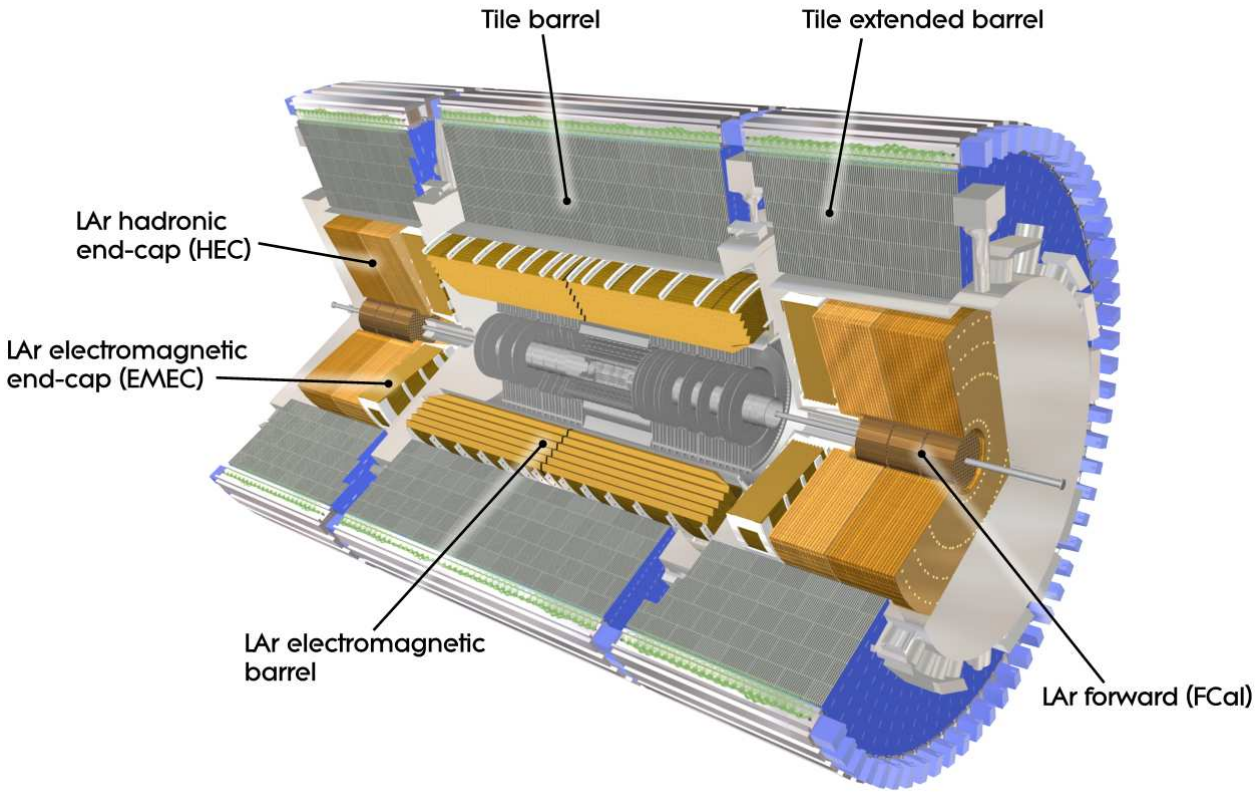


Figure 23: Drawing of the calorimeters in the ATLAS detector [7, fig.1.3] **ATLAS Experiment @2012 CERN**

these will be small compared to Brehmstralung [24, p.160].
Similarly the energy loss of photons above the same energy threshold can mainly be contributed to pair production. As mentioned above the electrons and photons entering the calorimeter will start a shower. Initially the average number of particles will grow exponentially as function of length traversed in the medium, while the energy of each will decrease exponentially. Once the particles are below the threshold where pair production and brehmstralung dominates, the other interaction types will take over and stop the particles completely.
The LAr calorimeter elements have been design to have an accordion geometry, as seen in figure 24. This has been found to optimize energy resolution in all directions.
The processes involved in hadronic showers are vastly different than those of electromagnetic. These include multiparticle production of e.g. pions and secondary particle ionization. Since it is likely that π^0s will also be produced, which decay almost instantaneously to two photons, the hadronic shower will also be mixed with an electromagnetic, increasing the complexity of describing the process. This is a demanding field where Monte Carlo simulations are paramount for any accurate description.
An interesting point is also that muons do not interact strongly, but has a mass 200 times that

of the electron, diminishing the brehmstralung cross section with $\sim 4 \cdot 10^4$. This means that it is very difficult to stop muons, and also explains why the muon spectrometer is situated on the outside of the detector.

6.4.2 The calorimeter systems

Combining the total system, the number of interaction lengths a particle traversing the calorimeter system will experience can be seen in figure 24.
Since this analysis focuses on leptons, the electromagnetic calorimeter will mainly be used. The design uncertainty of an energy measurement is: ([7], Table 1.1)

$$\sigma_E/E = 10\%/\sqrt{E} \otimes 0.7\% \qquad [\text{GeV}] \tag{6.3}$$

Comparing to the expression given for the inner detector resolution, it is seen that at low energy, tracking will give the best resolution, but above ∼30 GeV the calorimeter will give the better result. It should moreover be taken into account that an electron bending in a magnetic field, will tend to emit photons, which are undetected by the tracking system, but can mostly be taken into account using the calorimeter.
As seen in table 6 the calorimeter covers a far wider η region than the tracking.

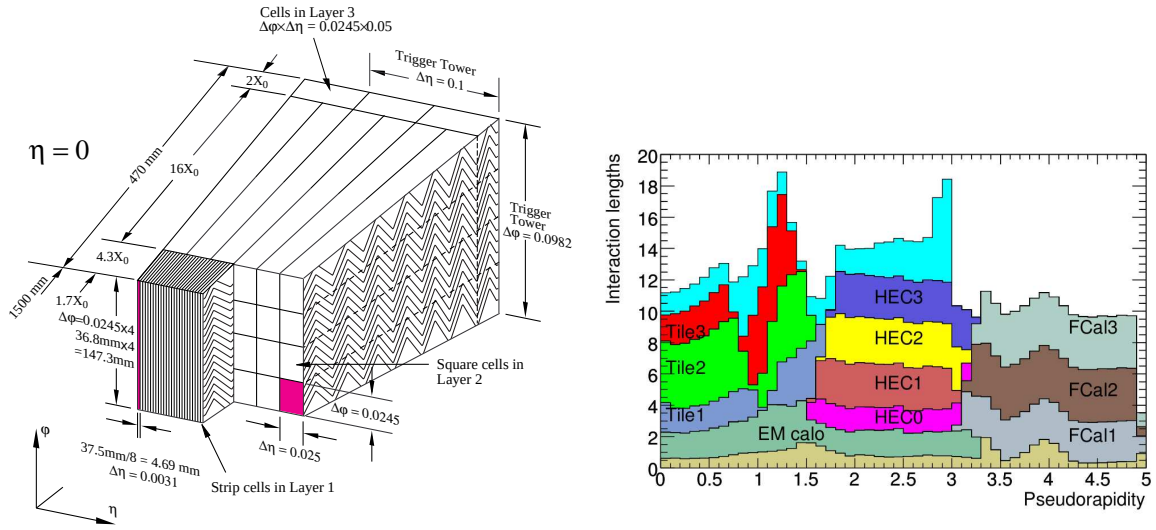


Figure 24: Left: Design of a barrel module of the LAr calorimeter [7, fig.5.4]. Right: Number of interaction lengths a particle passing through the different calorimeters will experience as a function of pseudorapidity. [7, fig.5.2]

6.5 Muonic Spectrometer

The muon spectrometer is found as the outermost component of the ATLAS detector. It is submerged in a toroidal magnetic field, which will bend particles in planes that also contain the beam axis. It is mostly muons that are able reach the central parts of the spectrometer, making both the muon selection efficiency and background rejection high.

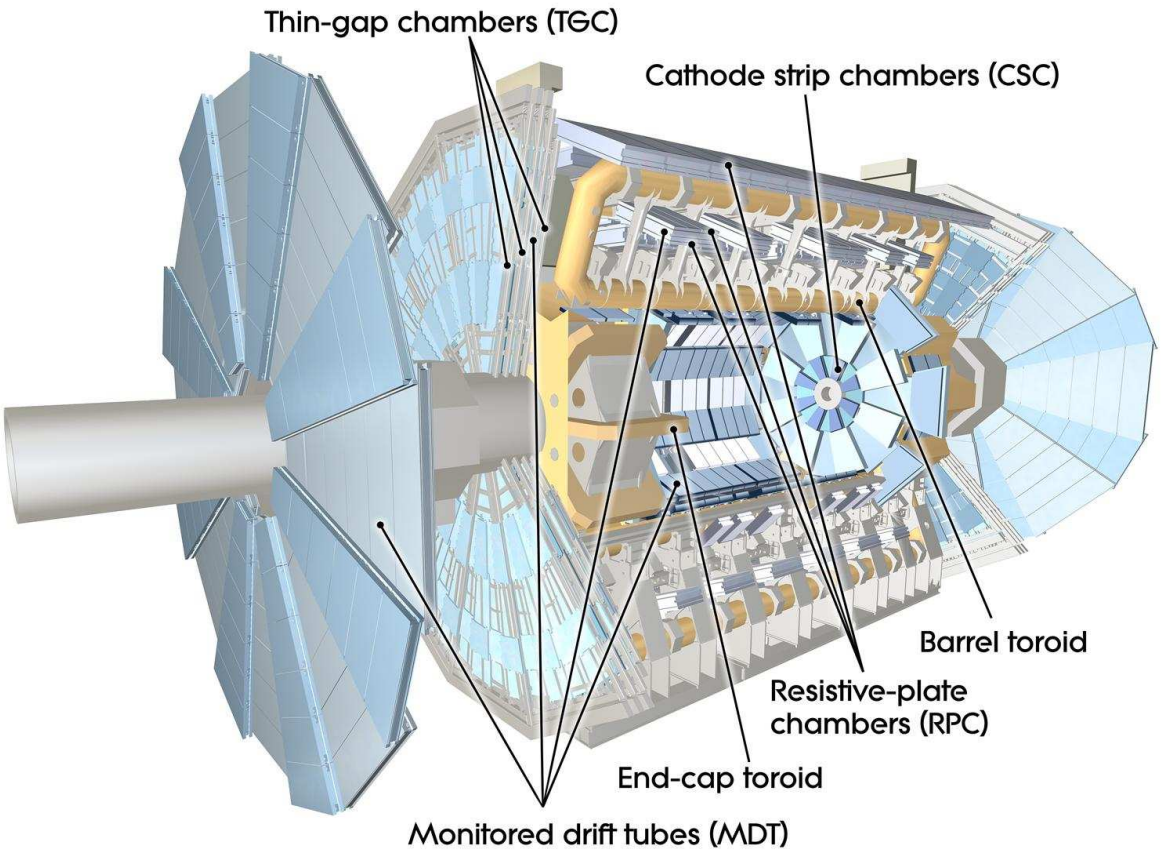


Figure 25: The muonic spectrometer of the ATLAS detector. **ATLAS Experiment @2012 CERN**

Part		Function	Aprox. η cov.
Monitored drift tube	MDT	Precision tracking	0.0 - 2.7
Cathode strip chambers	CSC	Precision tracking, radiation hard	2.0 - 2.7
Resistive plate chambers	RPC	Triggering, second coordinate	0.0 - 1.05
Thin gap chambers	TGC	Triggering, second coordinate	1.05 - 2.7

Table 7: Geometry of the muon spectrometer [7, table 1.4]

The tracking of the muon spectrometer works much in the same way as in the inner detector, while it is also implemented as part of the ATLAS trigger system.
The spectrometer consists of a number of different sub-detectors, which together with their function are listed in table 7, figure 25 shows the design.

Due to the spectrometers immense size, the bending power ($\sim B \cdot l$) of the toroid magnets is large yielding a high momentum resolution. The resolution expression is somewhat more complicated than for the inner detector, but can for a given value of η be parametrized in p_T

η region	p_0^{MS}(TeV)	p_1^{MS}(%)	p_2^{MS}(TeV^{-1})
$0 < \lvert\eta\rvert < 1.05$	0.25 ± 0.01	3.27 ± 0.05	0.168 ± 0.016
$1.05 < \lvert\eta\rvert < 1.7$	0	6.49 ± 0.26	0.336 ± 0.072
$1.7 < \lvert\eta\rvert < 2.0$	0	3.79 ± 0.11	0.196 ± 0.069
$2.0 < \lvert\eta\rvert < 2.5$	0.15 ± 0.01	2.82 ± 0.58	0.469 ± 0.028

Table 8: Resolution parameters of the muon spectrometer, [2]

[2]:

$$\frac{\sigma(p)}{p} = \frac{p_0^{MS}}{p_T} \oplus p_1^{MS} \oplus p_2^{MS} \cdot p_T \tag{6.4}$$

Dividing the spectrometer into four η intervals the resolution parameters can be found in table 8.

6.6 Trigger system

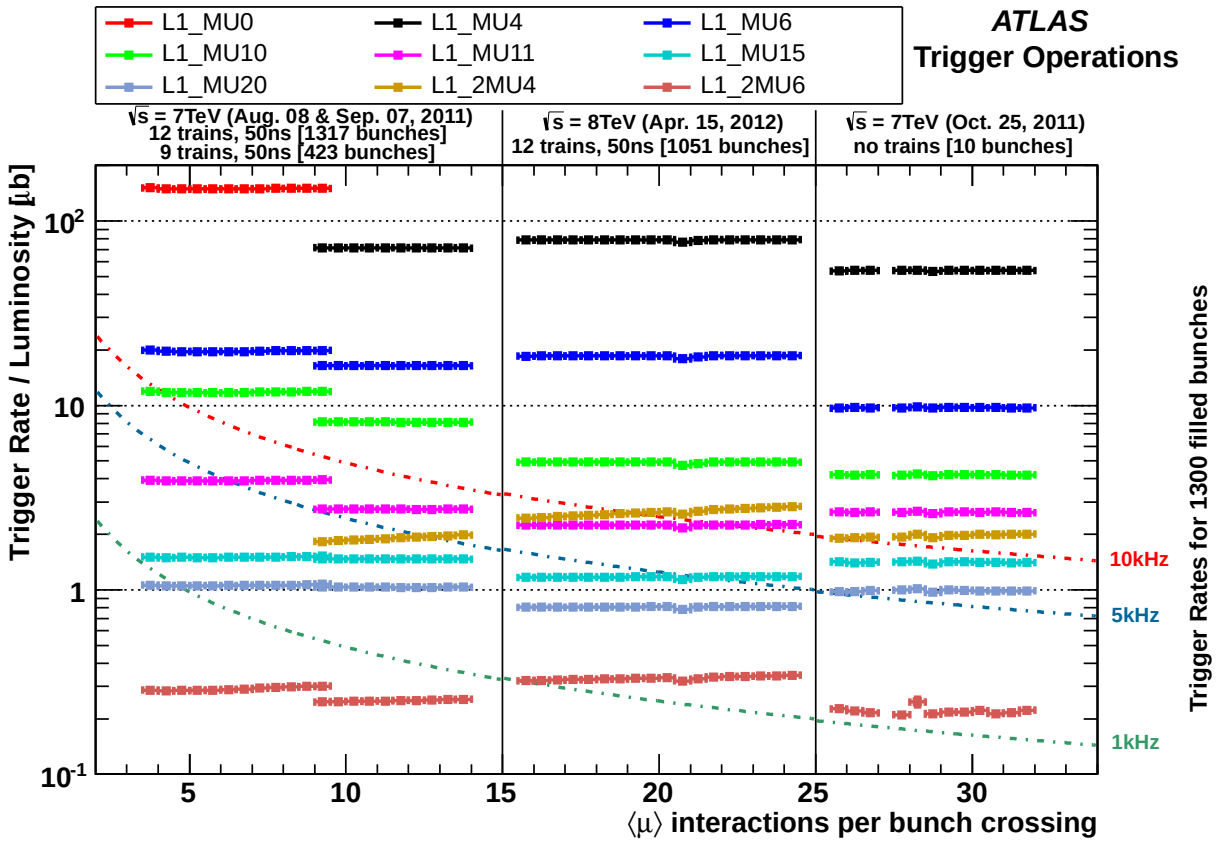
The number of interactions per second that ATLAS is able to handle, by far exceeds the computer systems ability to write information. This gives a need for a trigger system. What this does, is tell the computer system if an an event has occurred that is interesting for further analysis.

The ATLAS trigger system can be divided into three levels. The level 1 (L1) trigger looks for particles of high transverse momentum. If a particle hits the calorimeter or the muon spectrometer, with high energy/momentum it will define a region of interest which it passes on to the level 2 (L2) trigger. Some figures for the the level 1 muon trigger threshold can be found in figure 26 (top). It can be seen that the 15-20 GeV single muons or 6 GeV dimuons triggers , which are similar to those of the four lepton analysis, have a rate of ∼1-5kHz. In the discussion of luminosity above, it was seen that in 2011 and 2012 there were $20 \cdot 10^6$ bunch crossing per second. Figure 26, shows that the outgoing rate after the combined level 1 trigger is reduced to $\sim 10^5$, which is passed to the level 2 trigger. The L2 trigger then uses all available information from the detector to evaluate if an interesting event has occurred, this is possible because the L1 trigger has reduced the areas to analyze to approximately 2% [7, p.15]. The level two trigger reduces the examined event rate by more than an order of magnitude and $\lesssim 10^4$ is passed on to the event filter which reduces the the rate to $\sim$ 200 Hz. The raw data at this point contains an amount of information equivalent to 1.3Mb [7, p.14].

6.7 Lepton identification

Successful lepton identification and energy/momentum measurements will be paramount for the analysis.

In conclusion, the ATLAS detector uses a combination of tracking and calorimetry for electrons. The energy measurements of the tracking performs better than the calorimeter at low energies, but is overtaken at ~ 30 GeV. The electrons can be identified by the particle shower they deposit when interacting with the calorimeter. This can be combined with the inner detector information, e.g. the fraction of high threshold hits in the TRT, to give a cleaner sample. The tracking extends to $|\eta| = 2.47$ while the calorimeter extends to $|\eta| = 4.9$.
Muons are measured with a combination of inner detector tracking and the muon spectrometer. Muon identification tend to be more clean than electron identification, since the amount of background that permeates the muon spectrometer is smaller than the amount of backgrounds found in the calorimeter. The muon spectrometer covers the pseudorapidity range $|\eta| < 2.7$.
The ATLAS muon spectrometer does not perfectly cover $|\eta| < 0.1$. The analysis will make use of the characteristic narrow energy deposit of muons in the calorimeter systems to identify muons in this range. Muons that are identified in the muon spectrometer, but not in the inner detector tracking ('standalone muons'), $2.47 < |\eta| < 2.7$ will also be included.
The lepton identification is moreover improved by isolation criteria. In Jets and most decays containing leptons, the energy deposit in the calorimeter will be spread over a wider region than is the case for the final state $H \rightarrow ZZ^{(*)} \rightarrow 4l$ leptons. Likewise, Jets and decays will tend to have several tracks in a small region in the inner detector. By demanding that the electromagnetic shower energy from the lepton candidate is isolated and that its track is not accompanied by other high momentum tracks, the background rejection can greatly be improved. Figure 27 summarizes the electron and muon identification efficiencies.



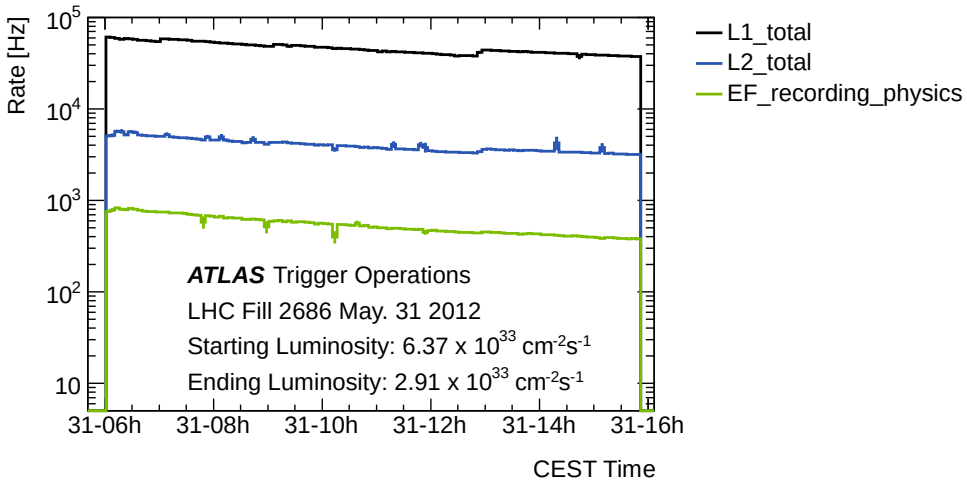


Figure 26: Top figure:Trigger rates for different Level 1 muon triggers. The L1_MU20 and L1_2MU6 which are activated by a 20 GeV muon or two 6 GeV muons respectively are similar to those used in the H→4l search. Bottom figure:Combined trigger rates for level 1, 2 and recorded event filter. It can be seen that the rate of the recorded event filter is around two orders of magnitude less that that of of the level 1 trigger.

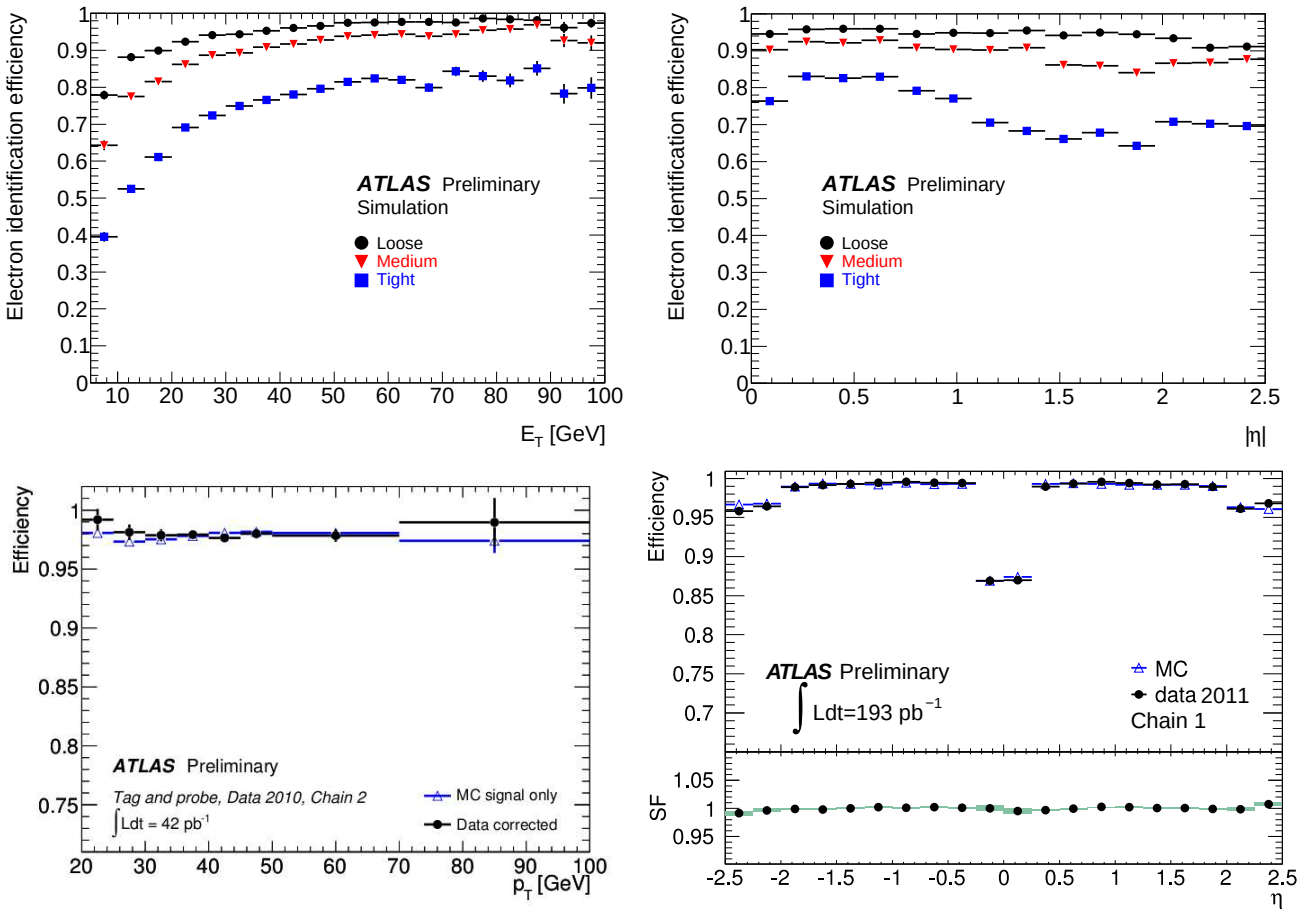


Figure 27: Top row: Electron identification efficiencies for electrons as a function of transverse energy and pseudorapidity [8]. The efficiencies are plotted for loose, medium and tight electron criteria. The loose criteria is based on the electromagnetic calorimeter shower shapes and energy leakage to the hadronic calorimeter. The medium selection combines the loose selection with tracking information, while the tight selection extends the medium by TRT requirements, and more stringent calorimeter cuts. It is furthermore demanded that the tight electron hits the innermost pixel layer to remove conversion electrons [6]. Bottom row: A data-Monte Carlo comparison of muon reconstruction efficiencies as function of transverse momentum and pseudorapidity, [10] (eta), [4] (pt).

Analysis

7 Overview

The Higgs to $ZZ^{(*)}$ diboson decay, in the four lepton final channel, is together with the diphoton decay channel what is going to lead the search of the Higgs boson.
For the four lepton search to be successful, as is the case with every analysis, it is paramount that all potential backgrounds are well understood, such that the size and purity of the signal can be maximized without introduction of unknown backgrounds. The work will take the Higgs work group 2 (HSG2) analysis and selection as a vantage point [11].
From this there are three distinct directions in which the search can be optimized, selection efficiency, acceptance and distinction from irreducible backgrounds.
Attempts to improve the selection efficiency and acceptance is described first, but the main focus on the analysis lies on the separation from the irreducible background.
The four lepton search channel can be subdivided into a 4μ, a $4e$ and $2e2\mu$, $2\mu2e$ channel. It is customary to write the lepton pair with invariant mass closest to the nominal Z mass first. The acceptance times efficiency for electrons are distinctly lower than for muons. For a Higgs signal of 125 GeV taken at a center of mass energy $\sqrt{s} = 7$ TeV the expected number of events for the three categories are 0.2 evt/fb^{-1}(4μ), 0.2 evt/fb^{-1}($2e2\mu/2\mu2e$) and 0.08 evt/fb^{-1}($4e$) [11, table.7]. The potential gain is thus much higher for electron optimization than for muons.
As described in the ATLAS section, one design feature of the transitional radiation tracker is electron identification. The efficiency optimization will try to make use of the TRT by lowering other requirements and imposing TRT identification instead. In subsequent times the TRT extension has been superseded by a similar approach that combines calorimeter shower shapes and inner detector silicon hits. The discovery of a Higgs-like particle at 125 GeV however has resulted in the need of electron identification at low transverse momenta, possibly giving the TRT extension a rejuvenation.
It is required in the selection that all electron candidates are within the inner detector coverage. The acceptance optimization therefore tried to use the LAr end-cap and LAr forward calorimeters, which extends to much higher pseudorapidity regions than does the tracker, to identify electrons. The lack of tracking from $|\eta| \geq 2.47$ reduces the background rejection considerably, and it is found that sophisticated methods for calorimeter electron identification are needed for a successful implementation.
The main background for $H \rightarrow ZZ^{(*)} \rightarrow 4l$ is the Standard Model production of a neutral current diboson pair $ZZ^{(*)}$ also decaying to four leptons [11, table.7]. The final state of these two processes are identical from a detector point of view, meaning that changing the selection will not alter the ratio of signal to background events in any notable way. The main focus of the analysis section will however be a discussion of a possible separation between these two types of events using slight differences of their production mechanisms and angular momenta.

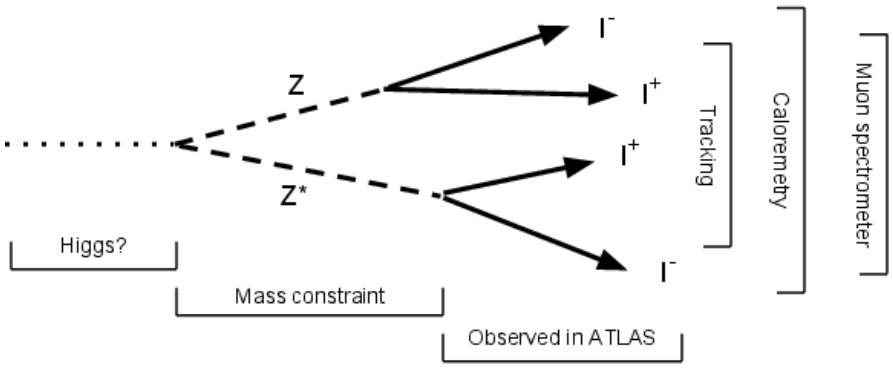


Figure 28: Conceptual drawing of Higgs decaying to a neutral current diboson in the four lepton final state. Also shown is the ranking of acceptance of used sub-detectors of ATLAS.

Figure 28 depicts two Z bosons decaying to two leptons each. The four leptons are identified by the ATLAS detector and can be combined by charge and lepton flavor to reconstruct the two Z boson masses. These can again be combined to reconstruct what could potentially be a Higgs boson.
There are extremely few things in the Standard Model, that will have four isolated leptons in its final state that are kinematically compatible with a Higgs boson. The potential background can be divided into categories. The true isolated four lepton final state only contains the Standard Model production of $ZZ^{(*)}$[16]. Next, the production of a single Z boson, together with something that fakes two leptons could be possible and will be the next greatest concern. Notably among these types of events is the Z$\rightarrow$ ll+b$\bar{\text{b}}$ production, where the bottom decay contains muons.
Finally a potential background is the t$\bar{\text{t}}$ decay, where each decay to a W boson and a quark and the W boson subsequently decays leptonically.
The left column of table 11 lists the cross section for each of these types of backgrounds. Some of these are orders of magnitude greater than that of the Higgs production. The selection has to be extremely well adapted to handle these kinds of events. The terminology used in the discussion is listed in table 9.

8 Event selection

The starting point of the event selection will be the recommended selection of the ATLAS Higgs working group 2 (HSG2), focusing on H$\rightarrow$ZZ$^{(*)}$ and Z(H$\rightarrow$ invisible), for summer 2012 [40]. The recommendations are then extended in several ways in an effort to improve the selection acceptance and efficiency, but it should be noted that some of these are at this

[16]Other things can have four isolated leptons as its final state, but these will not have momenta and combined dilepton masses that resembles a Higgs decay.

Name	Symbolic representation
Higgs candidate	H, 4l
On shell Z boson	Z-Z_{01}
Off shell Z boson	Z^* or Z_{23}
On shell Z leptons	l_0 or l_Z^-, l_1 or l_Z^+
Off shell Z leptons	l_2 or l_{Z*}^-, l_3 or l_{Z*}^+

Table 9: Summary of the terminology used in the analysis. l_0 and l_2 were chosen to represent the negatively charged leptons.

point irrelevant as the methods have been overtaken by similar approaches proposed by other members HSG2.
It has been chosen to restrict this analysis to potential Higgs masses in the range 110 GeV to 160 GeV, and all efficiencies and cross sections will be given correspondingly. Notice that invariant mass plots are depicted in the 100 GeV to 170 GeV mass window, such that unpredicted tail behavior of e.g. fits can be avoided.
The selection can be considered to consists of two types of requirements, a kinematical and a particle identification (PID).
The kinematical uses measured momenta, masses and angles to reject backgrounds. The PID uses detector information such as the electromagnetic calorimeter shower shapes or number of inner detector silicon hits to select electrons and muons and reject e.g. jets and photons.

8.1 The fiducial volume

In order to understand the magnitude of backgrounds a very crude analysis is performed which only ask that there are four lepton-like particles reconstructed in the detector, where tracking is possible, and that it has a minimal amount of transverse momentum. A set of cuts on the kinematics of the system, making an event to be Higgs-like is made. The total list of requirements can be seen in table 10, and can be considered to be the kinematical constraints imposed by the recommended HSG2 analysis.
The list of potential backgrounds used will be the one recommended by HSG2 [40, Summer 2012]. A list of the relevant ATLAS Monte Carlo samples and can be found in appendix A.
Notice that the Higgs$\rightarrow ZZ^{(*)} \rightarrow 4l$ samples include the tauon in the Z decays. For this reason all referred cross section times branching ratio of signal samples will also include τ decays. This will again be reflected in the listed selection acceptances. Cross sections and acceptances/efficiencies can be converted to only include electrons and muons by a multiplication of $\frac{4}{9}$ and $\frac{9}{4}$ respectively, while the expected number of events per integrated luminosity is of course the same.
Since there is an off change that two leptonically decaying τ will be used as either an electron or muon pair in the selection, it was believed this approach would reflect the real world more accurately.

It is moreover demanded that any lepton quadruple consist of two same flavor, opposite sign leptons pairs (SFOS), i.e. electrons with positrons and muons with anti-muons.

Kinematical constraints on Higgs candidates	
Muons	
Transverse momentum	$p_T \geq 6$ GeV
Pseudorapidity	$\lvert\eta\rvert \leq 2.7$
Calorimeter identified muons	
Transverse momentum	$p_T \geq 15$ GeV
Pseudorapidity	$\lvert\eta\rvert \leq 0.1$
Electrons	
Transverse momentum	$p_T \geq 7$ GeV
Pseudorapidity	$\lvert\eta\rvert \leq 2.47$
Quadruple	
Leading track	$p_T \geq 20$ GeV
Next to Leading track	$p_T \geq 15$ GeV
Next to next to leading track	$p_T \geq 10$ GeV
J/Ψ veto	No SFOS pair with $M_{ll} < 5$ GeV
Overlap removal: Same/opposite flavor	$\Delta R(l_i, l_j) > 0.10/0.20$
Z mass	$M_Z \in [50.0, 116.0]$ GeV
$Z^{(*)}$ mass	$M_{Z*} \in [17.5, 115.0]$ GeV
$Z^{(*)}$ mass (M_{4l} dependent)	$M_{Z*} \geq 17.5$ GeV $\leftrightarrow$ 50 GeV

Table 10: List of requirements an event must pass to be considered to be in the fiducial volume. The four lepton mass dependence on the $Z^{(*)}$ cut can be found in [40, Summer 2012, Secondary Di-Lepton Mass Cut]. For $M_{4l} = 125$ GeV it is 20 GeV.

Input	σ [fb]	Acc.	$N_{\mathrm{Exp}}/\mathrm{fb}^{-1}$
Z$\rightarrow$ ll + $\mathrm{N_{Jets}}$	$3.2{\cdot}10^6$	0.002847 $\pm$ 0.000019	9000 $\pm$ 60
ZZ$\rightarrow$ 4l	92	0.02936 $\pm$ 0.00019	2.7 $\pm$ 0.017
$\mathrm{t\bar{t}}$ (4l filter)	520	0.10087 $\pm$ 0.00055	52 $\pm$ 0.28
Z$\rightarrow$ ll + $\mathrm{b\bar{b}}$ + $\mathrm{N_{Jets}}$	160	0.04702 $\pm$ 0.00035	7.7 $\pm$ 0.057
H$\rightarrow$ ZZ $\rightarrow$ 4l (125 GeV)	4.8	0.1339 $\pm$ 0.0011	0.65 $\pm$ 0.0056
H$\rightarrow$ ZZ $\rightarrow$ 4l (150 GeV)	10	0.1883 $\pm$ 0.0015	1.9 $\pm$ 0.015

Table 11: Acceptance and cross sections for background samples with the same kinematical characteristics as a potential Higgs signal. Also listed is expected values for Higgs production at 125 GeV and 150 GeV. Notice that the listed uncertainties only reflect statistical uncertainties in the examined samples and do not express cross section uncertainties. Cross sections are from [40, Summer 2012, Useful cross sections]. The purpose of the further analysis is to use the ATLAS PID to reduce backgrounds to a compatible size with the Higgs expectations.

Table 11 and figure 29(left) shows the acceptance of potential Higgs signals and other backgrounds, given the kinematical configuration listed in table 10. It is seen, that even though it is difficult to land in this configuration, a selection has to be made that maximizes jet rejection.

8.2 HSG2 Requirements (Summer 2012)

The H$\rightarrow$ZZ$^* \rightarrow 4l$ selection and data format has been continuously updated until the particle state discovery in July 2012. The selection used in the analysis has therefore highly been inspired by the HSG2 selection, but is cannot be considered a complete copy.
To reduce the number of accepted Jets, the kinematical requirements are extended in a number of ways.

8.2.1 Preselection

Before anything else was examined, it was demanded that any event should contain a vertex with at least three associated tracks. Various triggers were then imposed. It was demanded that an event should contain either of; one 10 GeV electron and two 6 GeV muons, one 10 GeV electron and one 10 GeV muon, one 22 GeV electron, two 12 GeV electrons, two 10 GeV muons, or one 18 GeV muon.
Events were furthermore removed if it was LAr error flagged.

8.2.2 Muon selection

The muon selection allowed three different types of muons, combined, stand alone and calorimeter identified muons (Calo muons).

The combined muons are muons both reconstructed in the inner detector and muon spectrometer. For these it was demanded that they were associated with the primary vertex (the vertex with the highest scalar sum of momenta), and that they made a a number of hits in the inner detector.
The stand alone muons are muons only reconstructed in the muon spectrometer. A muon was only allowed to be stand alone if it was not in the inner detector coverage. For these it was demanded that they made a number of hits in the muon spectrometer.
The calorimeter identified muons are not reconstructed in the muon spectrometer. The muon spectrometer does not cover $|\eta| < 0.1$. It was therefore allowed for muons to be identified using the inner detector and calorimeter, if their pseudorapidity were in the spectrometer crack. The same requirements for inner detector hits and distance to the primary vertex was imposed on the calorimeter identified muons as the combined muons. Table 12 lists all cuts imposed on the muon selection.

8.2.3 Electrons

The selected electrons were required to be flagged as loose++. The loose++ flag is based on EM shower shapes and hadronic energy leakage an demands at least 1 pixel hit and 7 silicon hits. The pixel and silicon tracks are demanded to match the shower in the η direction ($|\Delta\eta| < 0.015$) [17, sect.:Loose++ (2011 and 2012)]. This gives a very good electron efficiency and background rejection.
The distance to the primary vertex was required to be less than 10 mm in the longitudinal direction and less than 6.5 standard deviations of the transverse distance measurement. Table 12 lists all cuts imposed on the electron selection.

8.2.4 Overlap removal

If a number of inner detector hits were reconstructed as both an electron and a muon, the electron was discarded. In the case where two reconstructed electrons shared the same inner detector track, the highest E_T electron was used in the further selection.

8.2.5 Quadruple selection

The selected leptons were finally combined into quadruples containing two same flavor, opposite sign lepton pairs, where the kinematical constraints described above were imposed.
In the case where the quadruple contained two muons, it was also required that it had only one stand alone or one calorimeter identified muon. Tracking and calorimeter isolation was then imposed. The isolation sums all momentum or energy in a cone ($\Delta R = 0.20$) around the lepton track as measured either by tracking or calorimetry. If the scalar summed momentum was 15% of the lepton momentum the lepton was not considered isolated. The corresponding number for energy was 30%.
If any of the other leptons were found inside the ΔR cone, their momentum or energy contribution were subtracted, explaining why isolation could not be imposed on the single lepton

Detector constraints on lepton candidates	
Combined muons	
Transverse distance to primary vertex	$d_0 \leq 1$mm
	$d_0 \leq 3.5 \cdot \sigma_{d_0}$
Longitudinal distance to primary vertex.	$z_0 \leq 10$mm
Muon reconstruction algorithm	Author = 6,7
Innermost pixel hit, if expected.	$N_{B-layer} \geq 1$
Inner detector hits.	$N_{pixel} + N_{pixel-deadsensors} \geq 2$
	$N_{SCT} + N_{SCT-deadsensors} \geq 6$
	$N_{Pixel-holes} + N_{SCT-holes} \leq 3$
TRT hits, if muon is in appropriate η.	If $\lvert\eta\rvert < 1.9 : N_{TRT} \geq 6$
Only 90% TRT hits are outliers.	$N_{TRT-Outliers} \leq 0.9 N_{TRT}$
Stand alone muons	
Muon reconstruction algorithm	Author = 6
Only accepts stand alone if muon is not in inner detector coverage.	$\lvert\eta\rvert \geq 2.5$
Muon spectrometer hits.	$N_{CSC-Eta} + N_{CSC-Phi} > 0$
	$N_{MDT-EM} + N_{MDT-EO} > 0$
Calorimeter identified muons	
Only if in muon spectrometer crack.	$\lvert\eta\rvert \leq 0.1$
Muon reconstruction algorithm	Author = 16
Multivariate calorimeter muon identification.	Calo muon likelihood > 0.9
Same inner detector requirements as combined muons	
Electrons	
Inner detector hits and EM cluster shape	loose++
Object quality	goodOQ
Transverse distance to primary vertex.	$d_0 \leq 6.5 \cdot \sigma_{d_0}$
Longitudinal distance to primary vertex.	$z_0 \leq 10$mm

Table 12: List of requirements leptons must pass to be used in a Higgs candidate

Quadruple selection	
Track isolation	All leptons should be isolated in tracking s.t. its p_T contribute 85% of the total p_T in an ΔR-cone of 0.20.
Calorimeter isolation	Same as tracking isolation, but here track E_T should contribute 70% of the total E_T in an ΔR-cone of 0.20.
Isolation overlap	Energy/momentum contribution from other leptons are removed in isolation.
Stand alone and Calo muons	Any quadruple is only allowed to contain one stand alone or one calo muon.

Table 13: List of requirements quadruples of leptons must pass in order to be considered an Higgs candidate.

Input	σ [fb]	Acc.· Eff.	$N_{\text{Exp}}/\text{fb}^{-1}$
Z→ ll + N_{Jets}	$3.2 \cdot 10^6$	0.00000014 ± 0.00000013	0.43 ± 0.42
ZZ→ 4l	92	0.01218 ± 0.00012	1.1 ± 0.012
$t\bar{t}$ (4l filter)	520	0.000067 ± 0.000015	0.034 ± 0.0077
Z→ ll + $b\bar{b}$ + N_{Jets}	160	0.000411 ± 0.000034	0.067 ± 0.0055
H→ ZZ → 4l (125 GeV)	4.8	0.0992 ± 0.0010	0.48 ± 0.0049
H→ ZZ → 4l (150 GeV)	10	0.1579 ± 0.0014	1.6 ± 0.015

Table 14: Backgrounds and potential signals expectations after HSG2 recommended requirements. Notice the extreme efficiency by which the Z+Jets has been reduced as compared to the fiducial volume acceptance. Notice also that the listed uncertainties only reflect statistical uncertainties in the examined samples and do not express cross section uncertainties. [40, Summer 2012, Useful cross sections]

selection. The quadruple selection is listed in table 13.

If several quadruples passed the requirements, the one with reconstructed Z masses closest to the nominal Z mass was kept. Applying the selection on the simulated background and signal samples resulted in table 14, where it is seen that the main concern is the Standard Model production of $ZZ^{(*)}$.
The resulting expected invariant mass distributions can be seen in figure 29.

8.3 Selection extension 1: Loose + TRT

To improve the electron efficiency it was examined if the transitional radiation tracker could improve the selection. At the time this was tested, the electron selection demanded medium reconstructed leptons instead loose++.
This was realized by allowing one electron to be loose if it also behaved electron-like in the TRT. The extension is motivated by figure 27, where it is seen that the loose and medium selection efficiencies are $\sim 95\%$ and $\sim 90\%$ respectively for any given η, while the identification probability drops significantly at lower p_T.

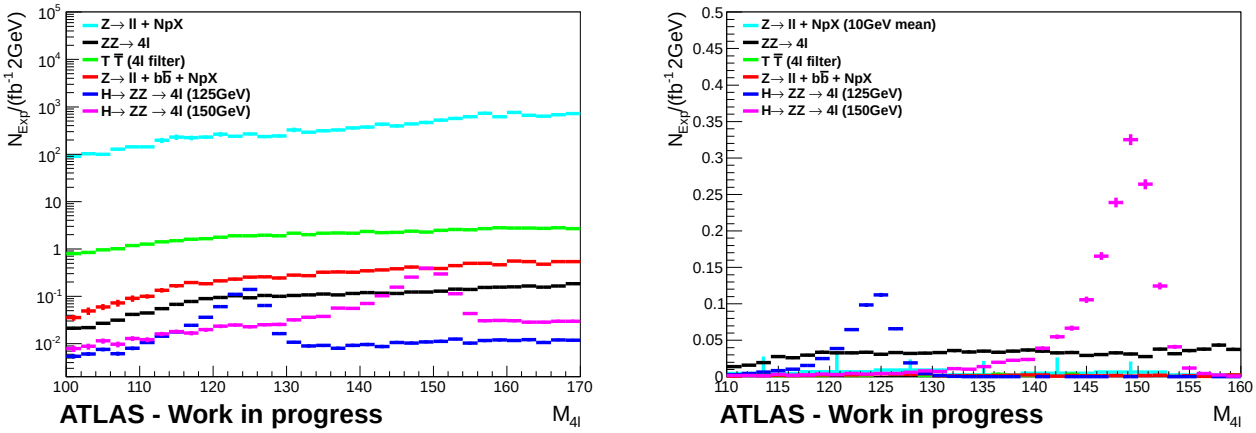


Figure 29: Signal and background mass distributions. Left: Mass distributions for all events passing kinematical requirements to be considered Higgs-like. Right: Mass distribution for all events passing full HSG2 selection as described in [40, Summer 2012]

The loose selection criteria is based on the middle layer of the electromagnetic calorimeter, where cuts are applied on the lateral shower width and shape. Moreover a cut is applied on leakage to the hadronic calorimeter. This gives a very high efficiency but also a relatively low background rejection[16, sect.1.3].
The medium criteria extends the loose, looking for second energy maxima around the calorimeter cell with the highest energy deposit. The search window is $\Delta\eta \otimes \Delta\phi = 0.125 \otimes 0.2$. If the energy distribution contains a second maximum, this is an indication of a $\pi^0 \rightarrow \gamma\gamma$ decay. Again requirements on the electromagnetic shower shape are imposed together with pixel hits, silicon hits and transverse impact parameter. [16, sect.1.3]. Note that in general will any medium electron automatically also be loose.
It is also worth mentioning that it is not of great interest to tighten the selection, since the dominant background contains the same particles as does the signal, figure 29 (right). The object is solely to loosen the selection, allowing more signal, but still rejecting the reducible backgrounds to the same degree.
As described in the ATLAS section one of the design features of the TRT was electron identification. The first thing to look into is how the ratio of the number of high threshold hits to the total of number TRT hits (fHT) distributes for electrons and hadronic tracks. Figure 30 (left) shows a clear separation in the two distributions, and that a cut of $\sim$0.15 seems appropriate.

The right plot in figure 30 show the $Z \rightarrow ee$ peak, where one electron is identified according to the standard selection (here meaning medium), while the type of the other is indicated in the legend. It is evident that a $\sim$10% gain could be obtained. The selection used for this plot only makes requirements on the individual leptons, similar to those found in table 12.
It is estimated that the loose++ identification efficiency is similar to the loose, while its rejection is higher. The change from medium to loose++ thus made this simple extension all but obsolete, figure 27. If time allowed it could however be interesting to allow the TRT

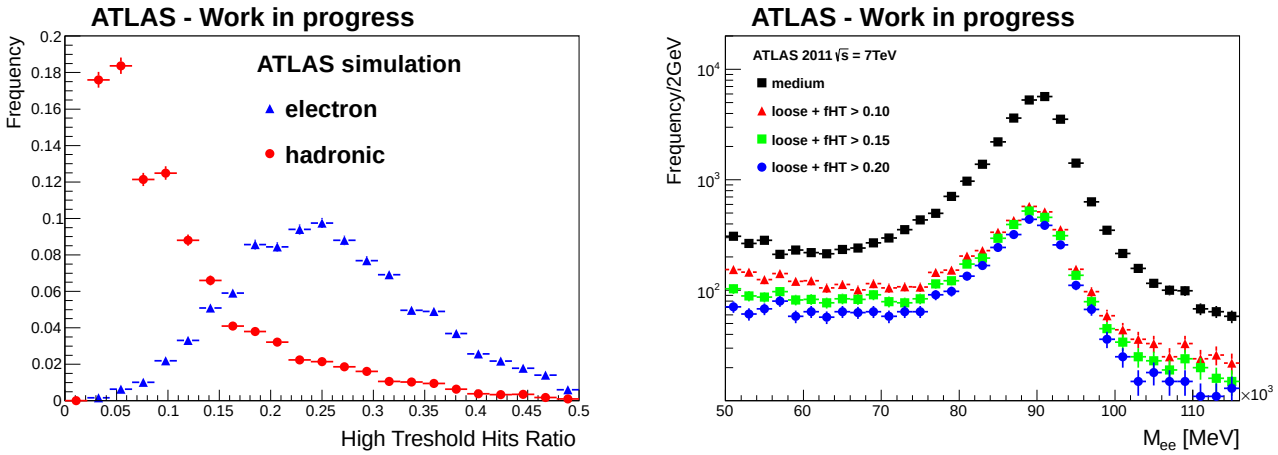


Figure 30: Left: Fraction of high threshold hits to the total number of TRT hits for electrons and hadronic tracks. Derived from ATLAS Z→ee+N_{Jets} simulation, where Z electrons are used as signal, and objects identified with a high p_T quark within a ΔR cone of 0.05 has been used for the hadronic distribution. Right: Z mass peak for medium selection and loose + fraction of high threshold hits. Plot is produced with 2011 data from ATLAS and indicates that a ∼10% gain in electron efficiency could be obtained this way without raising the signal to background ratio notably.

identified electron to have lower momentum.
The extension was appended for electrons of both the on-shell and off-shell Z boson but having applied the loose++ selection, the resulting figures were statistically indistinctable from those of table 14.

8.3.1 Systematic uncertainties

Due to the extension being obsolete it was not considered reasonable to assign to much work to the evaluation of systematic uncertainties. If however a study of systematics was to be conducted, it could be done by examining the effect on the extremely large Z→ee data sample that has been collected. This way dependencies of e.g. pileup, can be tested on real data, reducing the amount of assumptions going into the analysis, as compared to the simulation based.

Figure 31 shows the Z→ee mass spectrum, where one electron is demanded to have fHt > 0.10, while the other is selected to be medium. All other requirements are equal to the ones found in table 12. The distribution is plotted for collisions with less that 7 vertices and more than 10. Both the width of the peak changes as well as the signal to total number of events.

8.4 Selection extension 2: End cap and Forward electrons

A way to extend the acceptance of the analysis is to expand the pseudorapidity range in which electrons are accepted. The primary analysis is only using electrons in the pseudorapidity

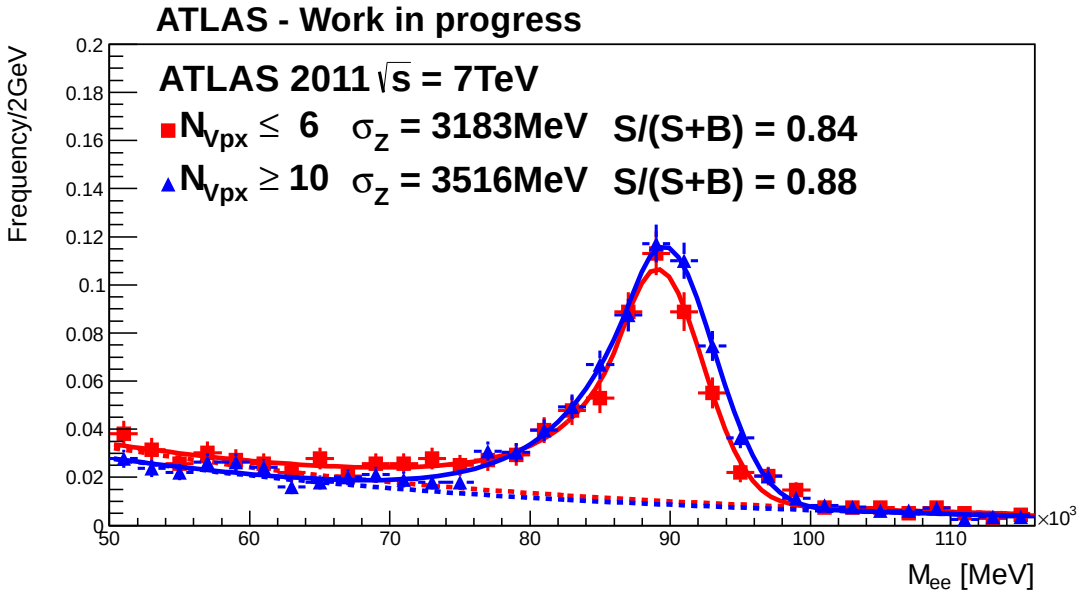


Figure 31: Pileup dependence of $Z \rightarrow ee$ mass spectrum with fHt > 0.10 for one electron. The distribution changes somewhat with pileup, an effect that would have to be included in the systematic uncertainty. The signal contribution to the total integral has been calculated between 81 GeV and 101 GeV.

coverage of the inner detector, table 10. It is however possible to identify electrons using the calorimeter, by examining the shape of the deposited shower. More precisely the calorimeter is able to separate electrons and photons from things that are not electrons and photons by geometrical features of the shower, how the energy density evolves as function of depth and energy leakage to the hadronic calorimeter. By only allowing one electron from the on-shell Z boson to be beyond the inner detector the restriction of the mass window reduces the amount of background vastly.
Figure 32 (left) shows the expected pseudorapidity distribution for electrons coming from a Higgs boson through the on-shell Z boson.

If the same electron efficiency and background rejection could be obtained for the calorimeter as the tracking calorimeter combination it would be possible to increase the Higgs acceptance by 23% taking η correlations into account. This, of course, is to idealistic.
The lack of tracking reduces the jet rejection and makes electron photon separation virtually non existent. In order to understand the detector response to backgrounds it is again necessary to examine data rather than simulation. This could be done by the tag and probe method, figure 32 (right). In figure 33 a Z is identified by an electron passing the tight requirements (the tag electron) and a reconstructed electron in the non-central region passing the loose requirements. For comparison the distribution for Z candidates where both electrons are central is also shown. To reduce backgrounds further, it was also demanded that the tracks have a transverse momentum of 20 GeV and that they are isolated $\sum_{i \in \Delta R=0.2} E_T^i < 0.2 \cdot E_T^e$.

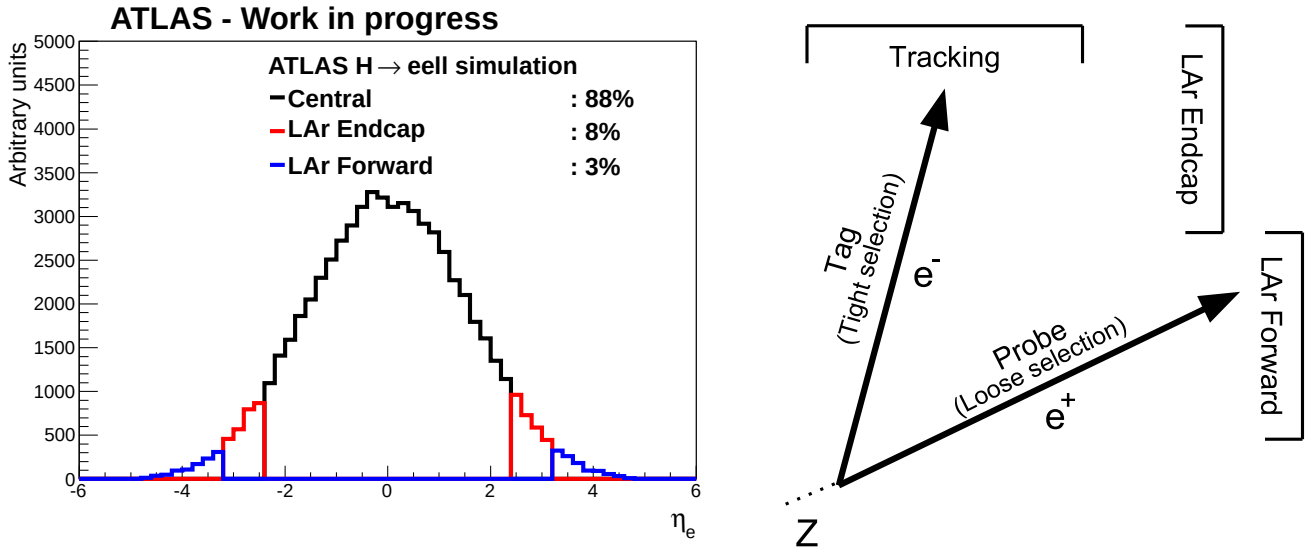


Figure 32: Left: Simulated electron pseudorapidity for Higgs event. In black: Electrons that hits the part of ATLAS covered by tracking. In blue and red are seen the η regions covered by the liquid Argon end-cap and forward calorimeter respectively. Right: Tag and probe. The tag electron passing tight requirements is identified using tracking and calorimetry. The properties on the probe electron can be examined while signal to background estimations can be derived from their invariant mass spectrum.

It is evident that the simple selection of forward electrons described here accepts to much background to be acceptable. It does however not represent the full potential of the forward calorimeters, since only a crude cut on the shower shape and a single isolation parameter has been used. In the following section, a more elaborate selection will be attempted.

8.5 Selection extension 3: Electron MVA

It was discussed above that muon selection performs significantly better than the electron selection. The list of things that is able to fake electron behavior is larger than that of muons. It is hence worth investigating if a more advanced selection could be derived.
The aim is to train a boosted decision tree on simulated Monte Carlo samples, and evaluate the performance on real Z→ee data, eliminating the introduced error from the detector simulation. Energy deposits and tracks ΔR matched with a high p_T quark at simulation truth level, was used for training backgrounds from Z→ee+$b\bar{b}$ and $t\bar{t}$, while Z→ee simulation electrons provided signal distributions. Both have very high statistics.
The way to proceed was to divide the electron identification into several parts. In the following these are; tracking, calorimeter energy and cluster shape, calorimeter isolation and track-calorimeter matching, table 15.

Furthermore, based on considerations of the detector design, the samples were subdivided

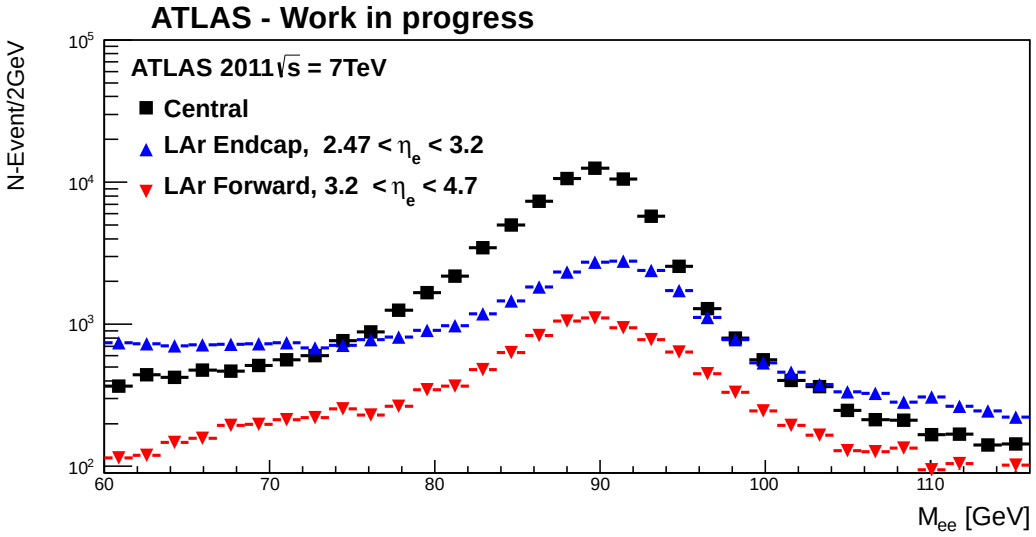


Figure 33: Z→ee invariant mass for central-central, central-LAr end cap and central-LAr forward electron configurations. It is clearly seen from comparing the sidebands $\lesssim 75$ GeV and $\gtrsim 105$ GeV to the center that the relative background rejection is worse in the LAr end cap and LAr forward regions.

Electron MVA categories	
Category	Related variables
Tracking	Pixel, SCT, TRT hits and TRT hT fraction, trk. isolation.
Calorimeter	Energy deposit in different LAr layers, cluster shape.
Calorimeter isolation	Energy deposit in ΔR cones of 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, relative to cluster energy.
Track-Calorimeter matching	$p_{T-\text{Track}}/E_{T-\text{Calo}}$, $\Delta\eta_{\text{Track}-\text{Calo}}$, $\Delta\phi_{\text{Track}-\text{Calo}}$ (corrected for curvature and brehmstralung)

Table 15: List of categories and related variables used in the electron BDT

into η regions, table 16. Taking into account that tracking does not extend to the furthest regions, this gives a total of 24 different trainings.

Figure 34 shows the resulting output for the four different categories in the first η region; 0.0 to 0.9. It is seen that separations defined as:

$$\frac{1}{2}\int dx \frac{(p_{sig}(x) - p_{bkg}(x))^2}{p_{sig}(x) + p_{bkg}(x)}, \tag{8.1}$$

between 0.36 and 0.76 are obtained[17]. After training it is desired to use the resulting BDT

[17]For comparison, two perfectly separated distributions have a separation of 1, while two Gaussian distributions with equal widths (σ) and means distanced 1σ apart have a separation of 0.2

Electron MVA η regions			
$\|\eta\|$		Contains (Barrel region)	Contains (End cap region)
0.00	0.90	Pixel, SCT, TRT, LAr	
0.90	1.37	Pixel, SCT, LAr	TRT
1.37	1.52	Pixel, SCT	TRT, LAr
1.52	1.95	Pixel	SCT, TRT, LAr
1.95	2.47		Pixel, SCT, LAr
2.47	3.20		LAr
3.20	4.90		Forward Cal

Table 16: List of η regions used in the electron multivariate analysis. Notice that the forward most regions do not contain tracking.

structures on real data to evaluate its performance. This could be done by the tag and probe method from the large Z→ee statistics. Requiring that the probe electron is electron-like in three of the four categories, should ensure that a very clean signal is obtained, a claim that can be examined by studying the invariant mass spectrum, much as in figure 31. The electron candidate score can then be examined in the final of the four categories. Conversely, objects that are not electrons can be identified by using the same method, and by possibly demanding that the fake Z candidate mass does not match nominal Z mass expectations. Conservation of charge can moreover also be used, in the η regions with tracking, to match opposite charged electron candidates for obtaining a pure Z signal, and same charged electron candidates for estimating backgrounds.

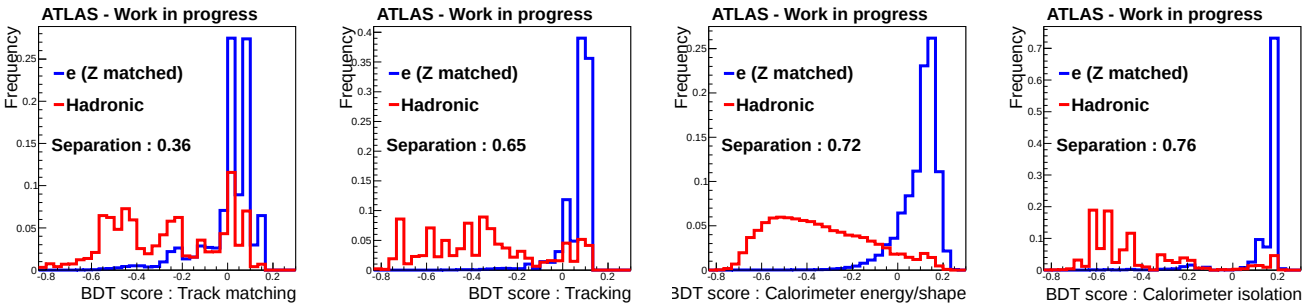


Figure 34: Separation of different electron MVA categories, for $0.0 \leq \eta \leq 0.9$. From left to right are separations obtained by, track matching, tracking, calorimeter energy deposit in different layers and calorimeter shower shape, and calorimeter isolation.

This method of tag and probe will only work if the BDT scores of the individual categories can be considered uncorrelated. The linear correlation factors for electrons and backgrounds can be seen in table 17 for the central most η region. It is assumed that the simulation to a reasonable degree will represent the correlation factors, and it is seen that this direct approach will not allow one to study the BDT behavior using the tag and probe method, which would render the implementation on real data questionable.

	Trk match		Tracking		Calo E/shape		Calo iso	
	e (Z)	Had.	e (Z)	Had.	e (Z)	Had.	e (Z)	Had.
Trk match	1.00	1.00	0.63	0.17	0.20	0.76	0.66	0.51
Tracking	0.63	0.17	1.00	1.00	0.33	0.14	0.55	0.16
Calo E/shape	0.20	0.76	0.33	0.14	1.00	1.00	0.25	0.56
Calo iso	0.66	0.51	0.55	0.16	0.25	0.56	1.00	1.00

Table 17: Linear correlation factors for BDT scores of the different electron identifier categories. The values are very different incompatible with zero correlation, rendering the Tag and Probe method for evaluation on data problematic.

It is still possible however that an uncorrelated set of boosted decision trees can be constructed using a more elaborate set of variables. This would require a very deep understanding of all measured effects, and for the time being it was deemed to great a challenge, given the time perspective, to continue in this direction.

9 Separation from the Standard Model ZZ background

According to the Standard Model it is possible to produce a ZZ diboson pair without an intermediate Higgs boson. As can be seen from figure 29 (right) this is what constitutes the largest background to the production of a Higgs boson in the relevant decay channel. It would seem that these have the exact same properties as a pair coming from a Higgs Decay. It turns out however that ZZ bosons coming from a Higgs will have their spin correlated in a way that is not present for the Standard Model production. Moreover the production mechanisms are different for the two types of events. To understand this, one can examine the Feynman diagrams giving the largest contribution to their respective cross sections. For the Standard Model this is the quark exchange, and for Higgs production it is the gluon fusion as seen in figure 36.

While the total angular momentum of the Standard Model ZZ* production is not confined to a single value, the mediation of a Higgs boson ensures that the system is produced with zero angular momentum. The idea of the analysis is thus to find as many variables as possible, that expresses both the differences in production mechanisms and the differences in angular momenta. These can then be used as input in a multivariate variate separation that accumulates the discrepancies.
The strategy of this analysis is to only consider the Standard Model ZZ production as background when choosing variables and training the boosted decision tree. Behavior of other backgrounds will be evaluated afterwards. This choice has been made both because the Standard Model ZZ* production dominates the background at the considered masses. But also because the statistics size of the other backgrounds samples, that passes the Higgs selection is very small in comparison to the ZZ*, resulting in a method very sensitive to overtraining.

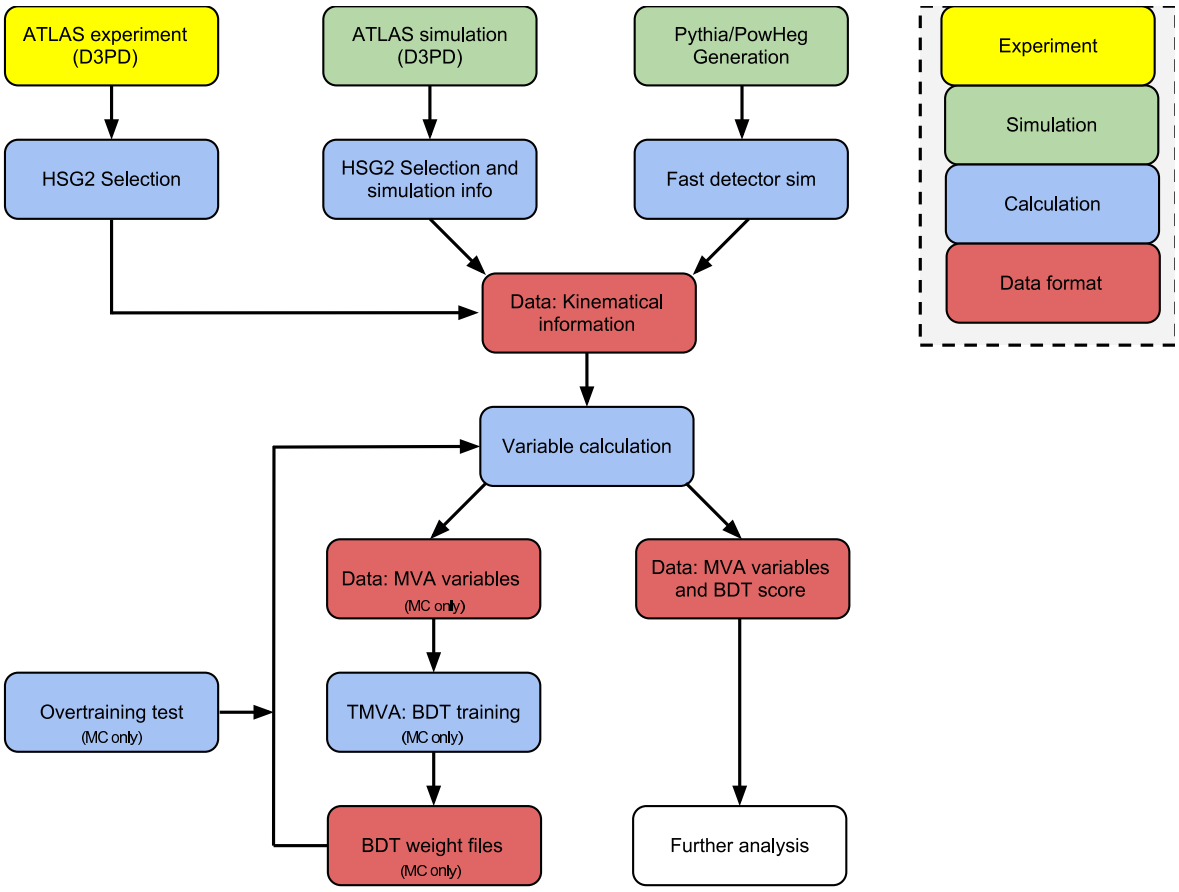


Figure 35: Flow chart depicting the progress of the multivariate analysis. Three different kinds of input are examined; ATLAS simulation containing a very precise description of the detector, a fast simulation where an effective detector model is used and experimental data. Both the ATLAS simulation and the fast simulation are used to train a boosted decision tree. The fastsim training is used to evaluate theoretical and systematical uncertainties, while the ATLAS simulation of signal and backgrounds is compared to the experimental data.

If on the other hand the trained BDT is taken as the expression for the best possible separation between Higgs and Standard Model $ZZ^* \rightarrow 4l$ production, behavior of other backgrounds can be evaluated in the BDT structure and their corresponding score distribution implemented in the further analysis.

Figure 35 summarizes the various steps performed in the analysis. The HSG2 selection is imposed on the Monte Carlo simulations and data, and the analysis variables are calculated from the resulting events. A boosted decision tree is subsequently trained on the simulation, whose structure is saved after being tested for overtraining. The BDT distributions for the selected events are then calculated and added to the list of variables for both data and simulation, before being implemented in the further analysis.

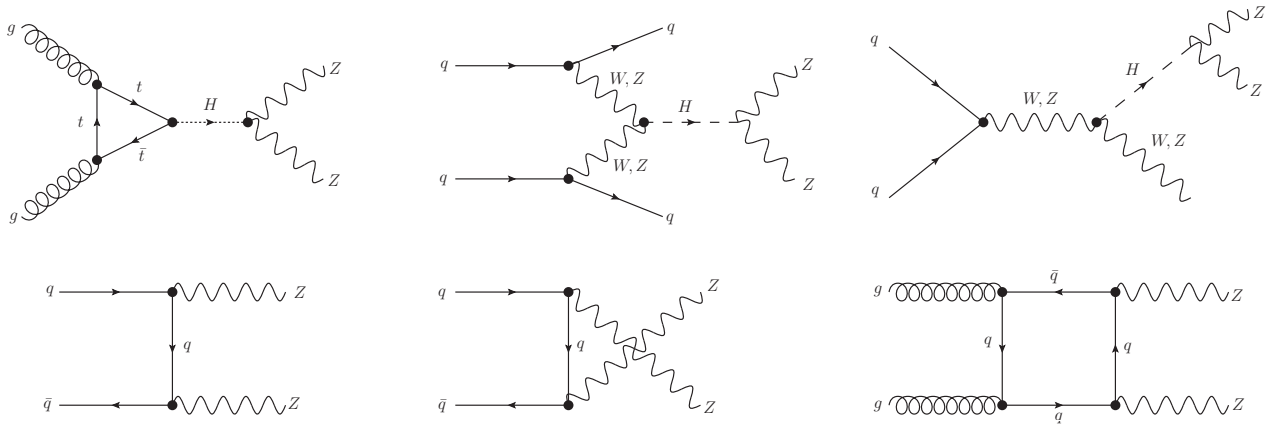


Figure 36: Top row: Higgs production. The rightmost diagrams represents gluon fusion, constituting ∼90% of the total cross section for a Higgs boson of 125 GeV. The center diagram shows vector boson fusion ∼7%, while the rightmost depicts Higgs strahung ∼3% [28]. Bottom row: Standard Model production of ZZ diboson without an intermediate Higgs boson. The two leftmost diagrams constitutes the leading order contributions. Simulations of the rightmost production mode were not obtained and was therefore not included in the analysis.

9.1 Variables

For this analysis a large list variables have been assembled that expresses differences between the Standard Model production of neutral current dibosons and Higgs production decaying to the same. These can be divided into three categories, production related, Z mass related and decay related. A summary of the variables found that expresses differences in the two types of events can be found in table 18.

It is worth noting that many of these can be expressed as functions of other variables, and that the separation power of each varies a lot. Also due to limited statistics of the simulated samples on which this will be trained the list of variables were reduced to limit complexity and not contain highly correlated parameters.
The final choice of variables should then omit weak variables and only contain a set were none of the variables indirectly can be approximated by the others.
There are a lot of different conflicting aspects when choosing variables. Intuitively a better separation could be obtained the more variables with separation power the BDT contains. But since statistics are finite, the separation of stronger variables is wasted because a more complex BDT is needed the more variables is used as input. If the tree is too complex relative to the size of the training samples it will be overtrained, making its predictions biased. The complexity of the BDT can also be kept down by ensuring that the set of input variables are as little correlated as possible. The simplest example of this is if two variables are completely uncorrelated. Here a cut on each variable would suffice, corresponding to two trees each containing one node. If all variables used were completely uncorrelated, or only linearly correlated, the use of a BDT would be unnecessary and simpler MVAs could be used.

Title		Separation
Production related		
Transverse momentum of candidate	p_T	0.08
Acoplanarity	$\Delta\phi_{ZZ*}$	0.10
Rapidity of four lepton system	y_H	0.02
Pseudorapidity of four lepton system	η_H	0.10
Z mass related		
Z/Z* masses	M_Z, M_{Z*}	0.01/0.02
Relative Z/Z* masses	M_Z/M_{4l}, M_{Z*}/M_{4l}	0.10/0.03
Phase space/relative difference	$1-(M_Z \pm M_{Z^*})/M_{4l}$	0.10/0.05
Decay related		
Angle between Z and Higgs direction	$\cos\theta_{HZ}$	0.01
Angle between l$^-$ and Z direction	$\cos\theta_{Zl0}$	0.01
Angle between l$^-$ and Z* direction	$\cos\theta_{Z*l2}$	0.01
Angle between decay planes of Z bosons	χ	>0.01

Table 18: List of potential variables found that could be implemented in the multivariate analysis. Notice that competing variables, indicated by being clustered together, are highly correlated and only one of each group should be implemented. The variables were chosen, as discussed in the text, are underlined. The separation power, $\frac{1}{2}\int dx \frac{(p_{sig}(x)-p_{bkg}(x))^2}{p_{sig}(x)+p_{bkg}(x)}$, of each individual variable is also listed. This value does not take correlations into account and therefore only serves as a crude estimation of the variable power when used in the BDT.

There does not exist a definite answer to how the choice of variables should be made but for the most part it will here be based on what gives the best performance, and what is most protected against overtraining.
All discussion below is based on the ATLAS simulations found in appendix A.

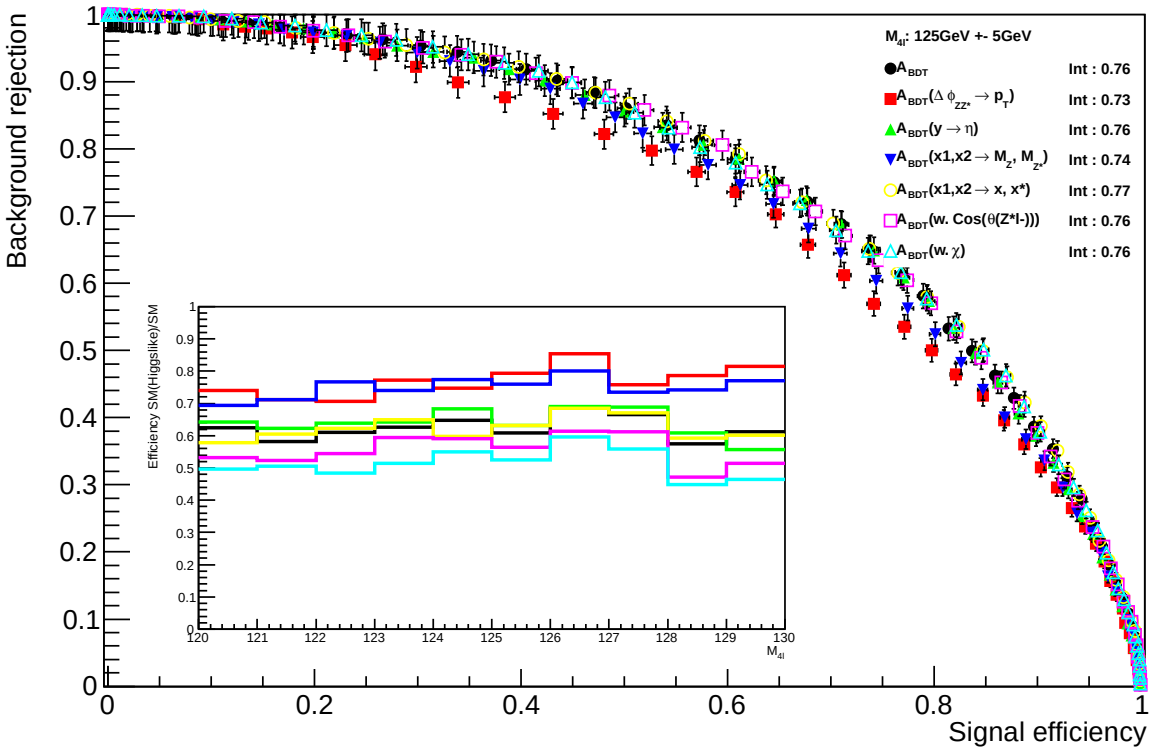


Figure 37: ROC curves and corresponding integrals (Int.) for different variable sets. The result is very stable. Insert: Number of Higgs like neutral current diboson pairs divided by total number of neutral current diboson pairs as function of mass. An event is considered Higgs like if its BDT score is above where the maximum of $S/\sqrt{S+B}$ is found. This serves as a crosscheck to see if the trained BDT selects a preferred mass for background events.

9.1.1 Transverse momentum of four lepton system and Acoplanarity.

The transverse momentum of the diboson system will be different such that the production of a Higgs boson predicts a higher value, than in the Standard Model scenario. Strongly related to this is the acoplanarity ($\Delta\phi_{ZZ*}$) of the system, which measures the angle between the two reconstructed Z bosons in the transverse plane of the ATLAS detector, figure 38. Considerations from special relativity suggest that for given values of $\cos(\theta_{HZ})$, p_T and $\Delta\phi_{ZZ*}$ will be strongly correlated, and hence only one will be used. This is because the opening angle

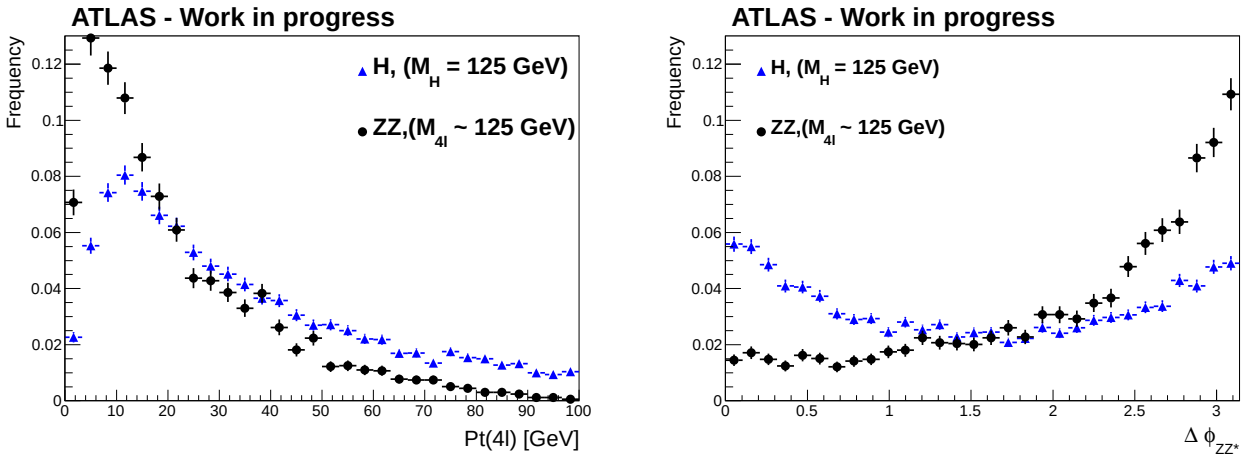


Figure 38: Higgs and neutral current spectra for transverse momentum (left) and acoplanarity (right).

between the Z bosons is π in the Higgs frame of reference. Given a boost of $(M_{\mathrm{H}}, p_{T-\mathrm{H}})$ the opening angle in the ATLAS frame of reference will only depend on the angle between the boost vector and the direction of decay of the two Z bosons. It should however be noted that $\Delta\phi_{ZZ*}$ only considers the transverse plane, while $\cos(\theta_{HZ})$ is measured in three dimensional space, such that a bijective map from two of the parameters to one can not be constructed. On the other hand, there is little that suggest that the extra information in how the H→ZZ decay plane is oriented compared to the ϕ plane contains separation power.

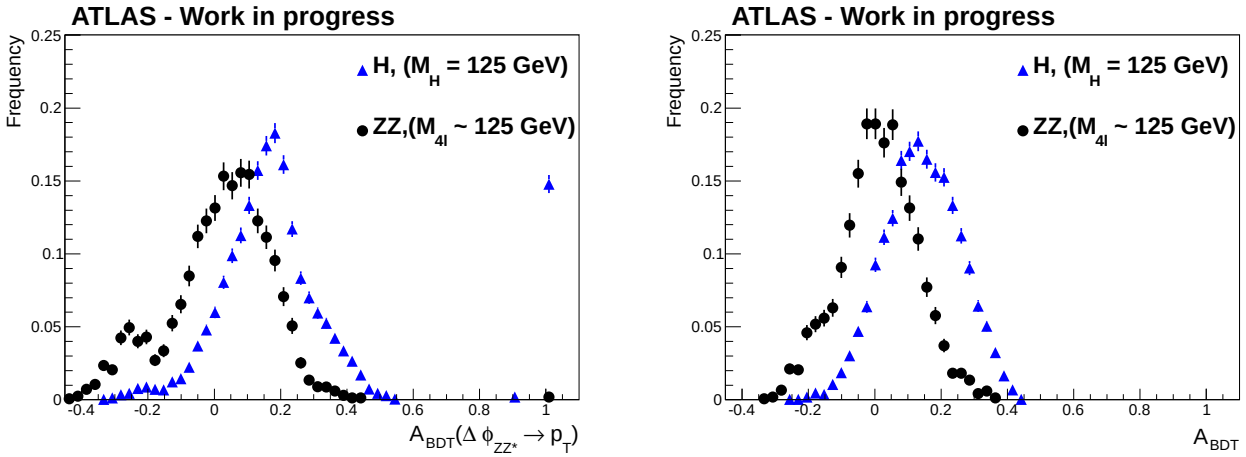


Figure 39: BDT scores when using p_T(left) and $\Delta\phi_{ZZ*}$(right) as input. Illustrates problem with using unbounded distributions as input for the BDT. The far right peak in the p_T case corresponds to high transverse momenta.

As can be seen from figure 37(left) the pt_T performance is slightly worse. An interesting

feature general for BDT training can moreover be seen when comparing the BDT PDFs for training with p_T and $\Delta\phi_{ZZ*}$ respectively, keeping all other things equal. From figure 38 it is seen that $\Delta\phi_{ZZ*}$ is bounded in a well defined interval, ensuring that both signal and background samples contain information for all values, this not being the case for high transverse momentum background in the p_T training. The result is the undesirable peak to the right of the p_T trained BDT in figure 39, corresponding to p_T values greater $\gtrsim$100 GeV. It seems untrustworthy that any four lepton event with $p_T \gtrsim 100$ GeV would return a likelihood around 99% of being signal [18].
In the training test sample comparison it was also seen that using $\Delta\phi_{ZZ*}$ results in a better agreement, indicating that the BDT finds that a simpler tree will give the same separation with less complexity.

9.1.2 Rapidity and Pseudorapidity

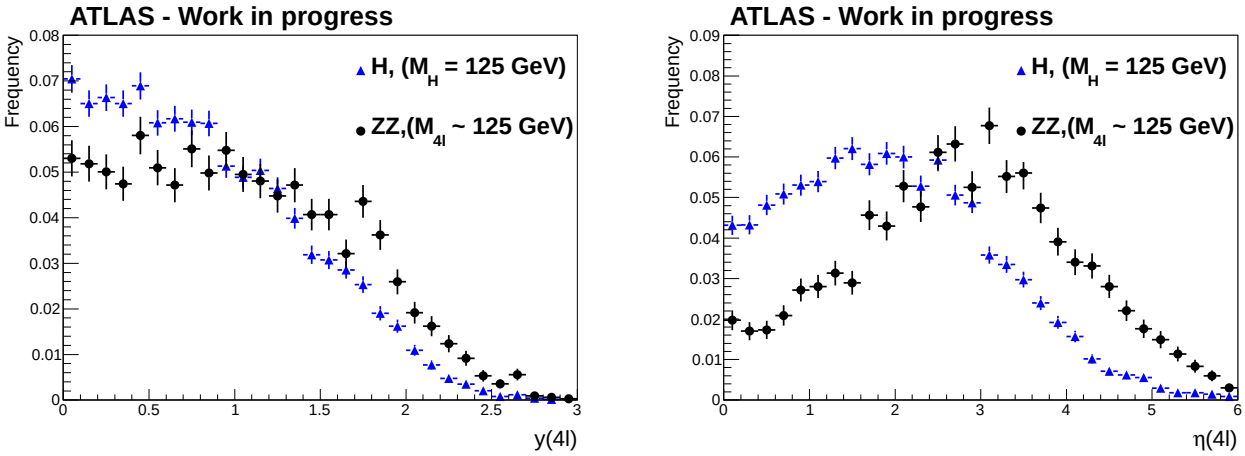


Figure 40: Higgs and neutral current distributions for transverse rapidity (left) and pseudo rapidity (right)

The probability density function for the rapidity of the signal (Higgs) and background (SM) expresses a difference in the production mechanism. To see why this is the case, note that the leading order ZZ production of the Standard Model is produced by a quark-antiquark while Higgs bosons will mainly be produced by gluon fusion.
The rapidity dependence on the difference in parton momentum fraction is given by [27, eq.3.7]:

$$|y| \sim \ln \frac{\sqrt{s}}{M} x \qquad x = x_1 - x_2 \ , \tag{9.1}$$

where x_i is the longitudinal momentum fraction from one proton and M is the system invariant mass. Examining figure 5 it seems reasonable that $\langle x \rangle$ will be higher by choosing a quark,

[18] This being the case for two of the 2012 Higgs candidates in the 125 GeV region results in a 3σ effect, which seems highly incredible

anti-quark pair than for a gluon pair. This will result in a tendency for a wider rapidity distribution for the ZZ* production than that of the Higgs production.
The Closely related pseudo rapidity distribution was also examined.
In figure 40, it seems by eye measure that the pseudorapidity distribution differs more that the rapidity, but from figure 37 it is seen that either choice gives the same separation, indicating that they contain the same amount of information. Test and training comparison also gives similar behavior. Seeing no difference in behavior η was chosen per default.

9.1.3 Z/Z* masses and variations.

For a given four lepton mass interval, the reconstructed Z and Z* masses will in general not be the same, for the Higgs production and the Standard Model ZZ* production. A number of different parameters that expresses this have been examined, figure 41. First it could be reasonable to use the Z masses directly as input. This however has the disadvantage that they give worse separation than the variable choices discussed below. Moreover it would seem that if it was possible to construct variables, less dependent on the reconstruction resolution of the Higgs mass, which the Z masses are correlated to, this would be preferred.
To partially correct for reconstruction resolution and make the distributions more BDT friendly, i.e. bounded on a well defined interval, the possibility of correcting the Z masses with the four lepton masses was also examined:

$$x_1 = \frac{M_Z}{M_{4l}} \qquad x_2 = \frac{M_{Z^*}}{M_{4l}} \tag{9.2}$$

For low Higgs masses ($< 2M_Z \sim 180$ GeV) the lower boundary on the Z* mass is approximately linearly dependent on the Higgs mass. Dividing the using x_1 and x_2 will keep the boundary constant for different masses and thereby reduce the variable correlation to the Higgs mass.
Finally a set of parameters that expresses the available phase space of the system and relative differences in Z masses were constructed:

$$x = 1 - \frac{M_Z + M_{Z^*}}{M_{4l}} \quad x^* = 1 - \frac{M_Z - M_{Z^*}}{M_{4l}} \ ,$$

also partially corrected for the reconstruction resolution.
Figure 37 indicates that using Z masses directly gives the worst separation, while using x, x* by all means gives identical separation to x_1, x_2.
The Kolmogorov-Smirnoff comparison of the test and training samples return probabilities for x_1, x_2 of 0.65 and 0.02 for signal and background respectively, while the x, x* gave 0.25 and 0.01 respectively. This could be an indication of some undesirable tail behavior of the x, x* distributions, but the evidence for this is lacking. In the end the scaled Z masses x_1 and x_2, were chosen.

9.1.4 Angle between Higgs and Z direction of flight $|\cos(\theta_{HZ})|$

It is possible to use the fact that the Higgs boson manifestly is a spin-zero particle. In the rest frame of the Higgs boson, the Z bosons will not have a preferred direction of decay. This

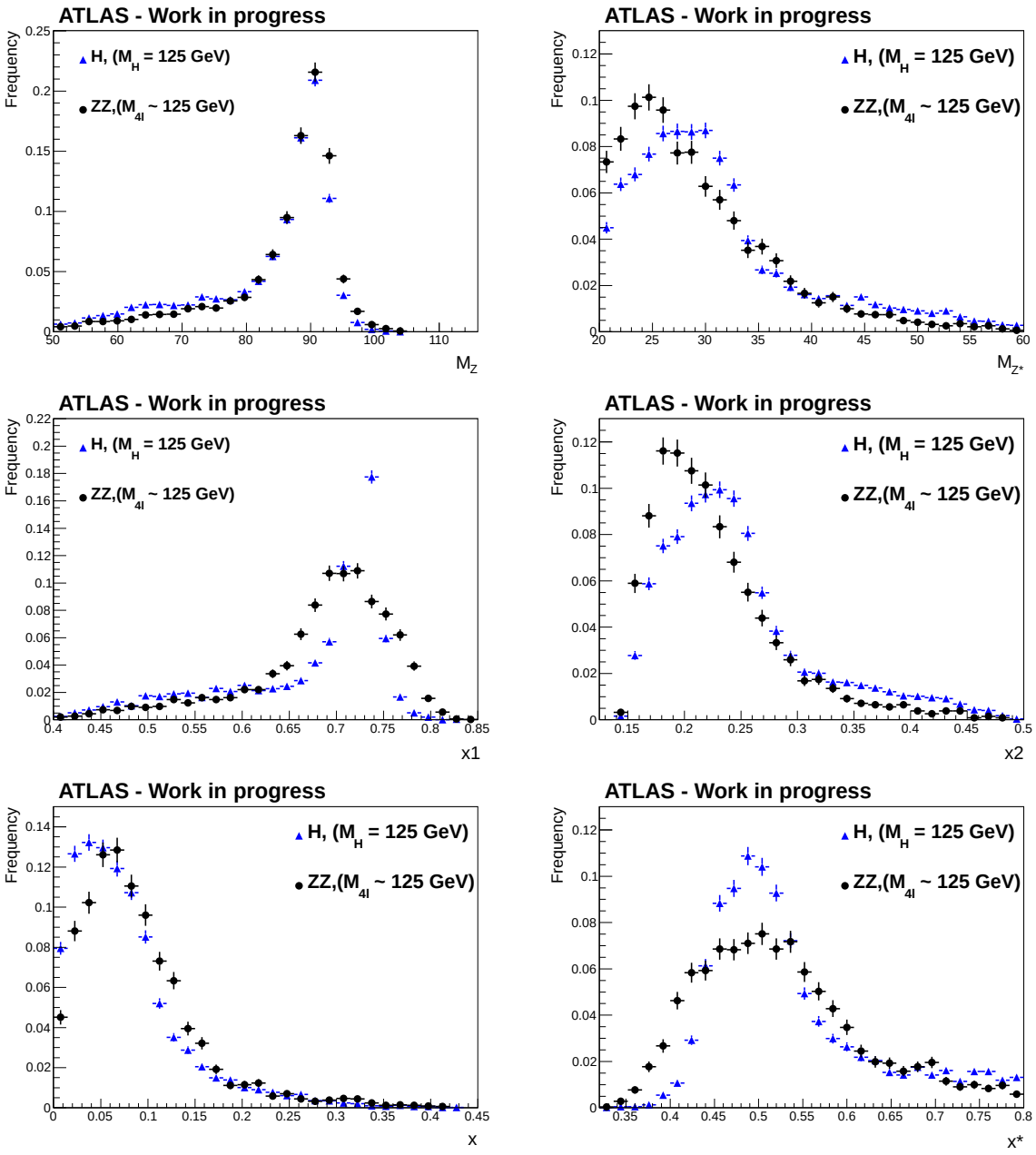


Figure 41: Higgs and neutral current spectra for Z/Z* related distributions. Top row: Z mass(M_Z) and Z* mass(M_{Z*}). Middle row: Relative Z mass(M_Z/M_{4l}) and relative Z* mass (M_{Z*}/M_{4l}). Bottom row: Available phase space $(1-(M_Z+M_{Z*})/M_{4l})$ and relative Z mass difference $(1-(M_Z-M_{Z*})/M_{4l})$

will on the other hand not be the case for the Standard Model production of ZZ diboson. The angle θ_{HZ} is measured between the Higgs(candidate) direction of flight in the ATLAS

frame and the Z boson direction of flight.
The concept of the Higgs direction of flight in the Higgs frame of reference does not truly make sense. In order to have a unanimous definition, the boost into the Higgs boson frame, is always done in a way that preserves the orientation of the ATLAS detector, e.g. up and beam direction, such that the direction of flight in the Higgs frame and ATLAS frame is the same. The Higgs and Standard Model diboson distributions can be found in figure 42.

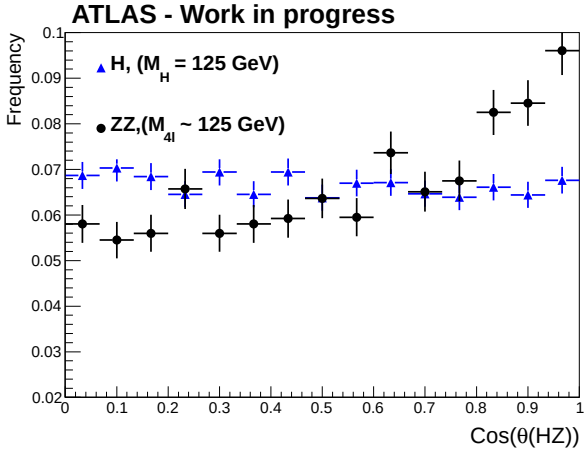


Figure 42: Higgs direction of flight to Z direction of decay correlations for Higgs and neutral current

9.1.5 Angles between Z and lepton direction of flight $|\cos(\theta_{Zl0})|$ and $|\cos(\theta_{Z^*l2})|$

The two variables also make use of the differences in angular momenta. This can be used if the direction of decay of one lepton is measured in the corresponding Z boson frame of reference. The angle is measured between the Z boson direction of flight in the Higgs frame and the negatively charged lepton in the Z frame, still keeping the orientation of the ATLAS detector constant. The result of these can be seen in figure 43 left and right for Z and Z* respectively.

Examining figure 43, the Higgs and Standard Model distributions only differs very slightly for $|\cos(\theta_{Z^*l2})|$ values greater that -0.4.
The analysis was run with and without $|\cos(\theta_{Z^*l2})|$ and it is seen from the ROC integral in figure 37 that the inclusion does not improve the result in any notably way. For this reason, $|\cos(\theta_{Z^*l2})|$ was omitted in the further analysis.

9.1.6 Angle between Z/Z* decay planes χ

Figure 44(top) shows that if the spin of the two Z bosons is considered perpendicular to their direction of flight, their spin will be completely correlated in the Higgs decay due to conservation of angular momentum, and uncorrelated in the spin 1 case.

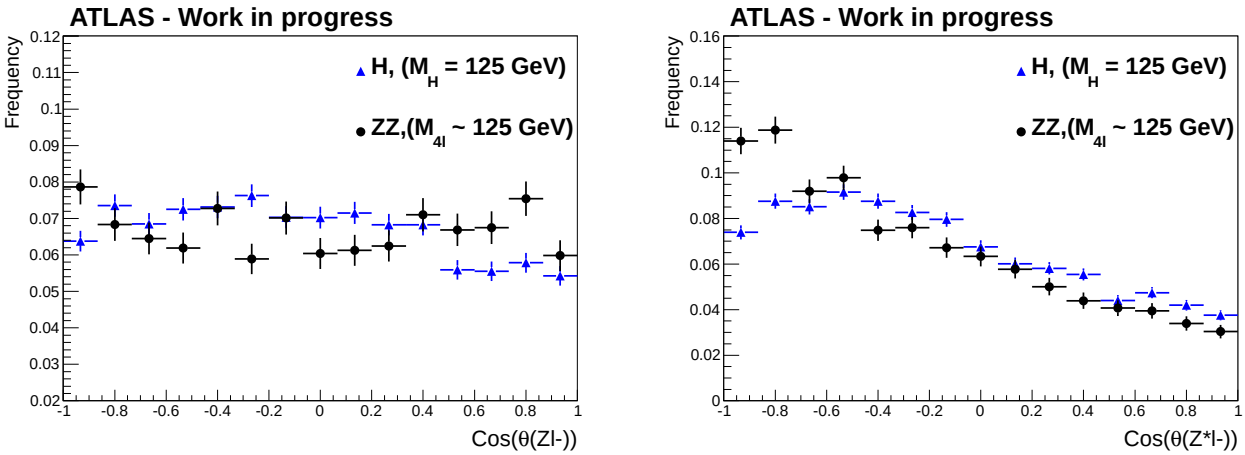


Figure 43: Z direction of flight to lepton direction of decay correlations for Higgs and neutral current

The angle χ is defined as the angle between the two vectors:

$$\vec{p}_{l_Z^+} \times \vec{p}_{l_Z^-} \qquad \vec{p}_{l_{Z*}^+} \times \vec{p}_{l_{Z*}^-} \tag{9.3}$$

Measured in the Higgs candidate frame of reference.
In the Z rest frame the decaying leptons will align their spin such that they combined add to the Z spin. The matter particle being left handed and the anti matter right handed results in a tendency for e.g. an electron to decay in the opposite direction of the Z spin while the positron will decay in the same direction. This means that the plane in which the leptons decay will be correlated if the Z bosons stem from a Higgs boson, and uncorrelated if the system has spin one. The result can be seen in figure 44 (bottom). The correlation is however quite weak, and figure 37 indicates that it can be omitted without any loss of separation power.
The variable is however not completely without influence as it is stronger in the separation between spin zero and spin two particles. A prospect that is interesting in the further study of the newly discovered boson.

9.1.7 Training

The list of variables that will go into the multivariate analysis has thus converged on six, as indicated in table 18.
While the variables discussion was based on arguments used on results obtained using a Boosted Decision Tree, other MVA methods were also examined to see if better separation could be obtained, or similar with less sensitivity to the size of the statistics available.
Both the neural network method CFMlpANN [1, p.91] and the analytical Fisher method [1, p.84] were examined. It was found that the neural network performance was the same as that of the BDT, but was far more demanding in terms of computational power while the Fisher method separation was worse. It can be shown that that the Fisher method [18, sect.4.4] gives the optimal separation if variables are only linearly correlated, which is not the case in

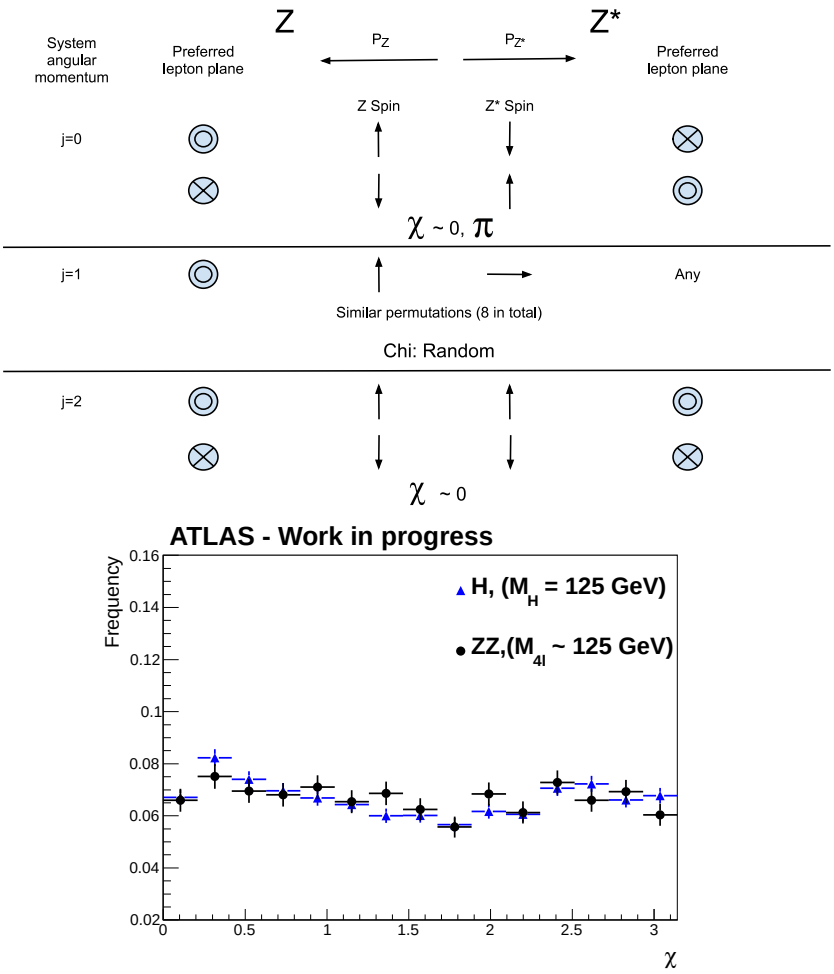


Figure 44: Top: Correlation chart for Z/Z* decay planes. If the system angular momentum (j) is 0 or 2, the orientation of the decay planes will be correlated ($\odot$ symbolizes a vector pointing out of the paper while $\otimes$ corresponds to a vector pointing into the paper). This is not the case for j=1. Bottom: Z/Z* decay plane correlations for Higgs and neutral current. It is seen that the separation of this parameter is very weak.

question. The stability of separation under exchange of variables in figure 37 also indicates that if information is available, the BDT will find use of it.

In the end, through trial and error, it was found that a BDT containing 150 trees with a maximal depth of two gave the optimal performance without overtraining. It was further demanded the each hypercube should contain at least 100 events. The maximal depth of two limits the BDT to only take correlations between two variables into consideration, but it was found, using the larger samples that was generated without the full detector simulation as discussed below, that there was not more information accessible using more complex trees. The sensitivity to overtraining on the other hand grows fast when allowing divisions into cubes

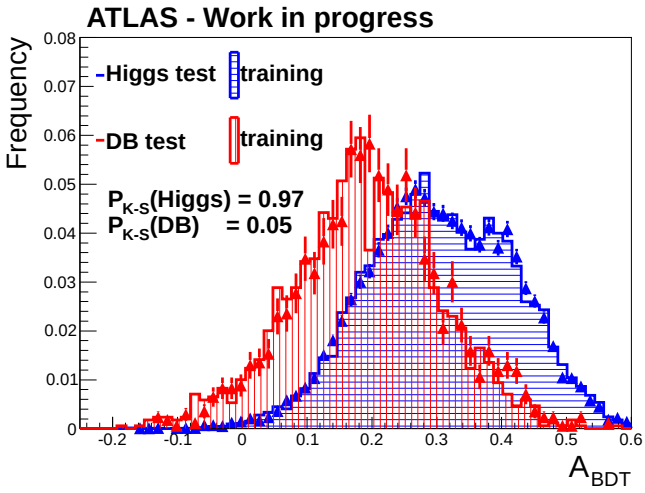


Figure 45: Boosted decision tree overtraining test. The BDT was trained with 125 GeV Higgs sample and Standard Model diboson sample in the mass window 125 ± 10 GeV. The points corresponds to the test sample while the filled areas shows the training samples. A Kolmogorov-Smirnoff comparison was conducted between the testing and training sample of each input, and returned values of 0.97 and 0.05 for the signal and background respectively.

and hypercubes of more dimensions. An example of an overtraining test can be seen in figure 45 [19]. The signal and background distributions are divided into a test sample and sample on which the BDT is trained. The test sample is evaluated using the found BDT structure. If the testing and training distributions are not equal it is an indication that the BDT is over trained, and its implementation can introduce biases.

Finally the correlation between a Higgs candidates invariant mass and BDT score was examined, both for signal and background, the result of which can be seen in figure 46. For a successful implementation in the analysis, it is of highest importance that the multivariate analysis will not select specific masses. If e.g. the background had higher score at a certain mass, a cut would create an unphysical peak, or as is the case in this analysis implementing the BDT in the mass likelihood fit would create biases.

9.2 Extension to full mass search

When the analysis was written, the Higgs search was still being performed. In the theory section it was discussed that it is not possible to calculate the mass directly, and only vague predictions gave favor of a Higgs mass of 90 ± 30 GeV, combined with exclusion from previous experiments up to $\sim$115 GeV as seen in figure 9. The indirect calculation was also performed under the assumption that the only missing massive particle in the Standard Model was the

[19] These are not from the same values as listed in the Z/Z^* discussion. The overtraining test is taken from the training using multiple signal masses as input as described below.

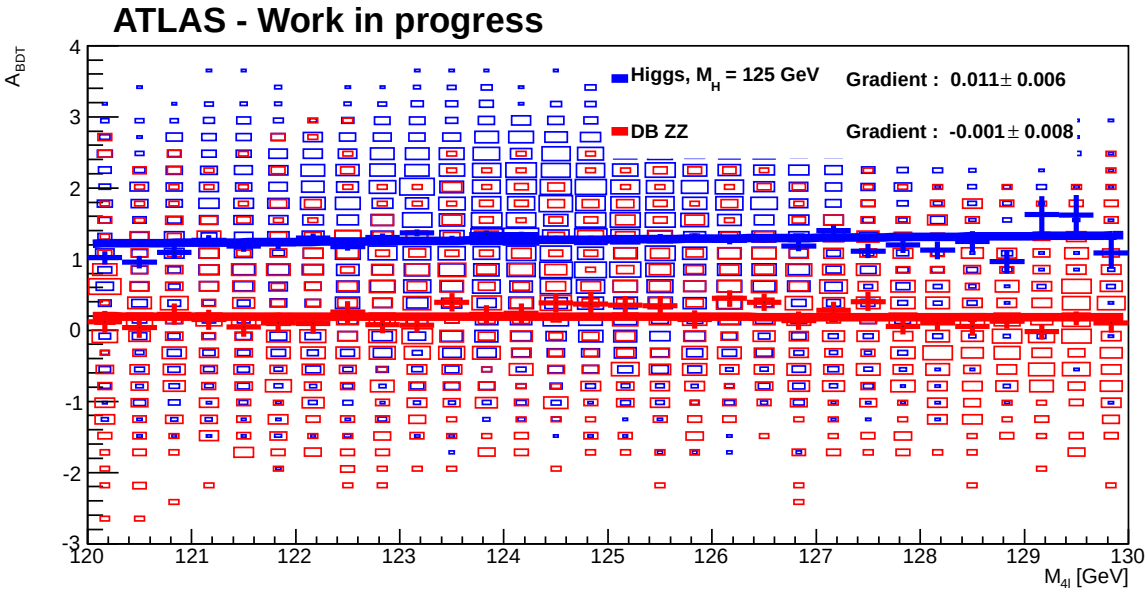


Figure 46: Dependence check between mass and BDT score distributions. The BDT profile, indicating the score mean for each mass bin is also shown. It is seen that there is no correlation between the parameters. Note that in the further implementation, the 125 GeV training sample will only be used in the interval 122.5 GeV to 127.5 GeV

Higgs boson. The official ATLAS $H \rightarrow ZZ^{(*)} \rightarrow 4l$ analysis was performed in a mass window from 110 GeV to 600 GeV [11]. It was however chosen that this analysis should only span the mass window 110 GeV to 160 GeV.
This choice was made partly due to simplicity, but also influenced by the 2011 indications of the interesting search areas. It was chosen to include search areas that were already excluded, cf. figure 9 to get a better handle on the background. If e.g the data shows distinct signal-like behavior in the BDT at the high masses of the search window it could be an indication of effects that are not understood.
The inclusion of the Z masses as separation variables also made it necessary to end the search window at masses some distance below $2M_Z \sim 180$ GeV. The discussion of the variable choices would in general be much different if the analysis extended a larger mass region.
As can be seen from figure 46 the BDT score does not correlate with the four lepton mass in a ± 5 GeV interval around the signal mass. This is not the case if the interval extended the 50 GeV that the search is being performed on. As an example, the score distributions for a BDT trained on an 125 GeV sample for signal and background candidates around 125 GeV and 150 GeV can be seen in figure 47 (left). The miss matched events are not nearly as separated.

It is therefore not possible to simply make one training where a single Higgs mass and the whole diboson background sample are used as input. First of all, if too wide a mass window is

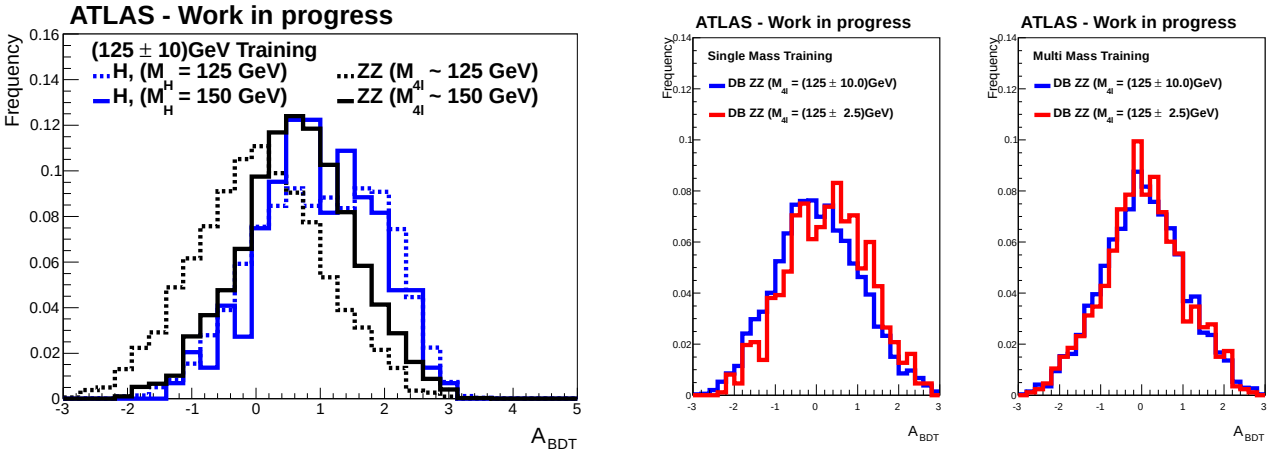


Figure 47: Left: Example of using a boosted decision tree trained on different masses that it is evaluated on. Here a tree is trained on a 125 GeV sample (dashed line). The scores of a 150 GeV sample can be seen to be much less separated. Right: Training the BDT on a ±10 GeV sample and evaluating it on a ±2.5 GeV sample could lead to mass dependence of BDT score. It is seen that if a single signal mass is used, this will indeed be the case, but not so using the more flat mass distribution of all signal samples.

used for the background sample, the BDT will use mass dependent discriminants to separate the signal and background, introducing a bias into the analysis. As an example, the allowed Z* masses are very restricted by the interval in which the four lepton mass is considered. Around 120 GeV, the bulk of the distribution only extend to ∼30 GeV, while around 160 GeV it extends as far as 60 GeV. Using 120 GeV signal and 110 GeV to 160 GeV background samples thus gives very high separation, when the four lepton masses are different and low separation when they are similar.

Even when restricting the background sample to the same masses as the signal sample, the BDT will still contain information about which mass it was trained on, since e.g. the width of the rapidity distribution is mass dependent [27, eq.3.7]:

$$\Delta y \sim 2 \cdot \frac{\sqrt{s}}{M} \;\Rightarrow\; \frac{\Delta y_{110\ \mathrm{GeV}}}{\Delta y_{160\ \mathrm{GeV}}} \sim 1.1 \tag{9.4}$$

This could make the distribution for a 110 GeV Higgs similar to the background distribution at 160 GeV.

Taking this into consideration it was chosen to train a BDT at 5 GeV intervals, using a mass window that extends 10 GeV in either direction starting at 110 GeV and ending at 160 GeV. This way the used mass windows should be small enough to be approximately decorrelated with the four lepton mass, and just wide enough that the background sample contains sufficient statistics to make a reasonable separation possible.

The problem with this is that a single signal sample will not contain statistics in the endpoints of training interval. To solve this, all mass samples were added together and every reconstructed mass in the given training interval was used, independent on which true mass

it stemmed from, figure 48.

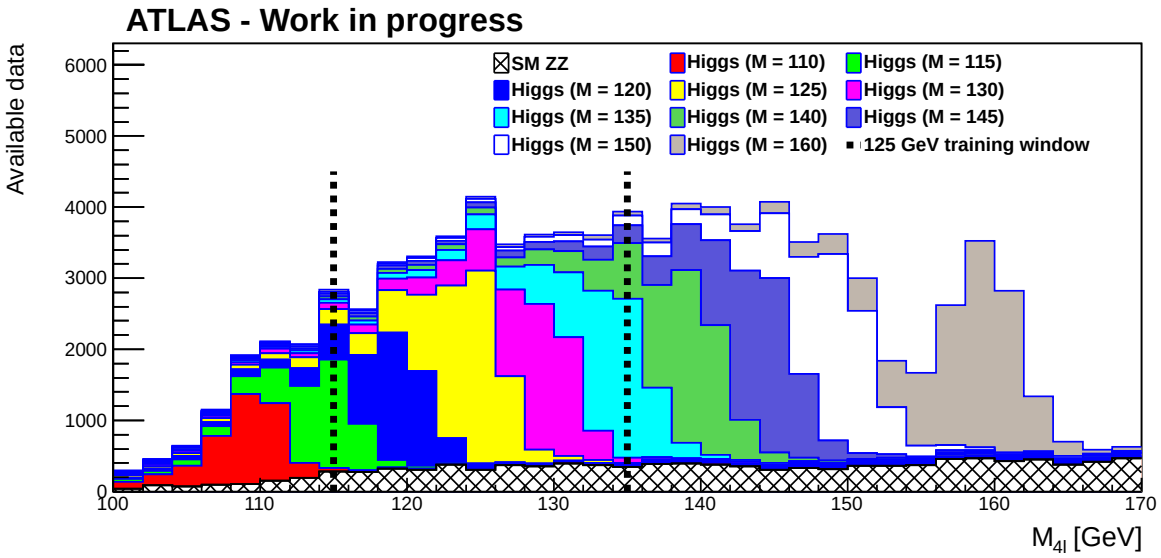


Figure 48: Invariant mass window used for the 125 GeV training. All candidates were used that fell into the interval independent on true four lepton mass. It is seen that except for the endpoints $\lesssim$110 GeV and $\gtrsim$160 GeV, this method will give a reasonable constant amount of data points as function of mass for each training.

Figure 47 (middle and right) indicates that this gives a very positive effect. The middle diagram shows the diboson distribution when only trained against the 125 GeV sample, while the right plot shows diboson candidates separation from all signal events reconstructed to have a mass around 125 GeV. The blue curve shows the distribution evaluated for candidates in the $\pm$10 GeV training window, while the red curve represents candidates in a $\pm$2.5 GeV window. The single mass training is thus able to distinguish the mass window while the multi mass input training is not.

To be able to have a continuous description over the total mass window of the search, it was chosen to transform the output such that the background distribution was described by a Gaussian with mean 0 and width 1.

This was done using the kernel density estimator of the signal and background distributions, a rarity transformation and a subsequent Gaussian transformation, as described in the statistical techniques section.

Any given candidate was given a score calculated using the BDT trained at the mass closest to the candidate mass. The resulting distribution for the full mass window can be seen in figure 49, together with a Higgs signal of 125 GeV.

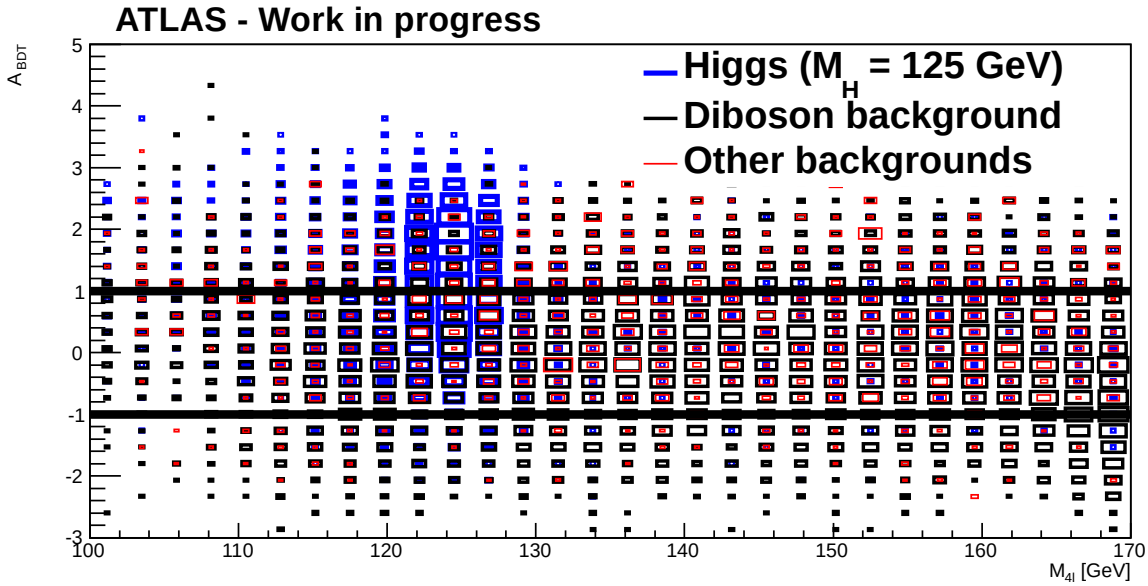


Figure 49: Mass vs. A_{BDT} correlation. The mass and boosted decision tree value for a Higgs signal at 125 GeV together with backgrounds. It can be seen that the construction of training BDTs at points of a mass grid, with 5 GeV intervals, renders the mass and BDT score uncorrelated to a high degree. The ± 1 bands approximately indicates the one standard deviation bands of the diboson sample.

9.3 Systematic uncertainties

The multivariate separation of a Higgs signal and the Standard Model production of a ZZ diboson seems promising.
In order to estimate the theoretical and systematical uncertainties of the analysis, it is necessary to change settings, rerun each simulation and examine the possible effects.

9.3.1 Generation

The following analysis of theoretical and systematical uncertainties was made using Pythia8 and the POWHEG-BOX simulator for the hard processes, [50], [43], [44], [48]. Only the processes contributing most to the cross section were simulated; gluon and vector boson fusion for the Higgs signal and quark exchange for the Standard Model diboson background. The POWHEG-BOX is able to do next to leading order processes, while Pythia8 only simulates leading order. Everything after the hard process, e.g. hadronization and radiation, was simulated using Pythia8.
It was chosen to use the Pythia8 generation with default settings as the nominal simulation and estimate the theoretical uncertainties by deviations from the resulting distributions.

9.3.2 Detector and selection effects

It is desired to test a number of different models, but it would be far to demanding to do a full detector simulation for each. Therefore an effective model was constructed which was able to describe acceptance, efficiency and smearing effects of the ATLAS detector.
The first step of the effective detector simulation was to impose acceptance and efficiencies on the individual leptons. The acceptance and efficiency model was derived using the large $Z \rightarrow$ll + N_{Jets} samples containing millions of events (appendix A). The individual lepton requirements as described in the HSG2 requirement section was imposed and the number of reconstructed divided by the total number of simulated leptons were saved in a p_T, η grid, figure 50. This way, both detector efficiencies and acceptance is accounted for, but also the specific analysis effects.
The reconstruction probability assigned to each lepton, in fast simulation, was then chosen from the appropriate p_T, η bin in figure 50.
If a lepton was 'reconstructed' the energy/momentum measurement was afterwards smeared by the resolution expressions for the inner detector and calorimeter for electrons and the inner detector and muon spectrometer for muons found in table 19.

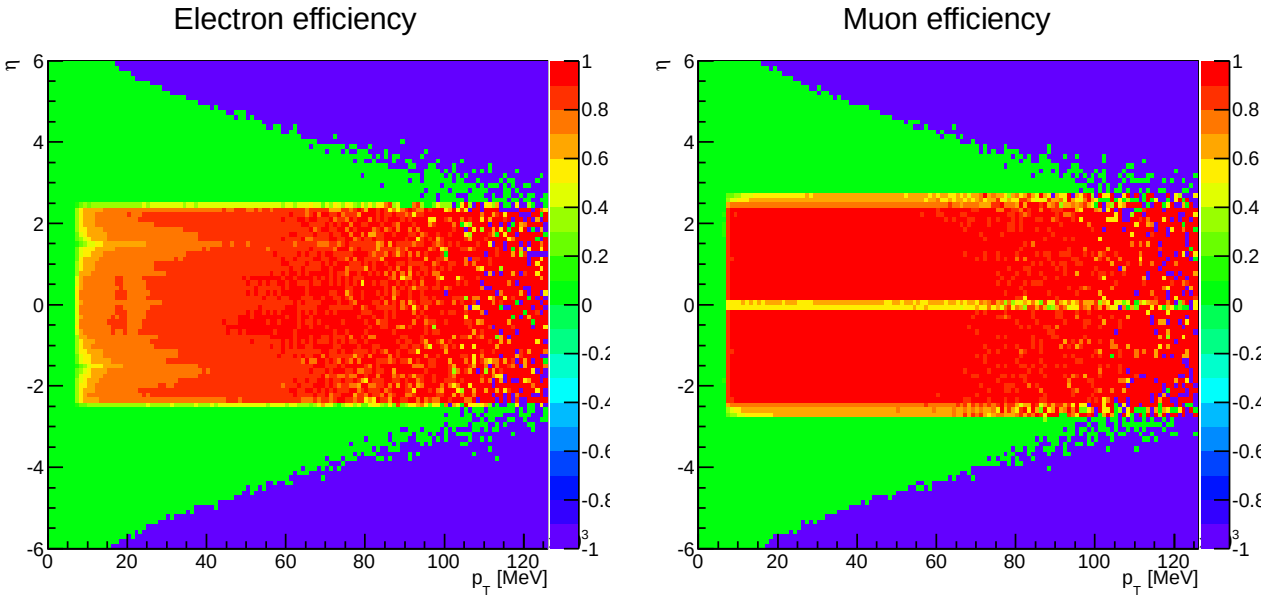


Figure 50: Acceptance times efficiency as function of p_T and η for electrons (left) and muons (right). Note that -1 refers to areas without statistics. Data were taken from the Z→ll + N_{Jets} samples, where the individual lepton requirements were imposed, tables 10, 12.

Finally the lepton and quadruple kinematical requirements, from table 10, were imposed on the fast simulated Higgs candidate.
To check whether this very simple approach gives an accurate description of the variables used for the analysis a full simulation-fast simulation comparison was constructed. Fully simulated

Inner Detector	e, μ	$\sigma_{p_T}/p_T = 0.05\% \cdot p_T \otimes 1\%$	[GeV]
Calorimeter	e	$\sigma_E/E = 10\%/\sqrt{E} \otimes 0.7\%$	[GeV]
Muon Spectrometer	μ	$\frac{\sigma(p)}{p} = \frac{p_0^{MS}}{p_T} \otimes p_1^{MS} \otimes p_2^{MS} \cdot p_T$	

Table 19: Smearing expressions imposed on the simulated leptons. The muon spectrometer values can be found in ATLAS detector section.

gluon fusion sample for a Higgs boson of 125 GeV was used for the test.
The fast simulation method was imposed on the 'true' lepton momenta, such that the output contained a 'true' simulated Higgs, a fast simulation Higgs candidate and a full simulation Higgs candidate. For each of these the multivariate analysis variables were calculated and compared, the result of which can be seen in figure 51.
It is seen that this method gives an accurate description of the relevant detector effects. A Kolmogorov-Smirnoff comparison was conducted comparing the fast and fully simulated variables and returned probabilities between 17% and 88% for the given amount of statistics, and can thus be used as a detector simulation tool imposed after each theoretical model was simulated.

It should be noted however that there are possible subtle variable correlations and other effects that are not accounted for nor expressed in the full simulation, fast simulation comparison described above. This is the reason why the fast simulation was deemed ill suited for the BDT training on which the experimental data was evaluated.
The method on the other hand was considered to be appropriate for uncertainty estimation. Since the systematic effects are expected to be influenced equally by any shortcomings of the fast simulation, this will render their relative discrepancies well described.

Below will be given a summary of the different changes imposed on the simulation process. The first discussed can be considered theoretical uncertainties, and consist of changes in the parton density function, renormalization scale of the hard process and the initial transverse momentum of the hard process. The following will be more oriented towards the calculational techniques of the simulation and include the choice of kinematical constraints imposed on the system, the simulated acceptance of the detector and the hard process simulator. The change of hard process simulator from Pythia8 to PowHeg also changes the order of perturbation of the hard process calculation, and could be equally well belong in the first category. The resulting BDT score distribution can be found in figure 52, while the individual distributions are shown in figure 53. Table 22 summarizes the discrepancies and its results are what will be used in further analysis.

9.3.3 Parton density functions

The proton parton density function is only known to a given precision. One way PDF uncertainties could affect the outcome, is that the probabilities for Higgs production through

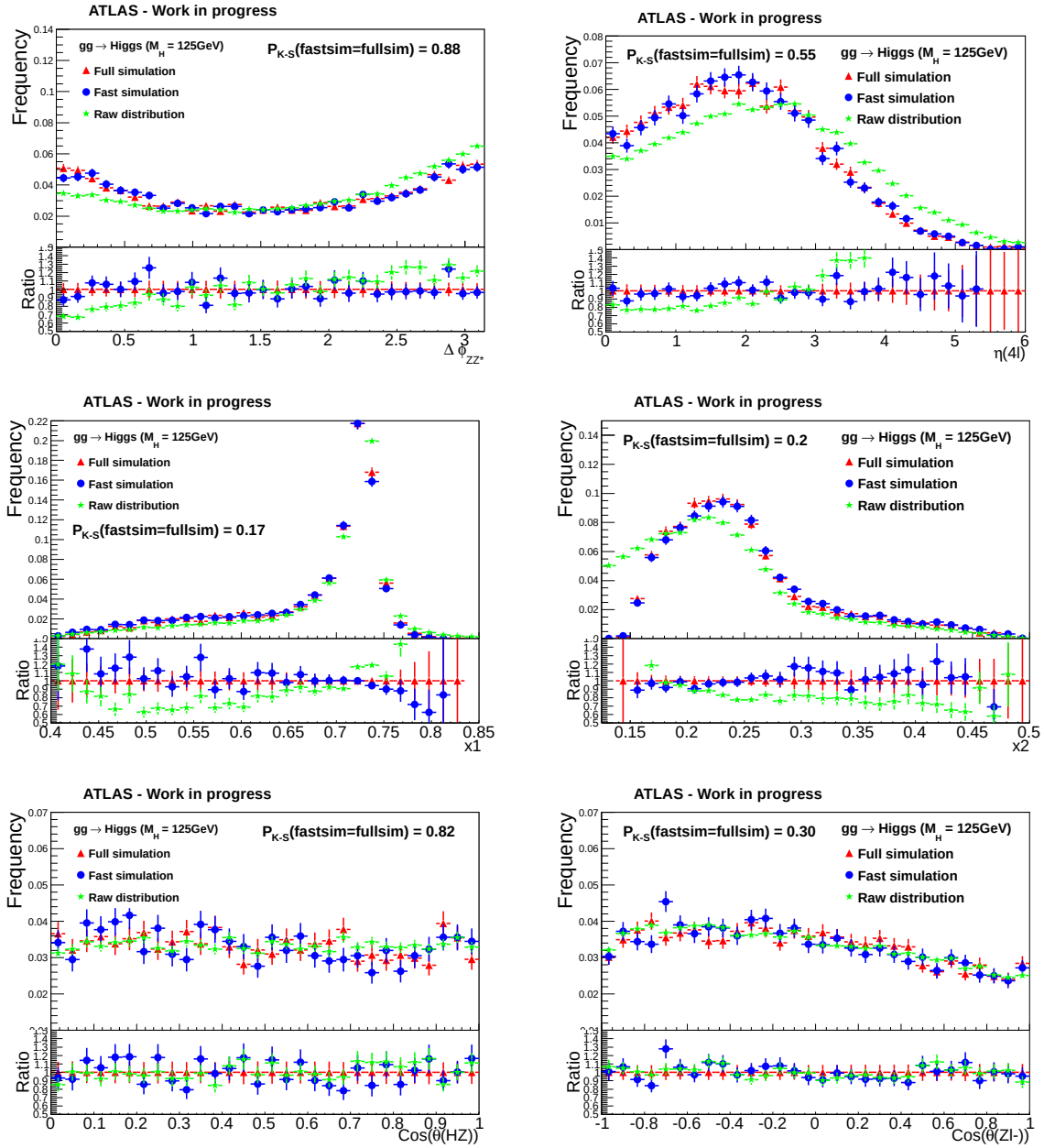


Figure 51: Fast simulation cross check for MVA variables. Top row: $\Delta\phi_{ZZ*}$, η_H. Middle row: x_1 and x_2. Bottom row: $\cos\theta_{HZ}$ and $\cos\theta_{Zl-}$. The full detector simulation distributions is seen in red, while the fast simulation is seen in blue. Generator level values are seen in green. The fast simulation gives a very decent description. A Kolmogorov-Smirnoff comparison of the fast and full simulations was also conducted with probabilities of the distributions being the same seen in the top.

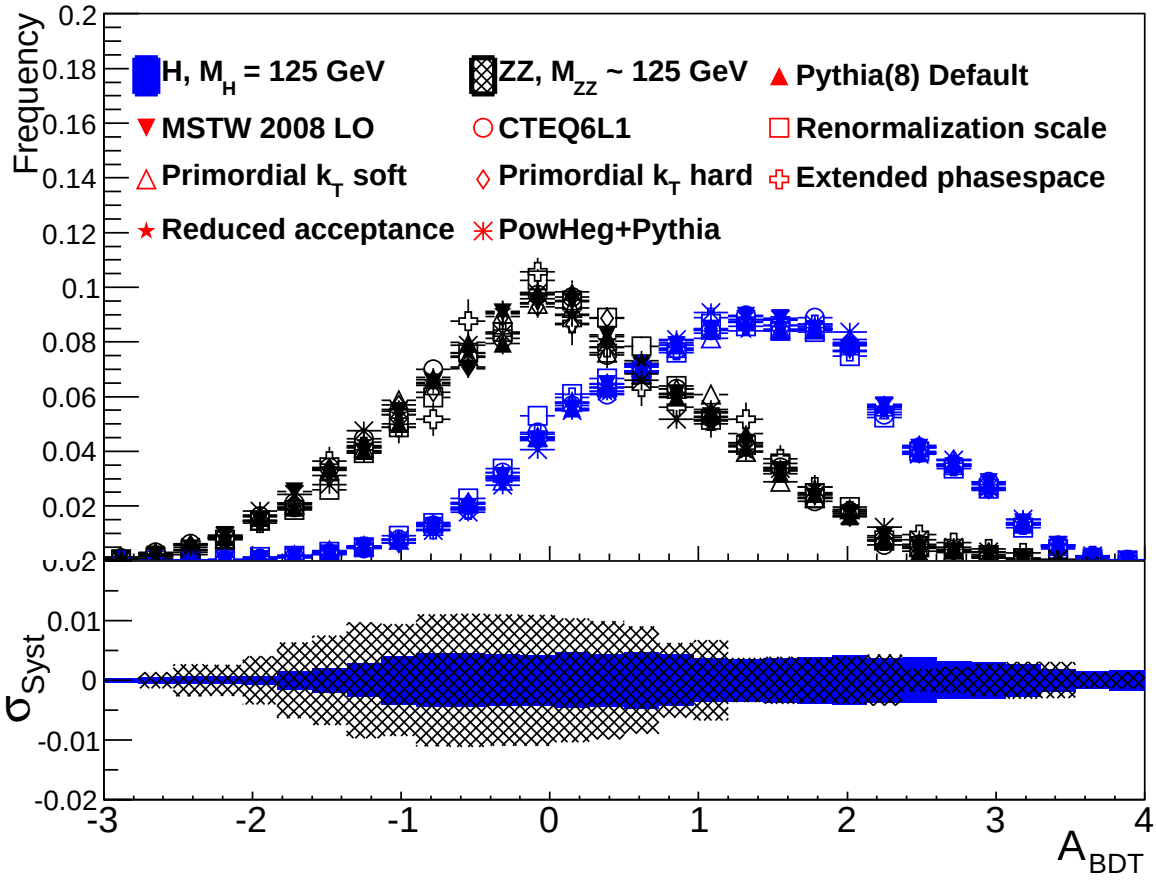


Figure 52: Evaluation of systematic uncertainties of the BDT score function A_{BDT}, for the different simulation models. Below are the differences between any models to the nominal generation added in quadrature subtracted by the estimated statistical uncertainty. The scale of the lower y-axis is 1/10 of the upper in the same units. The uncertainty at the central values are 10% and 7.5% for Standard Model diboson and Higgs production respectively.

gluon fusion and vector boson fusion are dependent on how probable it is to 'choose' a given parton at low x values. If the ratio the gluon distribution to a given quark was changed this would change the ratio of gluon to vector boson fusion events. This follows directly from the factorization theorem [27, e.q.3.9]:

$$d\sigma = Cf^1(x_1)dx_1 f^2(x_2)dx_2\hat{\sigma}_{\text{Process}} \ , \tag{9.5}$$

where f^1 and f^2 are the proton parton density functions for the relevant particles.
Apart from the final state of the two types of processes, the kinematics are somewhat different. The characteristic high η jets of VBF allow higher p_T. Another more directly way the PDFs

can influence the distributions is using [27, eq.3.7]:

$$x = 2m_T \ \sinh y/\sqrt{s} \qquad x \equiv x_1 - x_2 \qquad m_T^2 = m^2 + p_t^2 \ , \tag{9.6}$$

where both the system rapidity as well as the transverse momentum are affected by the choices of PDFs.
Pythia8 uses the CTEQ 5L, LO $\alpha_s(M_Z) = 0.127$ as the default distribution. This was changed to MSTW 2008 LO (central member), LO $\alpha_s(M_Z) = 0.13939$ and CTEQ6L1, LO $\alpha_s(M_Z) = 0.1298$ per recommendations from [47]. The value used for α_s at M_Z is also slightly different, resulting in different QCD couplings [51, PDF Selection].
In figures 53 and 52 an estimate of the theoretical uncertainty band on each distribution is also shown at the bottom. This was derived by adding the distance from each data point to the nominal Pythia8 distribution in quadrature. The normal standard deviation was not used since it is believed to give to small an estimate.
For the individual distributions this was based on the conclusion that the systematical changes does not affect variables equally. The choice of parton density function has e.g. no influence on variables related to the Higgs decay.
Similar conclusions for the resulting BDT score distribution can be reached. It can be seen from figures 52 that the distribution is not affected by the choice parton density function, in a way that it exceeds the statistical uncertainties. The examination of three different parton density functions will however change the standard deviation of each bin, since any effect that actually sticks out will be washed out by dividing by the number of models tested.
To iron out fluctuation, the summed distances were exchanged by the the corresponding mean of each, 4 bins in each direction.
The quadrature of distances will also depend on the statistical uncertainty of each bin. To account for this the statistical uncertainty, calculated from the number of entries in each bin was also added in quadrature, and exchanged by the mean 4 bins in each direction.
The estimated theoretical uncertainty was finally given as the difference between the summed distances and the summed statistical uncertainties. For the i'th bin, the expression reads:

$$(\sigma^i_{\text{Theo}})^2 \simeq \left(\frac{1}{9} \sum_{i-4}^{i+4} \sqrt{\sum_{j \in \text{model}} (\Delta^i_\mu(j))^2} \right)^2 - \left(\frac{1}{9} \sum_{i-4}^{i+4} \sqrt{\sum_{j \in \text{model}} (\sigma^i_{\text{Stat}}(j))^2} \right)^2 \tag{9.7}$$

This expression will probably tend to overestimate the uncertainty somewhat. A complete derivation, given plentiful statistics, should examine what changes can be considered independent.

9.3.4 Renormalization scale

As a feature of quantum field theory, there will naturally arise infinite integrals in the hard process calculations. These are subdued by a renormalization of the calculation, where another infinite integral is subtracted. This method introduces an arbitrary scale, the boundary of the subtracted integral. This scale is usually set according to the transverse momentum of the hard

process ($Q = p_T$). These were changed to $Q = 2p_T$ for the Higgs production and $Q = \frac{1}{2}p_T$ for the diboson hard process, per recommendations of [47]. This change will move the p_T spectrum, and equivalently acoplanarity, spectrum of the signal and background distributions closer together.

9.3.5 Primordial k_T

A complicating factor, when considering hadron collisions at high luminosities is multiparton interactions. Even though the hard process is only characterized by simpler kinematics, the possible rescattering of partons introduces a degree of freedom for the momentum in the transverse plane. Since the QCD calculations involving multiple interactions are not possible to perform directly, Pythia8 introduces an effective description where the parton is given a primordial k_T selected randomly from a Gaussian distribution with width [51, Beam Remnants]:

$$\sigma_{kt} = (\sigma_{soft} \cdot Q_{half} + \sigma_{hard} \cdot Q)\frac{(m_{half} + m)}{(Q_{half} + Q) \cdot m} \tag{9.8}$$

The parameters σ_{soft} and σ_{hard} defines the width of the contributions from soft and hard scattering respectively. Q is the hard process renormalization scale, m the mass of the system, and Q_{half}, m_{half} defines the halfway scale/mass where an interaction is considered soft or hard.
This being an effective model, it is worth examining the impact on the analysis, where it is expected that the four lepton transverse momentum will be affected. The default Pythia8 setting is a hard scale width $\sigma_{hard} = 2.0$ GeV which was tried changed to 1.0 GeV and 3.0 GeV per recommendations of [47].

9.3.6 Phase space

For any given simulated process, it is necessary to impose restrictions on the phase space of the system to avoid divergences. There is however the possibility that this cut on phase space could change the distributions, if they are sufficiently close to areas the analysis is sensitive to. As an example, in order to avoid divergences in the generation of Standard Model diboson pairs, a lower cut on the generated Z^* mass helps ignore the Z/γ interference, whose cross section grows fast at low masses [3, fig.9.9]. Another more technical problem arises because Pythia8 considers Z bosons and photons two distinct particles. Without a clear definition that distinguishes these, double counting of particles could potentially happen [51, Phase Space Cuts].
Due to the finite detector resolution, a reconstructed Z mass could be calculated too high, making it insufficient to impose the same kinematical acceptance as given in table 10. Tail effects of masses lower than the acceptance but reconstructed higher will enter into the accepted analysis. The list of phase space requirements are given in table 20.
It should also be noted that it is of interest to set these as high as possible without affecting the analysis, because all events that are simulated to be in the region between the phase space

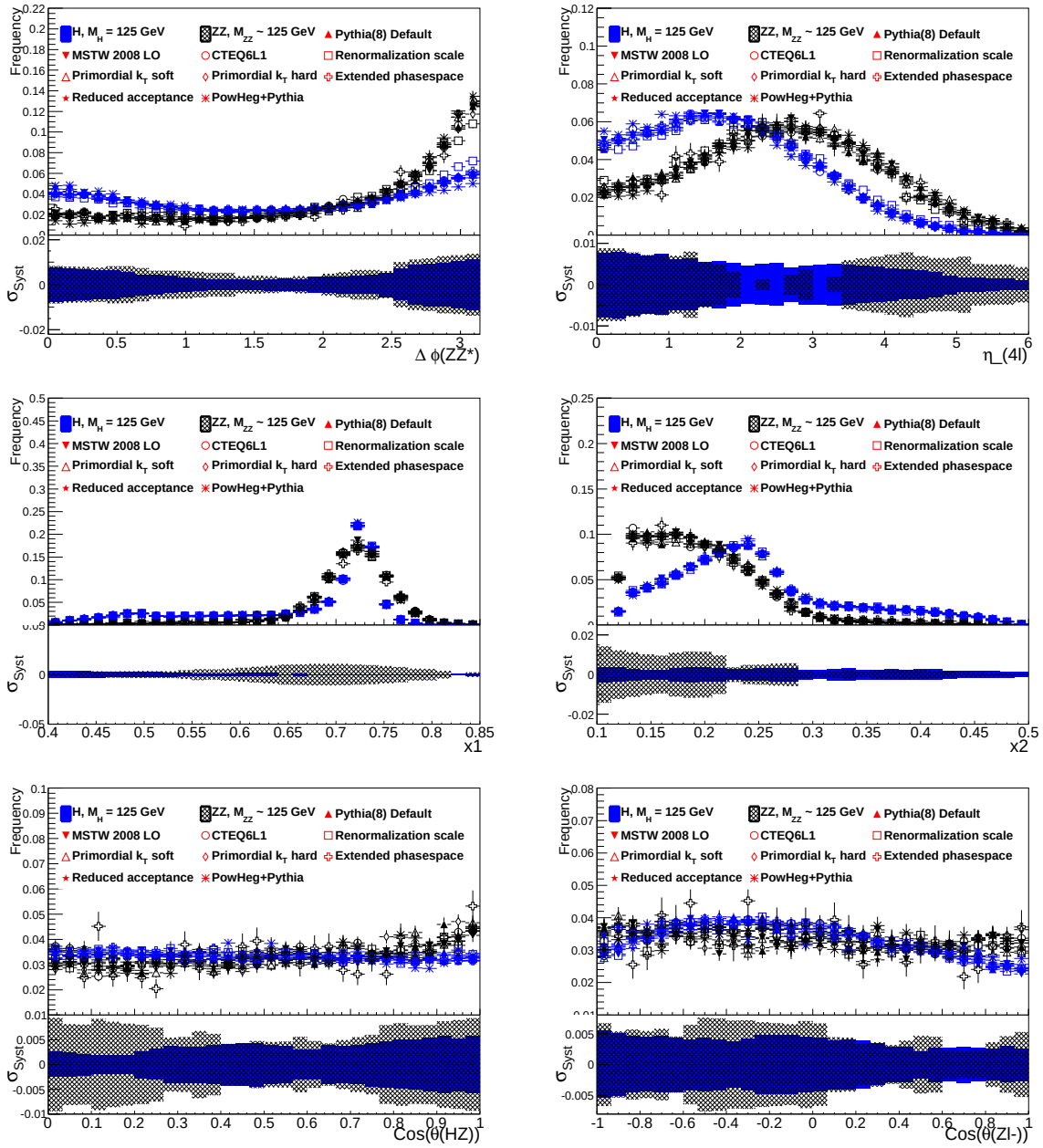


Figure 53: Evaluation of systematic uncertainties of individual variables. Top row: $\Delta\phi_{ZZ*}$ and η_H. Middle row: x_1 and x_2. Bottom row: $\cos\theta_{HZ}$ and $\cos\theta_{Zl-}$ are shown for the different simulation models. As in figure 52

cut and the analysis kinematical requirements are thrown away in later steps.

	Higgs		Diboson	
Simulation	$p_T(4l) >$ [GeV]	$M_Z >$ [GeV]	$p_T(4l) >$ [GeV]	$M_Z >$ [GeV]
Pythia fastsim.	1.0	10.0	1.0	10.0
Extended PS	1.0	2.0	1.0	5.0
ATLAS fullsim.	0.3	2.0	0.0	12.0

Table 20: Simulation requirements imposed in order to avoid divergences. The top row shows the choice for the fastsim Higgs and Standard Model diboson simulation, while the next row shows how these were extended to estimate the effect on the systematic uncertainty. The bottom row show the ATLAS simulation settings.

	e		μ	
Simulation	$p_T >$ [GeV]	$\lvert\eta\rvert <$	$p_T >$ [GeV]	$\lvert\eta\rvert <$
Nominal	7.0	2.47	6.0	2.7
Reduced	7.35	2.35	6.3	2.57

Table 21: The nominal row is the known ATLAS values for the imposed η and p_T requirements from the analysis. The known acceptance was reduced to estimate the BDT score dependence. Due to limited statistics it was chosen to reduce the acceptance far more that what the known uncertainties are.

9.3.7 Acceptance and resolution

In order to account for potential imperfections in the detector simulation, the acceptance and resolution of the lepton transverse momentum measurement was changed. It was assumed that the ϕ, η resolutions were negligible compared to the p_T measurement and was therefore not altered. The fiducial volume for both electrons and muons were reduced by 5% in both p_T and η while the smearing factor imposed on the lepton momenta, as described in table 19, was increased by 5%. The changes were made in a direction that expectantly would decrease the separation of the two distributions. A summary of the changes imposed can be found in table 21.
The reconstruction efficiencies were assumed to be the same. This was done mainly because an overall scaling of the efficiency model used, as described above, would leave all distributions uncharged. A more elaborate analysis on real data would be needed in order to improve the understanding of this.
This reduction is quite excessive as compared to what the real uncertainty on the ATLAS acceptance is. It however only removes 1.5% of the sample, in the electron case. The choice was made to examine whether any effect at all could be seen above the statistical.

9.3.8 Hard process generator/Order of calculation

Pythia8 was used as the default generator, whose unchanged setting were considered as nominal values for the distributions. Since Pythia8 uses leading order calculations, it is worth

examining the next to leading order correction to the variables used. For this, the POWHEG-BOX was used to simulate the hard processes; gluon fusion and vector boson fusion for the signal, who were added in accordance to their respective cross sections and $q\bar{q} \rightarrow ZZ \rightarrow 4l$ for the background. The output of the POWHEG-BOX was sent on to Pythia8 who handled decays and QCD effects such as multiparton interactions, final state radiation and hadronization. Any discrepancy between the two distributions could thus both be from differences in method as well as precision of calculations.

9.3.9 Evaluating theoretical uncertainties

Estimating the size and effect of the systematic uncertainties poses several questions. It is not intuitively clear by which method the differences of the different models should be propagated through the boosted decision tree, resulting in numbers that can be used in the later analysis. Keeping in mind that the BDT only takes single events as input, and is therefore not able to evaluate distributions with uncertainties, it is clear that uncertainty bands on the individual distributions is of no help.
An approach could be to train a BDT on each model, transform the output such that e.g. the background distribution was described by the same PDF, and compare the signal models. This offers the problem however, on which BDT structure the actual data is going to be evaluated in the end. To solve this a nominal BDT could be chosen, but it seems uncertain how the individual data point would fare in the other model structures, even though their overall PDFs are similar.
It was evaluated that the uncertainties would best be represented by choosing a nominal model, the Pythia8 default setting. A BDT was trained on the nominally sample, and all other systematic model samples were evaluated using this BDT structure. This way, one can think of the BDT as a constant expression, and the discrepancies among the different models as uncertainties in what the experiment will give.
The background PDF of the BDT were chosen to be the unit Gaussian with mean zero for the nominal model. Using this approach, what is interesting in further analysis, since the signal distribution is also Gaussian like, is the mean and width of each distribution. Using the same logic as when calculating the systematic uncertainty bands of the distributions, all deviations from the Pythia8 default distributions have been added in quadrature.
The result can be seen in table 22.

Using the standard formula for error propagation, the theoretical uncertainty on the Gaussian separation (S) is estimated to be:

$$\boxed{\frac{\sigma_S}{S} = 9.2\%} \tag{9.9}$$

It is worth mentioning that the BDT distributions describe what probability a single signal or background A_{BDT} measurement is expected. The effect of the uncertainty on several measurements will be implemented below.

	Higgs		Diboson			
Theoretical Input	μ_A	σ_A	μ_A	σ_A	$\frac{\mu_s-\mu_b}{\sigma_s^2+\sigma_b^2}$	$\int$ROC
Pythia(8) Default	1.204±0.006	1.00	0.049±0.013	1.01	0.81	0.79
MSTW 2008 LO	1.219±0.006	0.99	0.045±0.013	1.01	0.83	0.79
CTEQ6L1	1.214±0.006	0.99	0.020±0.013	1.03	0.84	0.80
Renormalization scale	1.154±0.006	1.01	0.090±0.013	1.00	0.75	0.77
Primordial k_T soft	1.206±0.006	1.00	0.034±0.013	1.02	0.82	0.79
Primordial k_T hard	1.212±0.006	0.99	0.047±0.013	1.02	0.82	0.79
Extended phasespace	1.205±0.006	0.98	0.110±0.028	1.05	0.76	0.78
Reduced acceptance	1.213±0.006	1.00	0.034±0.013	1.02	0.83	0.79
PowHeg+Pythia	1.244±0.009	0.98	0.042±0.015	1.06	0.84	0.80
Cumulative Dev.	0.067	0.03	0.082	0.07		

Table 22: Mean and width of Boosted decision tree score for signal (Higgs) and background (Diboson). The uncertainty on the mean reflects the generated amount of statistics for each model. The cumulative difference is calculated as the difference between any model and Pythia8 added in quadrature. The left column shows the Gaussian separation and the ROC integral of each distribution. The most influential changes are the choice of renormalization scale and the generator phase space cut values.

9.4 The Likelihood fit

The boosted decision tree method of distinguishing the production of a Higgs boson to the Standard Model diboson production in the four lepton final state has been seen to be successful and stable under different theoretical models.
In the statistical techniques section, it was described how a likelihood of a series of mass measurements could be obtained for the best fit and background only hypothesis and subsequently could be turned into a significance of a signal.
It is the object to incorporate the additional information obtained from the angular multivariate analysis into this likelihood estimation, maximizing the probability difference between the best fit and the background only hypothesis.

9.4.1 Mass and Boosted decision tree PDFs

A simple incorporation is possible if the four lepton mass and MVA variable can be considered uncorrelated. As was discussed in the 'training' and 'extension to full mass window' sections this is a reasonable assumption. The training on small mass intervals together with the mixing of signal inputs successfully decorrelates the BDT score from the invariant mass. The likelihood for a series of mass, BDT score measurements (m_i, A_i) and a(f, m) hypothesis can be calculated by multiplying the individual probabilities:

$$\begin{aligned} \mathcal{L}(f,m) &= \prod_i \left[f \cdot p^m_{Sig}(m_i;m) \cdot p^{A_{BDT}}_{Sig}(A_i) + (1-f) \cdot p^m_{Bkg}(m_i) \cdot p^{A_{BDT}}_{Bkg}(A_i) \right] \\ p^\alpha_{Bkg} &= \frac{\sigma_{DB}}{\sigma_{DB}+\sigma_{RB}} p^\alpha_{DB} + \frac{\sigma_{RB}}{\sigma_{DB}+\sigma_{RB}} p^\alpha_{RB} \ , \end{aligned} \tag{9.10}$$

where the background PDF is the diboson (DB) and reducible background (RB) PDFs added in correspondence to their cross sections.
A very nice feature of this is that it does not introduce further degrees of freedom into the mass fit, and the subsequent calculated significance is therefore directly compatible with the significance obtained in the mass only measurements.
The mass probability density functions were estimated by a fit to a Crystal Ball plus a Gaussian function for each mass point in the grid of signal samples, a sixth and a third order polynomial for the diboson and reducible background distributions running over the total mass window of the search. The signal mass distribution, where all normalization is absorbed into a, is given by:

$$p^m_{Sig}(m) = a\left(f \cdot \mathrm{CB}(m;\alpha,n,m_0,\sigma_{CB}) + (1-f)\cdot e^{\frac{(m-m_0)^2}{2\sigma_g^2}}\right) , \tag{9.11}$$

where the Crystal Ball PDF, constructed to help describe the tail of the electron momentum measurement, is given by:

$$\mathrm{CB}(m;\cdots) = \begin{cases} e^{-\frac{(m-m_0)^2}{2\sigma^2_{CB}}} & \text{if } \frac{(m-m_0)^2}{2\sigma^2_{CB}} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{|\alpha|^2}{2}} \cdot \left(\left(\frac{n}{|\alpha|} - |\alpha|\right) - \frac{(m-m_0)}{2\sigma_{CB}}\right)^{-n} & \text{if } \frac{(m-m_0)^2}{2\sigma^2_{CB}} \leq -\alpha \end{cases} \tag{9.12}$$

While this at first glance seems arbitrary, what is worth noting is that this is a Gaussian distribution, with a slowly decaying left tail. I.e. a Gaussian until the variable is α standard deviations to the left of the mean, and a power distribution decaying at rate $\left(\cdots \frac{(m-m_0)}{2\sigma_{CB}}\right)^{-n}$ hence forth. Its expression ensures that the function is continuous and differentiable per construction.

Figure 54 shows the fitted masses for the different signal samples used as input. It is seen that the fitting model gives a good description of the PDFs and is able to take into account the electron tail at lower masses than the peak.
If a mass hypothesis was tested that was not equal to a mass in the grid of generated signal distributions, the PDF corresponding to the mass closest to the hypothesis, was used where the central value was shifted accordingly. This of course assumes that the other parameters of the PDF changes slowly enough for this to give an accurate description. Figure 54 indicate, that this is reasonable. Moreover, if as is currently the case, around 20 candidates have been found in the 125 GeV mass area[11, fig.8], it seems unlikely that the fit will be able to distinguish two mass PDFs with widths of 2.6 GeV and 3.0 GeV as is the case for the 120 GeV and 125 GeV fits. The validity of this method will be tested below.
The diboson and reducible background probability density functions where given as:

$$\begin{aligned} p^m_{DB}(m) &= a_0 + a_1 m + a_2 m^2 + a_3 m^3 + a_4 m^4 + a_5 m^5 + a_6 m^6 \\ p^m_{RB}(m) &= a_0 + a_1 m + a_2 m^2 + a_3 m^3 \end{aligned} \tag{9.13}$$

Figure 55 (left) shows the fits for the diboson and reducible background distributions together with the expected distribution of a Higgs signal at 125 GeV, the fits returned $\chi^2/\mathrm{N_{DOF}}$

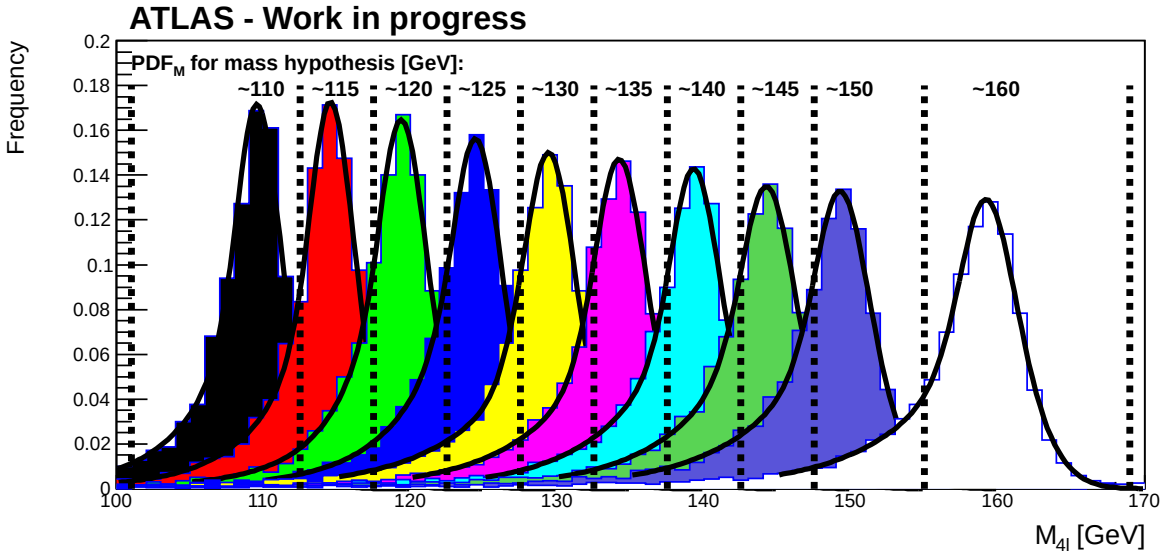


Figure 54: Mass fits using the Gaussian plus Crystal Ball function in the full mass range from 110 GeV to 160 GeV. The fitting function seemingly gives a good description of the distributions.

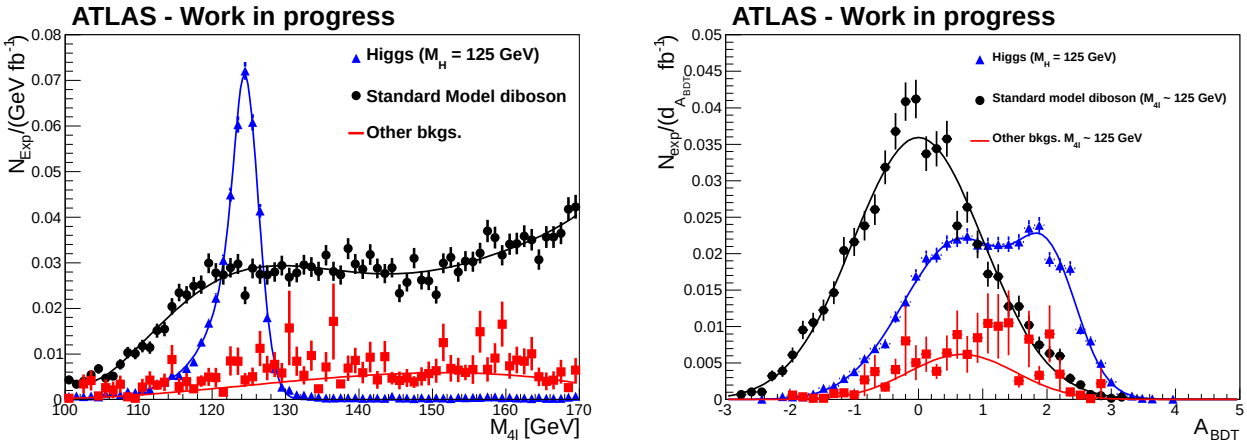


Figure 55: Mass and BDT score probability density function fits for a Higgs signal of 125 GeV and corresponding backgrounds. The Higgs mass distribution has been fitted with a combination of Gaussian and Crystal Ball function. The Standard Model diboson with a 6th order polynomial and the reducible backgrounds with a 3rd order polynomial. The two background BDT score distributions are fitted with a Gaussian, while the Higgs A_{BDT} distribution is fitted with two Gaussians. Notice that the reducible background distribution is a weighed sum of many samples with small but varying statistics, reflected in the highly changing error bars.

values of 95/18, 91/63 and 106/65 for Higgs, diboson and reducible backgrounds respectively. The large value for the signal fit, can partly be conferred to a poor description from 130 GeV to 135 GeV. The adequacy of the fitting functions will be discussed in the outlook section.

While the BDT score for the Standard Model diboson background was constructed to be a unit Gaussian, the signal distribution cannot be described accurately by this. The signal boosted decision tree distributions were fitted with two added Gaussian distribution, while both types of backgrounds were modeled by a single Gaussian. Figure 55 (right) shows the result for candidates with $m_{4l} \sim 125$ GeV. There is a good agreement between distribution and fit with $\chi^2/\mathrm{N_{DOF}}$ at 42/33, 60/35 and 29/28 for Higgs, diboson and reducible backgrounds respectively. Like the mass, the boosted decision tree score and corresponding signal/background probability for each candidate, was evaluated using the BDT structure closest to the candidate mass.

9.4.2 Significance estimation

Figure 56 shows a flow diagram expressing the different steps performed going from an experiment or a pseudo experiment to an estimate of the signal significance. The experiment will give a series of mass and BDT score measurements $\{m_i, A_{BDT\ i}\}$. Using the mass and BDT probability density functions found by fitting the generated samples, a probability for a given measurement either being signal or background can be calculated under a mass hypothesis. Combining the probabilities for the total sample of measurements, a likelihood can be calculated assuming a fraction of signal to background events.

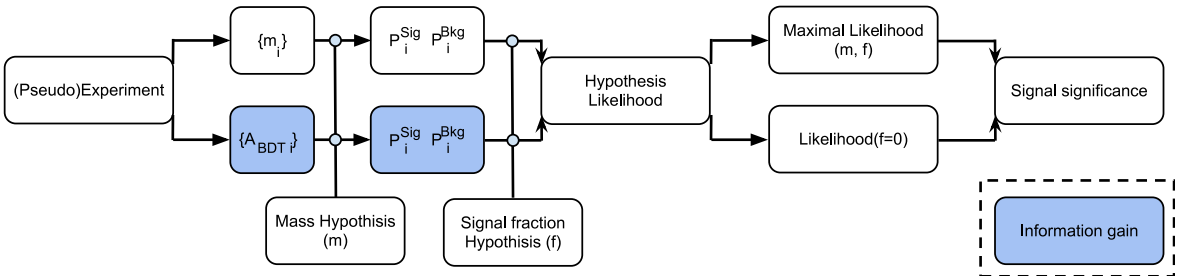


Figure 56: Maximal likelihood scan flowchart. Conducting the experiment, or a pseudo experiment, will give a sample of mass and boosted decision tree score measurements. Using the PDFs found by fitting the Monte Carlo simulation distributions, the probability for one event either being signal or background can be calculated for a given mass hypothesis. Assuming the ratio of signal to background events, a likelihood for the (m,f) hypothesis can be calculated. Two of these will be of special interest, namely the (m,f) that maximizes the likelihood and the f=0, i.e. the background only hypothesis. These are finally used to give a significance estimation. The information gain as compared to the mass only fit is depicted in blue.

Equation 9.10 shows maximization of the likelihood function will contain two free parameters, the fraction of signal to background events (f), and the invariant mass (m). Only having two free parameters, it is possible to avoid a complex fitting procedure, and a simple scan, where different (m,f) hypothesis are tested, will suffice. The mass window as described stretches from 110 GeV to 160 GeV, while the signal to background fraction per definition is confined between zero and one. It was deemed sufficient to use 1 GeV intervals while steps

in the fraction scan was performed in 0.025 intervals, resulting in 2000 different hypothesis to test.
In order to establish confidence in the method, a series of pseudo experiments were conducted. This is an effort to mimic what is to be expected from the actual experiment, and serves several purposes. First it is important to check whether the result is stable and that the pull distributions for mass and fraction measurements:

$$p(x_{meas}) = \frac{x_{meas} - x_{truth}}{\sigma_{meas}} \tag{9.14}$$

Return sensible distributions centered around zero and is of acceptable width.
Secondly it is possible to propagate the systematic uncertainties through the likelihood fit and finally it is possible to give estimates of how the signal significance will evolve in the future.
The pseudo experiments were performed by generating a number of events containing both signal and background. The number of each distribution was chosen as a random number following a Poissonian distribution with the expected cross section as mean. An example of such can be seen in figure 57, where the 'measured' mass and BDT score has been plotted. The plot can be seen as an instance of a Higgs signal, with invariant mass 125 GeV and a collected data sample of 5 fb^{-1}.
The (m,f) scan was performed on each pseudo experiment. The scan can be imagined as varying the peak center of the mass distribution, and changing the relative integral of the signal and background fits.
In the bottom of figure 57 shows the log-likelihood ratio is plotted for each (m,f) hypothesis.

The minimum, corresponding to the optimal (m,f) hypothesis is indicated by the two intersecting lines while the one and two σ_m, σ_f uncertainties are given by the dashed and full ellipses. The background only log-likelihood ratio can be seen in the bottom row, and is independent of mass. Comparing the minimum and the background only values will finally yield the significance of this pseudo experiment as described in the statistical methods section.
The experiment can now be run a number of times to test the pull distributions of the free parameters and see what the expected signal significance is. It is also of interest to compare these results to the mass only fit, to see if the multivariate analysis addition improves the analysis, and if so, by how much.
The mean and widths of the pull distributions can be found in table 23. The distributions where calculated for a 125 GeV Higgs signal at integrated luminosities of 5 fb^{-1}, 10 fb^{-1} and 20 fb^{-1}, to see if the measurements will improve as more statistics is gathered.
The uncertainties of the measurements were given by the one σ level in the likelihood scan, as in figure 57. The mass uncertainty was read at the found minimum for the fraction and vice versa.
It can be concluded that both variables will in general be measured within one standard deviation of its true value.
The fitted mass values will be measured too low, an effect that does not vanish for high statistics. It could be suspected that the asymmetric tail of the signal distribution may tend to include background events, but at low statistics this is not of great concern. The calculated mass uncertainty tends to improve with time.

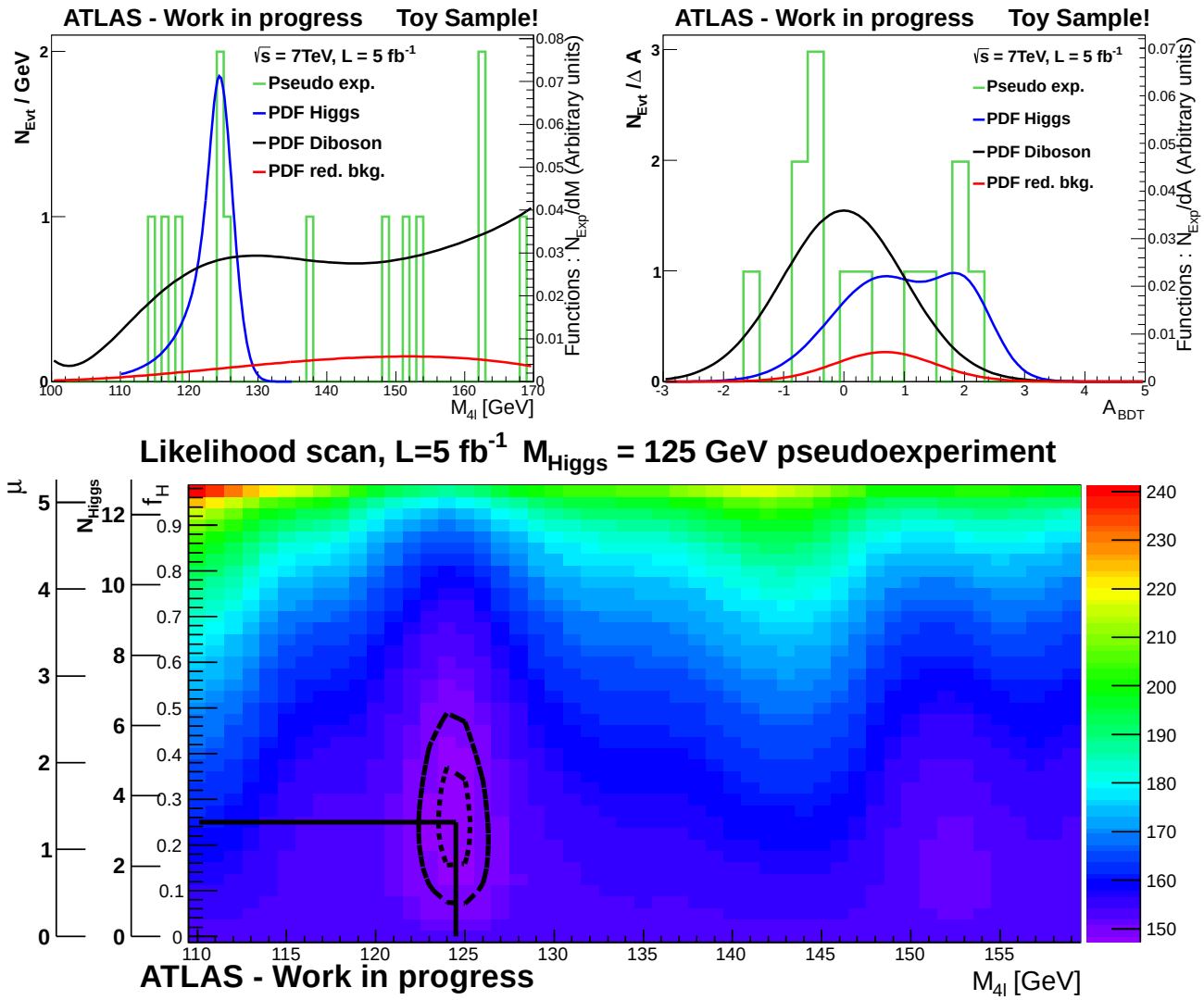


Figure 57: Top row: An example of a pseudo experiment. Here a number of events have been generated, for an integrated luminosity of 5 fb^{-1}, according to the background distributions and a signal at 125 GeV. The expected distributions are also shown. It should be noted that the integral of the fitting functions does not carry any physical interpretation, and only the relative signal to background size matters. It should also be noted that the BDT PDFs shown, are the ones corresponding to candidates of ∼125 GeV, candidates at other masses are evaluated using their corresponding PDFs. Bottom figure: Log likelihood ratio for each mass and fraction of signal to background hypothesis for the same pseudo experiment as above. The scan can be imagined, as varying the central value of the signal mass and relative integral of the signal and background PDFs in the figures above. The one and two sigma uncertainties on the measured mass and fraction are indicated by the full and dashed lines.

The fitted fractions of signal to background events is disposed to exceed the actual value, still

$M_H = 125$ GeV							
		5 fb^{-1}		10 fb^{-1}		20 fb^{-1}	
		μ	σ	μ	σ	μ	σ
Pull$_M$	m fit	-0.08 ± 0.09	0.86 ± 0.09	-0.29 ± 0.08	0.77 ± 0.07	-0.50 ± 0.11	0.89 ± 0.11
	m-A$_{BDT}$ fit	-0.36 ± 0.27	1.33 ± 0.29	-0.32 ± 0.07	0.75 ± 0.06	-0.64 ± 0.11	0.87 ± 0.13
Pull$_f$	m fit	0.62 ± 0.06	0.78 ± 0.06	0.34 ± 0.05	0.68 ± 0.04	0.26 ± 0.05	0.66 ± 0.04
	m-A$_{BDT}$ fit	0.40 ± 0.05	0.70 ± 0.05	0.25 ± 0.06	0.68 ± 0.04	0.18 ± 0.04	0.57 ± 0.03

Table 23: Means and widths of likelihood fit pull distributions. The distributions express how well the fitted mass and fraction of signal to background events compare to the input from the pseudo experiment. A standard deviation of one indicates that uncertainties of the parameters are well estimated

within one standard deviation. I suspect that the number is overshot due to the low statistics in the fit. If the number of signal events fluctuates low, the fit will probably tend to find an upwards fluctuation in the background somewhere else. The number of signal events are very low for a 125 GeV Higgs, and it is therefore not unlikely that a similar number of background events would cluster somewhere in the search window, introducing a bias in the positive direction [20] As an example, figure 57 (top left) shows that the background mass distribution has clustered around 152 GeV. This results in the valley seen to the right of the minimum in likelihood scan. Even such a small collection will result in a fraction of compatible size with the actual signal.

It is furthermore seen that the uncertainty on the estimated number of signal events is too large.

The mean and widths where estimated by fitting the pull distributions to a Gaussian. This was found to work well for the fraction distributions but the mass distributions presented some difficulties. In instances where there was a low fluctuation in the number of signal events, the fitted mass will be randomly distributed over the total search window, resulting in a flat contribution to the desired pull distribution. Moreover the fitted mass and uncertainties where based on the likelihood scan, having 1 GeV intervals between tests. This results in a too high probability for the found mass to be exactly equal to the input mass. A more thorough examination of the pull distribution would either use smaller distances between the mass test,s or even better fit, the valley in the likelihood distribution with a second order polynomial.

As a benchmark 200 pseudo experiments were created with Higgs masses at 125 GeV and 150 GeV with an integrated luminosities of 5 fb^{-1}, 10 fb^{-1}, corresponding to the statistics size in December 2011, July 2012. The experiment was of course run at 8 TeV rather than 7 TeV in 2012, and changes in the BDT variables, signal to background ratios etc. should be examined. The stability of the multivariate method under the change of experiment center of mass energy will be discussed in the outlook section. The only change imposed here was a change in the mean number of Higgs candidates, corresponding to the expected increase in

[20] The expected number of signal Higgs events at 5 fb^{-1} is 2.45, table 14. As reference, an Possionian with mean 2.45 will give 0 in 9% of instances, 1, 21% of the time and 25% of cases it will give 2.

Significance	$M_H = 125$ GeV		$M_H = 150$ GeV	
	5 fb^{-1}	10 fb^{-1}	5 fb^{-1}	10 fb^{-1}
m fit	1.38 ± 0.81	1.95 ± 1.03	2.75 ± 1.09	4.32 ± 1.23
m-A$_{BDT}$ fit	1.58 ± 1.00	2.35 ± 1.19	3.27 ± 1.21	5.08 ± 1.33
σ Gain	1.14	1.21	1.19	1.18

Table 24: Expected significances for the mass only likelihood fit and mass-BDT score likelihood fit methods, calculated from 200 pseudo experiments. The expectations are both given for Higgs signals at 125 GeV and 150 GeV, and 5 fb^{-1} and 10 fb^{-1} data samples. The gain indicates the expected increase in significance by the inclusion of the multivariate analysis.

cross section table [28].

The results found in table 24 finally shows the expected significances for both mass and mass-BDT score fits. The expected significance gain by the multivariate inclusion is 15%−20%, corresponding to an increase in statistics between 30% and 40%.

9.4.3 Systematic uncertainties and tests

In order to establish confidence in the method, a series of test were made, both to validate its stability and to propagate uncertainties. For simplicity and since it by all means is the most interesting, the following discussion will only focus on a 125 GeV Higgs signal.
The first series of systematic uncertainties to be propagated were the theoretical uncertainty of the boosted decision tree score. This was done by using the same mass and BDT PDF as found above, but the simulated input of the pseudo experiment were smeared with a Gaussian of mean and width as found in table 22. The smearing was only done on the Higgs signal and Diboson background, because these were the only known values and the reducible background contribution is small compared to the others. The signal and background were both scaled upwards and downwards and the likelihood fit was tested both in an optimistic scenario ($(\mu, \sigma)_{\mathrm{BDT}} \uparrow$), where the signal was scaled upwards while the background was scaled downwards and vice versa for a pessimistic scenario ($(\mu, \sigma)_{\mathrm{BDT}} \downarrow$).
Next a series of pseudo experiments were made, with the number of signal Higgs events scaled upwards and downwards 10% corresponding to the theoretical uncertainty of the Higgs cross section, table 2.
As mentioned above, if a candidate mass fell in between two points on the mass grid, the mass and BDT PDFs of the simulated mass closest to the candidate mass was used. To test the validity of this approach, the 125 GeV mass was removed from the initially fitted PDFs, while the pseudo experiment was given a 125 GeV Higgs input. This would reveal if biases were unintentionally introduced in e.g. the fitted mass. Notice that the distance to the nearest PDF is in this case 5 GeV rather than 2.5 GeV.
Finally the routine was run again with only background samples as input. This will reveal if significant peaks are systematically found in the background only sample, and at the same time also give indications of how the local significance for the peak can be extended to a global

$M_H = 125$ GeV at 5 fb^{-1}			
	Pull_M	Pull_f	σ_{local}
Nominal	-0.36±1.33	0.40±0.70	1.58±1.00
$(\mu,\sigma)_{\text{BDT}} \downarrow$	-0.16±0.84	0.36±0.72	1.47±0.98
$(\mu,\sigma)_{\text{BDT}} \uparrow$	-0.46±1.22	0.42±0.68	1.65±1.02
σ_{Higgs} low	-0.30±0.78	0.48±0.70	1.56±1.02
σ_{Higgs} high	-0.22±0.82	0.42±0.69	1.66±1.08
Removed grid point	-0.32±0.72	0.40±0.60	1.55±0.95
Combined $\mu \pm \sigma_\mu$	-	-	1.58±0.07
Bkg. Only	-	-	0.94±0.70

Table 25: Likelihood fit theoretical uncertainties, the listed value, from top to bottom are: Nominal; the theoretical prediction. $(\mu,\sigma)_{\text{BDT}} \uparrow$ and $\downarrow$, the BDT systematic uncertainties in an optimistic and pessimistic scenario, where in the first case the signal BDT scores have been scaled upwards and background downwards and both smeared as in table 22. In the pessimistic scenario the scores were scaled in the opposite direction. In σ_{Higgs} low and high the Higgs cross section has been scaled upwards and downwards by 10%, corresponding to the theoretical uncertainty of production cross section. In removed grid point, a 125 GeV Higgs input has been evaluated in a likelihood fit not containing the 125 GeV mass and BDT PDFs.

significance.
The results of these variations can be found in table 25.

It can be concluded that there is a good agreement between the results. The analysis systematic uncertainties are small compared to the statistical uncertainty reflecting that the number of signal and background are Poissonianly distributed.
Moreover it is seen that removing a grid point gives the correct value all the same, indicating that it is valid to use adjacent mass and BDT PDFs.
The background only pseudo experiment on average returns significances of $0.94\sigma \pm 0.7$. The signal and background expectations differ by $\sim 1\sigma$, i.e there is a $\sim$16% chance that the background will fluctuate upwards, giving a signal significance of the expected value.

10 Results

10.1 Selection extensions

In the outlook section it will be discussed how the multivariate analysis used for separation of Higgs→ZZ* →4l from ZZ* →4l events might be rewritten in an effort to measure the spin and parity of the recently discovered particle.
The differences between a spin 0 and a spin 2 Higgs boson, and likewise a boson with positive and negative parity, are naturally small, and for this reason it is of high interest to attain as many Higgs candidates as is achievable, still having as large ratio of signal to background events as possible.
Even though the use of the transitional radiation tracker has been surpassed, for now, by other methods of selection. It could be of interest for the further analysis to examine whether the boundary on the electron transverse momentum could be lowered, using the TRT electron identification.
If a successful and testable multivariate analysis is obtained for electron identification in the LAr end-cap and forward calorimeters, not covered by tracking, it could also increase the amount of Higgs candidates.
Not only will the reduction of allowed p_T and increase of allowed $|\eta|$ improve sensitivity, but also the variable measurements. Figure 51 indicates an example where this is the case. The imposed ATLAS acceptance changes both $\Delta\phi_{ZZ*}$ and η_H in the extremes of the variable domains. These are exactly the areas with most separation power[21], figures 38,40.

10.2 Separation from the Standard Model ZZ background

It has been found that it is possible to construct a multivariate analysis that is able to distinguish a Higgs signal from its irreducible ZZ* →4l background.
Using slight differences in production mechanisms and system angular momenta as input to a boosted decision tree, score distributions of the two types of events was produced with a ROC integral expressing their differences of 0.76.
It was estimated that the separation was stable to within 9.2% under changes in theoretical models and systematic changes.
The multivariate separation together with the invariant mass distributions was used in a likelihood fit, and the found signal significance was compared to the results obtained using only the mass. The expected significances were:

$$\begin{aligned} S_{\mathrm{m_H}} &= 1.38 \pm 0.81(\mathrm{stat.}) \\ S_{\mathrm{m_H-MVA}} &= 1.58 \pm 1.00(\mathrm{stat.}) \pm 0.07(\mathrm{syst.}) \end{aligned}$$

for a 5 fb^{-1} sample taken at $\sqrt{s} = 7$ TeV for the two types of analysis, suggesting an improvement of $\sim 15\%$ by the multivariate addition in the 2011 datasample. The statistical uncertainty reflects the Poissonainly distributed number of signal and background events. The results were stable under different theoretical models and tests of the likelihood fit method.

[21] at least in the Higgs-diboson separation

10.3 Unblinding the ATLAS 4.8 fb^{-1} at 7 TeV 2011 data sample

Having tested the boosted decision tree method and subsequent likelihood fit under various systematic changes, it is time to apply the analysis on real data. The data is the 2011 4.8 fb^{-1} at 7 TeV sample that followed the HSG2 set of requirements described above. For each candidate an invariant mass and boosted decision tree score was calculated. These were passed to the likelihood scan, in the same manner as the pseudo experiments.

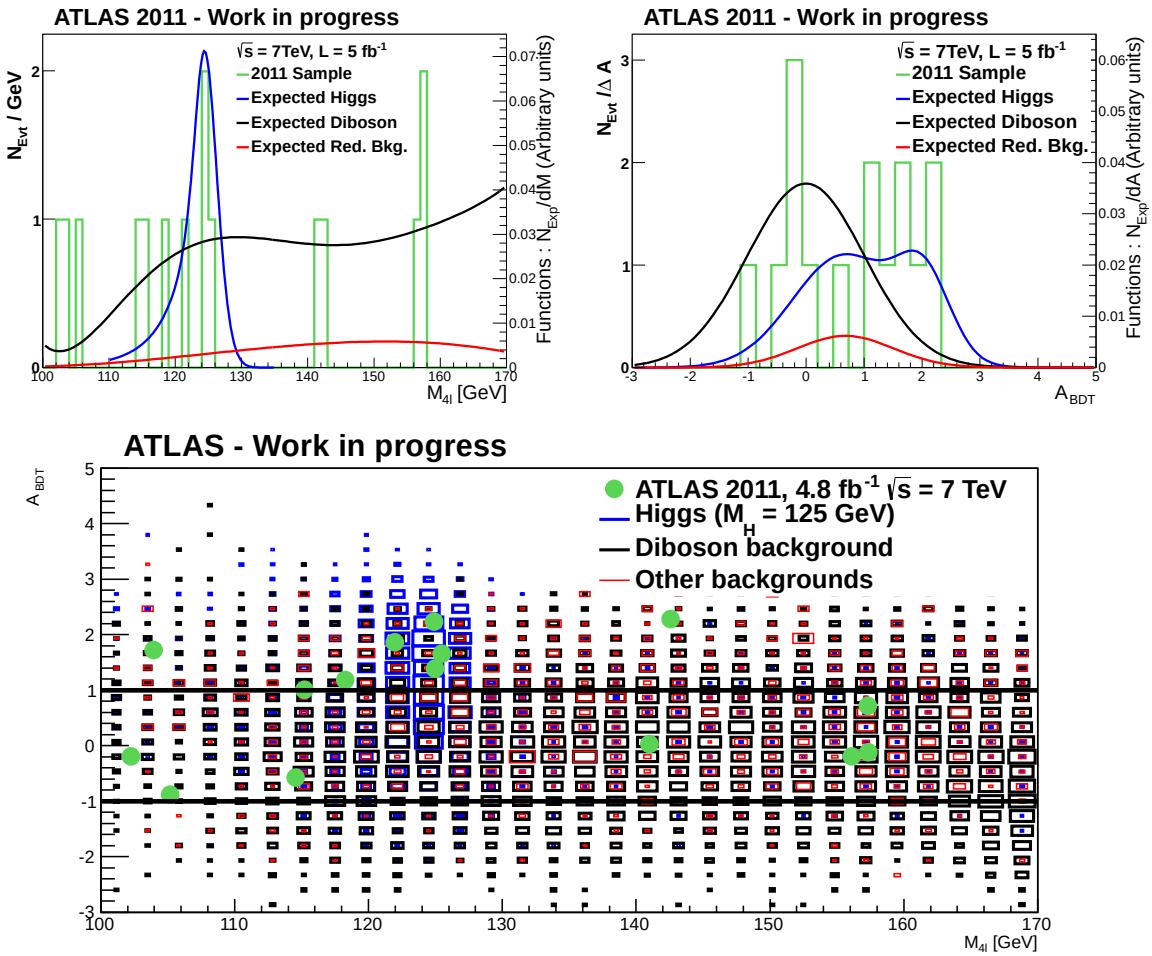


Figure 58: Top row: Mass and BDT score distributions for the 2011 ATLAS 4.8 fb^{-1} data set taken at a center of mass energy of 7 TeV (green). Together with the candidates is also shows the expected distributions for Standard Model Diboson production (black), the combined reducible backgrounds Z→ll+Jets, Z→ll+b$\bar{\text{b}}$+Jets and t$\bar{\text{t}}$ (four lepton filter) (red) and the predicted distribution of a Higgs signal at 125 GeV (blue). Bottom figure: Mass vs. BDT score plot for the same data sample and distributions as above. A cluster of events compatible with a Higgs signal is seen in the 125 GeV area.

Examining first the mass and boosted decision tree score, the individual distributions and

combined scatter plot are depicted in figure 58, together with the expected background distributions and the theoretically predicted distribution of a 125 GeV Higgs signal. Indications of a cluster of events in the 125 GeV area with BDT scores compatible with the Higgs distributions are seen.

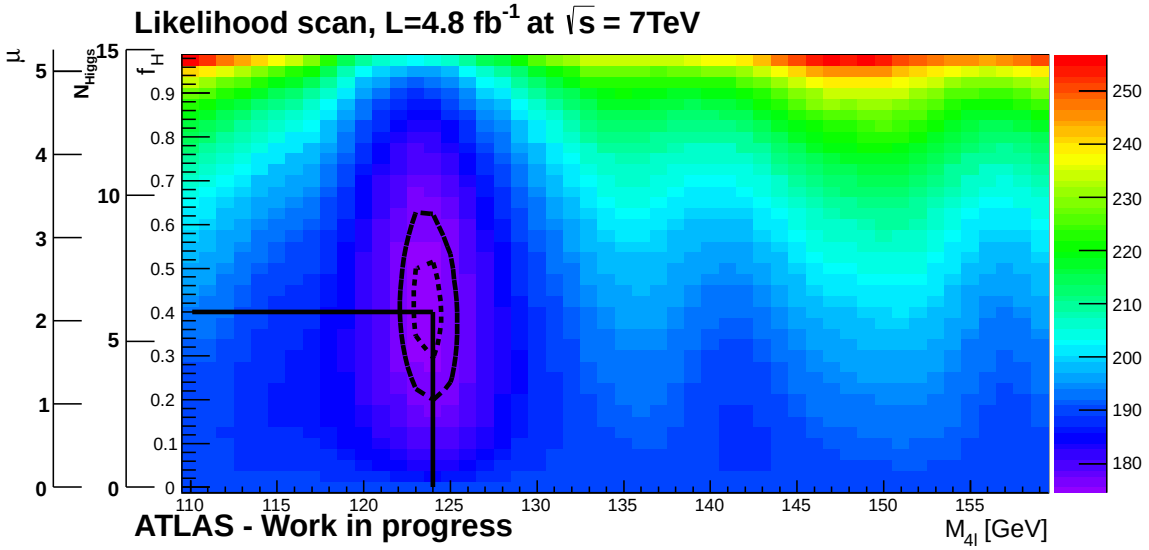


Figure 59: Mass-signal to background fraction likelihood scan of the 2011 ATLAS 4.8 fb^{-1} data set taken at a center of mass energy of 7 TeV. The dashed circles corresponds to the one and two standard deviation uncertainty bands while the line indicates the best fit values.

The likelihood scan was then performed on the data sample, figure 59.
Due to the large uncertainties in the expected number of reducible background events in table 14, the ratio of diboson events to reducible background events, in equation 9.10, where fixed at the value found in [11, table.7], still using the PDFs as above.
There is only one distinct valley around 125 GeV, with a signal to background fraction that is not compatible with the background only hypothesis. The resulting expected signal to background fraction is twice the Standard Model expectations. Assuming that the statistical uncertainty on the expected number of events is given by a Poissonian distribution this constitute an upwards fluctuation of 2 standard deviation of the number of Higgs events.

Finally the the local significance was calculated as a function of mass, where one distinct peak at 125 GeV clearly exceeds the background only expectations, figure 60. The inclusion of the multivariate analysis moreover gives a significance less compatible with the background only expectations.
It can be concluded that an excess of events at 124.0±1.2 GeV has been found in the 2011 data sample with with local significance of 3.3 for the mass BDT-score analysis. The corresponding value for the mass fit is 2.1. The zero signal expectations are 0.94±0.70 and 0.96±0.65 respectively. Taking the look elsewhere effect into account, estimated using 200 background

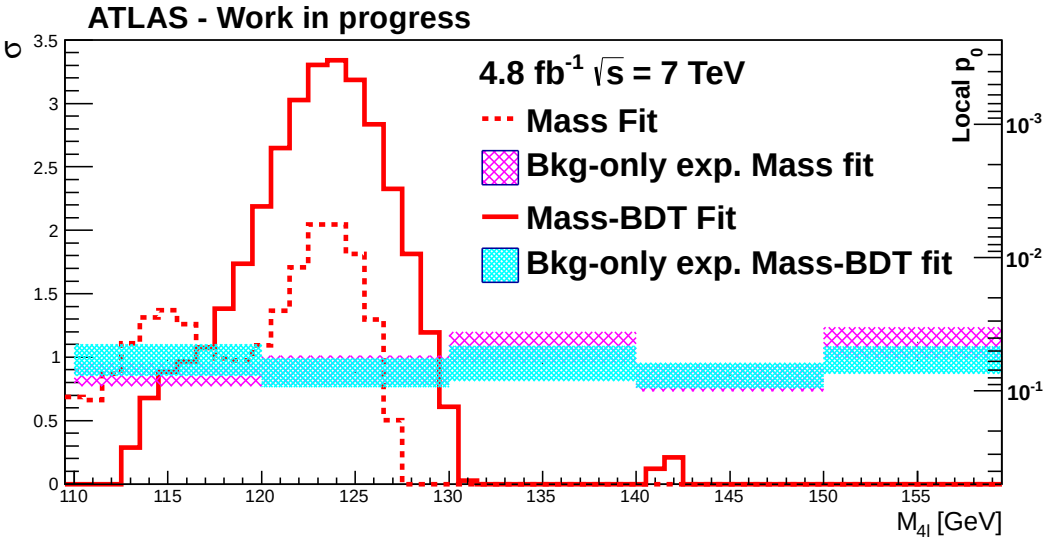


Figure 60: Local significance as function of mass for the 2011 ATLAS 4.8 fb^{-1} data set taken at a center of mass energy of 7 TeV. The dashed red line shows the values for the mass only fit, while the red line corresponds to the mass-BDT score fit. The magenta and cyan bands shows the expected significances in the background only case for the mass only and mass-BDT score fit respectively.

ATLAS, $\int$Ldt=4.8 fb^{-1} at $\sqrt{s} = 7$ TeV

	M_{4l}	p_0 (local)	σ (local)	$\langle N(c) \rangle$	p_0	σ
m fit	124.0±1.2	0.02	2.1	*0.15*	*0.17*	*1.0*
m-A$_{BDT}$ fit	124.0±1.2	0.0004	3.3	0.0037	0.004	2.6

Table 26: Results for the 2011 data set. Note that the used method for calculating the look elsewhere effect is approximate, and possibly overshoots the effect at low significances as found with the mass only fit, [20, fig.2]

only pseudo experiments, as described in the statistical techniques section, the significances for the mass-BDT score distribution is 2.6σ corresponding to a probability p_0=0.004, table 26.

11 Conclusion

The search for the Higgs boson in the four lepton final state poses two difficulties, low statistics and the irreducible background from the Standard Model diboson production in the four lepton final state.
The optimization of the Higgs selection and background rejection, in the ATLAS detector, has been of very high priority, and as such it has been seen that it is at this point very difficult to improve.
The use of a multivariate analysis to help distinguish the production Higgs from the Standard Model production of ZZ^* in the four lepton final state has been proven promising.
Applying the method on the data sample collected by the ATLAS detector in 2011 gives indication of a Higgs boson with an invariant mass of 124 GeV.

12 Outlook

12.1 Improvements

There are several ways the addition of the multivariate analysis to the search for the Higgs boson could be optimized to extract further information from the data, and give more reliable results.
One hindrance of the analysis was the lack of simulation statistics, most notably of the backgrounds. In figure 48 it was shown that the boosted decision tree training was performed on a window extending 10 GeV in either direction of the simulated mass. This choice was made due to two conflicting interests. In one perspective, the more events that is used in the training, the more complex the boosted decision tree is able to be, without being over trained. From another perspective, if the window is too wide, mass dependent effects might possibly enter such that the mass and BDT score can no longer be considered uncorrelated, rendering the likelihood fit method invalid.
The discovery of the new particle state at $\sim$126 GeV [12] will surely motivate the simulation of relevant signal and background samples in the time to come.
Figure 57 shows that the reducible background BDT score distributions tend to lie between the Higgs and Standard Model diboson distributions. One interesting possibility would be to include these into the multivariate separation.
This is at the moment problematic. When running the reducible background samples through the $H \rightarrow ZZ^{(*)} \rightarrow 4l$ selection, only a total $\sim$1000 events over the total search window will pass. It these are again divided into smaller mass windows for the BDT training it will become very sensitive to overtraining. This is only solved by reducing the complexity of the BDT structure at the cost of separation power.

Throughout the analysis the Higgs - Standard Model diboson separation was considered a single entity both in separation and in the significance estimation. For one thing, table 10 shows that the kinematical constraints differ somewhat for electrons and muons, while [11, table.7] both shows that the muon and electron efficiencies, $\epsilon 4e/\epsilon 4\mu = 0.37$, and the signal to

background ratios, $(S/B)_{4e} = 0.1$, $(S/B)_{4\mu} = 0.2$, differ remarkably for the 2011 sample in the low mass region. The $2e2\mu$ and $2\mu2e$ lies somewhere in between these two.
An interesting refinement would be to divide the analysis into three separate categories $4e$, $2e2\mu$+$2\mu2e$ and 4μ. In the 2011 data sample 6 4μ candidates where found in the 110 GeV to 160 mass region. In the method described, their signal likelihood estimation was subdued by a too large estimation of reducible background to Standard Model diboson events. The number of signal to total number of background events where considered a free parameter in the significance estimation and will thus not be affected. Due to the fact that the mass distributions of the Standard Model diboson and reducible backgrounds are not distinguishable by the low statistics of candidates, the division into lepton categories would probably not improve the mass fit. Considering figure 57 on the other hand, the categorizing of lepton final states could potentially improve the BDT score fit.
Moreover the analysis could be further refined by dividing the reducible backgrounds into the appropriate Z→ll+Jets, Z→ll+$b\bar{b}$ and $t\bar{t}$ categories. By not fitting samples combined of many small datasamples with highly varying statistics, this could certainly improve the PDF fit of reducible backgrounds, figure 55.

Related to this, the fitting functions used for the Higgs and diboson distributions, have also been found to be problematic. The diboson mass distribution was fitted with a sixth order polynomial, and gave a reasonable $\chi^2/NDF \simeq 1.5$. The χ^2 is usually a good indicator of the performance of the fit model, but in this case it falls short of an important point. There is a clear tendency for histogram bins to lie over the fitting function in the interval 115 GeV to 122 GeV. It was tried to change the order of polynomial as well as reducing the mass window, but a model that adequately described this feature was not found.
By eye measure it seems that the Higgs mass distribution fits gives a good description of the data, but returned $\chi^2/NDF \simeq 5$ for the 125 GeV fit. If given more time a constant term added to the signal mass PDF would probably give a better description.
Figure 61, a greatly enhanced version of figure 55, summarizes the difficulties in the fitting procedure, .

12.2 Towards the 2012 data sample

All simulations and tests so far has only been made using simulations reflecting the 2011 $\sqrt{s} = 7$ TeV conditions. The change of center of mass energy in 2012 to $\sqrt{s} = 8$ TeV could potentially change the input distributions of the multivariate analysis, making its predictions biased.
For instance, the relative cross sections on backgrounds may be influenced quite significantly by the change of center of mass energy. At the same time, the total expected number of events also change [11, table 7].
All parameters used in the multivariate analysis that are related to the Higgs decay are not expected to change[22]. The rapidity distribution is expected to have a logarithmical dependence on center of mass energy [27, e.q 3.7], while the system transverse momentum will definitely

[22] For the signal at least

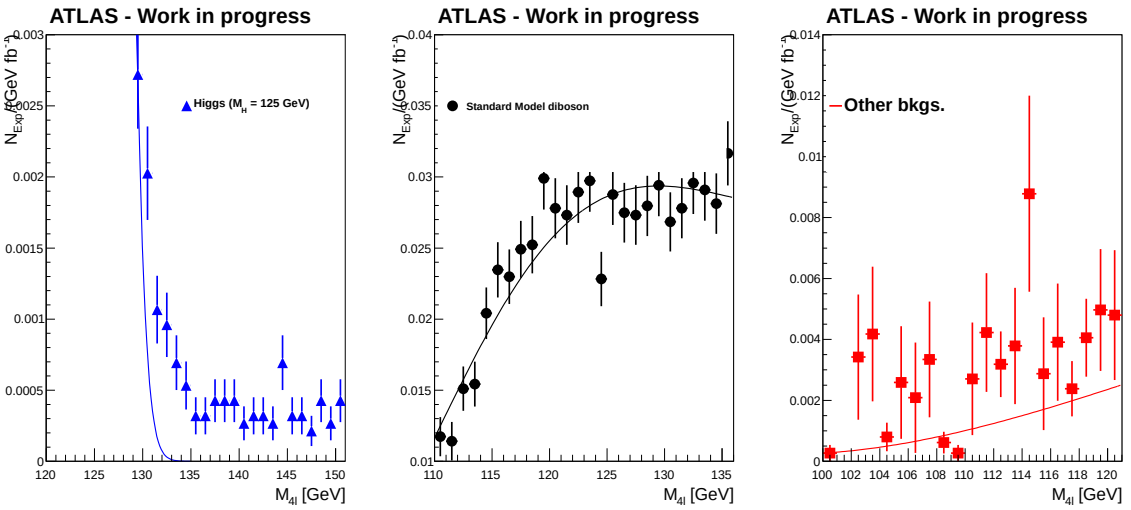


Figure 61: Problems involved in the mass PDF fitting procedure. The figures are a zoom of figure 55. Left: The mass PDF has problems with describing the tail distribution at masses $\gtrsim 10$ GeV from it central value. Middle: The sixth order polynomial used to describe the diboson invariant mass distribution has difficulties with datapoints systematically being over or under. Right: The greatly varying statistics size of the samples used for the reducible background distribution renders this method arguable.

also change due to the change in the ratio of gluon fusion to vector boson fusion produced Higgs.
While the analysis is set up, it would be a simple change to run everything again on simulations created for 2012 conditions, but not having the samples ready at hand it was deemed to dauntless to unblind the 2012 data sample using expectations derived under 2011 conditions.

12.3 Future: Measurement of spin and parity

Some of the parameters that entered the multivariate analysis made explicitly use of the fact that the Higgs boson is a spin zero particle, while the angular momentum of the background is in general not confined. This is most clearly seen in $|\cos(\theta_{HZ})|$, which measures the direction of decay of the Z boson with respect to the Higgs boson direction of flight. Being a spin zero particle the Higgs boson will have no sense of direction when decaying and as expected the $|\cos(\theta_{HZ})|$ distribution is completely flat.
Seeing that the diphoton channel indicates that the non integer and spin one cases are not allowed [12]. It is still possible that the newly discovered particle carries spin two instead of spin zero. The multivariate approach could help distinguish these two cases, figure 62 (left). Moreover some parameters are also influenced by the parity of the examined particle, 62 (right).
This possible extension of the analysis encounters a common problem in multivariate analysis. The MVA methods discussed can only maximize the difference between two types of input,

but if a candidate is found, there are now four distinct possibilities: Higgs boson, spin two boson, Standard Model diboson or a reducible background. It is the object to distinguish the first two in a data sample who, will also contain the last two as backgrounds.
One possibility, in e.g. the spin separation, would be to divide the MVA into two distinct categories, one independent of spin and one spin dependent. The first could then be used to separate the signal boson from backgrounds while the second could be used to separate the spin zero from the spin one case. It is however not clear if any of the variables are completely independent of particle spin, and that a separation subsequently can be obtained using only a small number of variables.
One other possibility would be to use all variables in the same MVA separating the spin zero and spin two hypothesis, and examine the background distributions in the subsequent BDT structures, much as was done with the reducible backgrounds in the analysis.

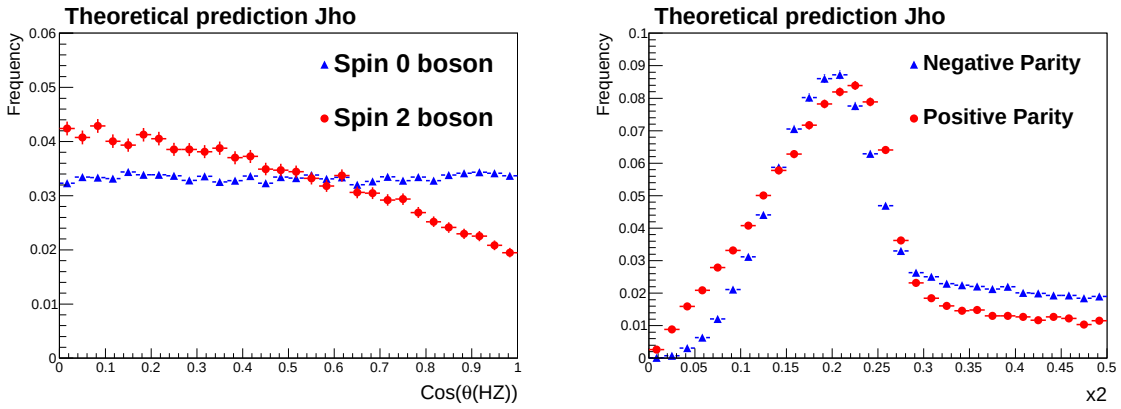


Figure 62: Left: Theoretically predicted $|\cos\theta_{HZ}|$ distribution a for spin-0 (blue) and a spin-2 boson (red). Right: Theoretically predicted x_2 distribution for a negative (blue) and positive (red) parity spin 0 boson. The distributions are the predicted PDFs without taking any detector effects into account. The expected separation due to these two distributions is therefore much lower than is what suggested by the figure.

The spin and parity measurements will be of highest priority in determining if the discovered particle state is indeed the Higgs, and the prospect of determining them using the angular distributions and production mechanisms in the four lepton final state is by all standards an exiting study for the time to come.

13 Acknowledgment

I would like to acknowledge all that through the last year have helped make this work possible. Thank you:

The Discovery Center and especially the HEP group
For education, facilities and in general most excellent working environment. It has been a great pleasure working here for the last year. Special thanks to my various office comrades for input, discussion and sociality, and Alex for providing simulations of spin two and negative parity bosons.

Andreas Hoeker
For expert insight in the subject of multivariate analysis techniques, and discussions on the theoretical error propagation.

Peter Skands
For expert counseling on physics simulation and theoretical uncertainties. Without your insight, the evaluation of theoretical uncertainties would have been in vain.

Stefano Rosati and Konstantinos Nikolopoulos
For endless discussions on multivariate analysis variables and their implementation into the $H \rightarrow ZZ^* \rightarrow 4l$ analysis.

Troels C. Petersen
Last but not least, thank you for the last year.
For inspiration on all aspects of particle physics and excellent counseling.
For giving me enough freedom to make my own errors, while directing me on the right path with perfect timing.
And for giving me the opportunity to work on the Higgs search in the most exiting times imaginable, it has been both an honor and a privilege.

References

[1] J.Stelzer J.Therhaag E. von Toerne H. Voss A.Hoecker, P.Speckmayer. TMVA 4, Toolkit for Multivariate Data Analysis with ROOT Users Guide. *arXiv:physics/0703039v5*, 2009.

[2] The ATLAS Collaboration Antonio Salvucci. Measurement of muon momentum resolution of the ATLAS detector. *arXiv:1201.4704v1*, 2012.

[3] G. Shaw B.R. Martin. *Particle Physics*. WILEY, 2009.

[4] The ATLAS collaboration C. Bini. Study of the performance of the ATLAS Muon Spectrometer. *ATLAS note:ATL-MUON-PROC-2011-008*, 2011.

[5] ATLAS collaboration. ATLAS public results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic`.

[6] ATLAS collaboration. EGamma - Electron identification. `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronIdentification`.

[7] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *IOPscience:2008 JINST 3 S08003 ,Institute of Physics Publishing and SISSA*, 2008.

[8] The ATLAS collaboration. Expected electron performance in the ATLAS experiment. *ATLAS note:ATL-PHYS-PUB-2011-006*, 2011.

[9] The ATLAS collaboration. Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011. *ATLAS-CONF-2011-116*, 2011.

[10] The ATLAS collaboration. Muon reconstruction efficiency measurement in the ATLAS experiment. *ATLAS note:ATL-PHYS-PROC-2012-11*, 2012.

[11] The ATLAS collaboration. Observation of an excess of events in the search for the Standard Model Higgs boson in the H$\rightarrow$ZZ$^{(*)} \rightarrow 4l$. *ATLAS Note:ATLAS-CONF-2012-092*, 2012.

[12] The ATLAS collaboration. Observation of an Excess of Events in the Search of the Standard Model Higgs boson with the ATLAS detector at the LHC. *ATLAS-CONF-2012-093*, 2012.

[13] The ATLAS collaboration. Search for excited electrons and muons with 5/fb of proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *ATLAS Note:ATLAS-CONF-2012-008*, 2012.

[14] The CDF Collaboration. Precise measurement of the W-boson mass with the CDF II detector. *Phys. Rev. Lett. 108, 151803(2012), arXiv:1203.0275*, 2012.

[15] The CMS Collaboration. Measurement of the weak mixing angle with the Drell-Yan process in proton-proton collisions at the LHC. *arXiv:1110.2682*, v2, 2011.

[16] Electron Gamma combined performance group on behalf of the ATLAS collaboration. Electron Indentification. `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronIdentification`.

[17] Electron Gamma combined performance group on behalf of the ATLAS collaboration. TechnicalitiesForMedium1. `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TechnicalitiesForMedium1`.

[18] Glen Cowan. *Statistical Data Analysis*. Oxford University Press Inc. New York, 1998.

[19] E.H. Simmons C.T. Hill. Strong Dynamics and Electroweak Symmetry Breaking. *arXiv:hep-ph/0203079v3*, v3, 2003.

[20] Ofer Vitells Eilam Gross. Trial factors for the look elsewhere effect in high energy physics. *arXiv:1005.1891v3*, 2010.

[21] J. Beringer et. *al* (Particle Data Group). DEVELOPMENTS IN HEAVY QUARKONIUM SPECTROSCOPY. *Phys. Rev. D86, 010001*, 2012.

[22] J.Beringer et. *al.* (Particle Data Group).

[23] R. Brout F. Englert. BROKEN SYMMETRY AND THE MASS OF GAUGE VECTOR MESONS. *PHYSICAL REVIEW LETTERS*, 13(9):321–323, 1964.

[24] Richard Fernow. *Introduction to experimental particle physics*. Press Syndicate of the University of Cambridge, 1986.

[25] Forthhommel. Standard Model. `http://en.wikipedia.org/wiki/CERN`, Source:**CERN**.

[26] J. Goldstrone. Field Theories with <<Superconductor>> Solutions. *IL NUOVO CIMENTO*, 19(1):154–164, 1961.

[27] Dan Green. *High P_T Physics at Hadron Colliders*. Cambridge University Press, 2005.

[28] LHC Higgs Cross Section Working Ground. Handbook of LHC Higgs Cross Sections 1. Inclusive Obervables), VOLUME = v6, year = 2011, journal = arXiv:1101.0593v3.

[29] LHC Higgs Cross Section Working Ground. Handbook of LHC Higgs Cross Sections 2. Differential Distributions), YEAR = 2012, journal = arXiv:1201.3084v1.

[30] The LEP Electroweak Working Group. `http://lepewwg.web.cern.ch/LEPEWWG/`.

[31] T.W.B. Kibble G.S. Guralnik, C.R. Hagen. GLOBAL CONSERVATION LAWS AND MASSLESS PARTICLES. *PHYSICAL REVIEW LETTERS*, 13(20):585–587, 1964.

[32] Peter W. Higgs. BROKEN SYMMETRIES AND THE MASSES OF GAUGE BOSONS. *PHYSICAL REVIEW LETTERS*, 19(16):508–509, 1964.

[33] Sofia Andringa Jorge C. Romão. Vector Boson decays of the Higgs Boson. *arXiv:hep-ph/9807536v1*, 1998.

[34] J.W.F.Valle. Neutrino masses and oscillations. *arXiv:hep-ph/0509262v1*, 2005.

[35] M. Muhlleitner R. Noriega-Papaqui I. Pedraza M. Sprira P.M. Zerwas M. Gomez-Bock, M. Mondragon. Electroweak Symmetry Breaking and Higgs Physics: Basic Concepts. *Journal of Physics: Conference Series*, 18:74–135, 2005.

[36] A. Vikhlinin L. David W. Forman C. Jones S. Murray W. T. Tucker M. Markevitch, A.H. Gonzalez. Direct constrains on the dark matter self-interaction cross-section from the mergin galaxy cluster 1E0657-56. *arXiv:astro-ph/0309303v2*, 2003.

[37] S.P. Martin. A Supersymmetry Primer. *arXiv:hep-ph/9709356v6*, v6, 2011.

[38] D.V. Schroeder M.E. Peskin. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[39] Stannered MissMJ, TriTertButoxy. Standard Model. `http://en.wikipedia.org/wiki/Standard_Model`.

[40] Higgs Working Group 2 on behalf of the ATLAS collaboration. HSG2 Higgs WG Subgroup: H→ZZ$^{(*)}$ and Z(H→invisible). `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HiggsZZstar`.

[41] P. Wagner on behalf on the ATLAS Collaboration. Performance of the ATLAS Transition Radiation Tracker Readout with High Energy Collisions at the LHC. *ATLAS note:ATL-INDET-PROC-2011-040*, 2011.

[42] P. Lebrun S. Myers O. Ranko J. Poole P. Proudlock O.S. Bruning, P. Collier. LHC Design Report. *CERN-2004-003-V-1*, 2004.

[43] Carlo Oleari Paolo Nason. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *arXiv:0911.5299v2 [hep-ph]*, 2010.

[44] Carlo Oleari Paolo Nason. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *arXiv:0911.5299v2 [hep-ph]*, 2010.

[45] Gioacchino Ranucci. The profile likelihood ratio and the look elsewhere effect in high energy physics. *arXiv:1201.4604v1*, 2012.

[46] M.Besancon M.Chemtob A.Deandrea E.Dudas P.Fayet S.Lavignac G.Moreau E.Perez Y.Sirois R.Barbier, C.Berat. R-parity violating supersymmetry. *arXiv:hep-ph/0406039v2*, 2005.

[47] Peter Skands. Private conversations.

[48] Raoul Rontsch Giulia Zanderighi Tom Melia, Paolo Nason. W+W-, WZ and ZZ production in the POWHEG BOX. *arXiv:1107.5051v3 [hep-ph]*, 2011.

[49] David Tong. Lectures on String Theory. *arXiv:0908.0333*, v3, 2012.

[50] Peter Skands Torbjorn Sjostrand, Stephen Mrenna. A brief introduction to pythia 8.1. *arXiv:0710.3820v1 [hep-ph]*, 2007.

[51] Richard Corke Stephen Mrenna Stefan Prestel Peter Skands Torbjorn Sjostrand, Stefan Ask. Pythia 8.1.60 manual. Included in pythia download, `http://home.thep.lu.se/~torbjorn/pythia8/pythia8165.tgz` (.../htmldoc/Welcome.html therein).

[52] Durham University. The Durham HepData Project. `http://hepdata.cedar.ac.uk/pdf/pdf3.html`.

[53] Steven Weinberg. A MODEL OF LEPTONS. *PHYSICAL REVIEW LETTERS*, 19(21):1264–1266, 1967.

A Monte Carlo Sample list

Background Samples

Hard Process		(N_{evt})	σ [fb]
$ZZ^* \rightarrow 4l$	mc11_7TeV.109292.Pythiazz4l_3MultiLeptonFilterElecMu	747957	91.54
$t\bar{t}$ (4l)	mc11_7TeV.109346.T1_McAtNlo_Jimmy_4LepMass_Mll60GeV12GeV	297947	515.2
$Z\rightarrow ee + N_J$	mc11_7TeV.107650.AlpgenJimmyZeeNp0_pt20	500000	827375
	mc11_7TeV.107651.AlpgenJimmyZeeNp1_pt20	429999	166625
	mc11_7TeV.107652.AlpgenJimmyZeeNp2_pt20	1944195	50375
	mc11_7TeV.107653.AlpgenJimmyZeeNp3_pt20	549949	14000
	mc11_7TeV.107654.AlpgenJimmyZeeNp4_pt20	134948	3375
	mc11_7TeV.107655.AlpgenJimmyZeeNp5_pt20	35000	1000
$Z\rightarrow \mu\mu + N_J$	mc11_7TeV.107660.AlpgenJimmyZmumuNp0_pt20	499999	822125
	mc11_7TeV.107661.AlpgenJimmyZmumuNp1_pt20	99996	166000
	mc11_7TeV.107662.AlpgenJimmyZmumuNp2_pt20	1719943	49500
	mc11_7TeV.107663.AlpgenJimmyZmumuNp3_pt20	549896	13875
	mc11_7TeV.107664.AlpgenJimmyZmumuNp4_pt20	120000	3500
	mc11_7TeV.107665.AlpgenJimmyZmumuNp5_pt20	50000	1000
$Z\rightarrow \tau\tau + N_J$	mc11_7TeV.107670.AlpgenJimmyZtautauNp0_pt20	499950	828125
	mc11_7TeV.107671.AlpgenJimmyZtautauNp1_pt20	279746	167375
	mc11_7TeV.107672.AlpgenJimmyZtautauNp2_pt20	1004847	50375
	mc11_7TeV.107673.AlpgenJimmyZtautauNp3_pt20	409997	13750
	mc11_7TeV.107674.AlpgenJimmyZtautauNp4_pt20	105000	3500
	mc11_7TeV.107675.AlpgenJimmyZtautauNp5_pt20	45000	1000
$^{\dagger}Z\rightarrow ee + b\bar{b} + N_J$	mc11_7TeV.116960.AlpgenHWfZeebbNp0_4LepM	150000	28.898
	mc11_7TeV.116961.AlpgenHWfZeebbNp1_4LepM	30000	26.324
	mc11_7TeV.116962.AlpgenHWfZeebbNp2_4LepM	11500	14.707
	mc11_7TeV.116963.AlpgenHWfZeebbNp3_4LepM	1500	10.226
$^{\dagger}Z\rightarrow \mu\mu + b\bar{b} + N_J$	mc11_7TeV.116965.AlpgenHWfZmumubbNp0_4LepM	149500	30.122
	mc11_7TeV.116966.AlpgenHWfZmumubbNp1_4LepM	90000	27.534
	mc11_7TeV.116967.AlpgenHWfZmumubbNp2_4LepM	11500	14.722
	mc11_7TeV.116968.AlpgenHWfZmumubbNp3_4LepM	1500	11.114

Table 27: List of Monte Carlo samples used in the analysis, cross sections are taken from [40, Summer 2012, Useful cross sections]. † indicates a that a four lepton filter has been imposed prior to analysis.

Signal Samples

Hard Process			(N_{evt})	σ [fb]
gg→H→4l	M_H =110 GeV	mc11_7TeV.116761.PowHegPythia_ggH110_ZZ4lep	49950	0.944
	M_H =115 GeV	mc11_7TeV.116762.PowHegPythia_ggH115_ZZ4lep	50000	1.69
	M_H =120 GeV	mc11_7TeV.116763.PowHegPythia_ggH120_ZZ4lep	50000	2.81
	M_H =125 GeV	mc11_7TeV.116764.PowHegPythia_ggH125_ZZ4lep	70000	4.26
	M_H =130 GeV	mc11_7TeV.116764.PowHegPythia_ggH125_ZZ4lep	50000	5.89
	M_H =135 GeV	mc11_7TeV.116765.PowHegPythia_ggH135_ZZ4lep	49999	7.46
	M_H =140 GeV	mc11_7TeV.116766.PowHegPythia_ggH140_ZZ4lep	50000	8.66
	M_H =145 GeV	mc11_7TeV.116767.PowHegPythia_ggH145_ZZ4lep	50000	9.24
	M_H =150 GeV	mc11_7TeV.116768.PowHegPythia_ggH150_ZZ4lep	49999	8.95
	M_H =160 GeV	mc11_7TeV.116770.PowHegPythia_ggH160_ZZ4lep	49996	3.86
VBF H→4l	M_H =110 GeV	mc11_7TeV.125063.PowHegPythia_VBFH110_ZZ4lep	30000	0.0665
	M_H =115 GeV	mc11_7TeV.125064.PowHegPythia_VBFH115_ZZ4lep	30000	0.124
	M_H =120 GeV	mc11_7TeV.125065.PowHegPythia_VBFH120_ZZ4lep	30000	0.214
	M_H =125 GeV	mc11_7TeV.125066.PowHegPythia_VBFH125_ZZ4lep	30000	0.337
	M_H =130 GeV	mc11_7TeV.125067.PowHegPythia_VBFH130_ZZ4lep	30000	0.481
	M_H =135 GeV	mc11_7TeV.125068.PowHegPythia_VBFH135_ZZ4lep	30000	0.627
	M_H =140 GeV	mc11_7TeV.125069.PowHegPythia_VBFH140_ZZ4lep	30000	0.751
	M_H =145 GeV	mc11_7TeV.125070.PowHegPythia_VBFH145_ZZ4lep	30000	0.823
	M_H =150 GeV	mc11_7TeV.125070.PowHegPythia_VBFH150_ZZ4lep	30000	0.819
	M_H =160 GeV	mc11_7TeV.125072.PowHegPythia_VBFH160_ZZ4lep	30000	0.094
WH→4l	M_H =110 GeV	mc11_7TeV.125265.PythiaWH110_ZZ4lep	20000	0.0416
	M_H =115 GeV	mc11_7TeV.125266.PythiaWH115_ZZ4lep	20000	0.0702
	M_H =120 GeV	mc11_7TeV.125267.PythiaWH120_ZZ4lep	20000	0.111
	M_H =125 GeV	mc11_7TeV.125268.PythiaWH125_ZZ4lep	20000	0.159
	M_H =130 GeV	mc11_7TeV.125269.PythiaWH130_ZZ4lep	20000	0.209
	M_H =135 GeV	mc11_7TeV.125270.PythiaWH135_ZZ4lep	19999	0.25
	M_H =140 GeV	mc11_7TeV.125271.PythiaWH140_ZZ4lep	20000	0.276
	M_H =145 GeV	mc11_7TeV.125272.PythiaWH145_ZZ4lep	20000	0.28
	M_H =150 GeV	mc11_7TeV.125273.PythiaWH150_ZZ4lep	20000	0.256
	M_H =160 GeV	mc11_7TeV.125275.PythiaWH160_ZZ4lep	20000	0.0973
ZH→4l	M_H =110 GeV	mc11_7TeV.125425.PythiaZH110_ZZ4lep	20000	0.0225
	M_H =115 GeV	mc11_7TeV.125426.PythiaZH115_ZZ4lep	20000	0.0382
	M_H =120 GeV	mc11_7TeV.125427.PythiaZH120_ZZ4lep	20000	0.0608
	M_H =125 GeV	mc11_7TeV.125428.PythiaZH125_ZZ4lep	20000	0.0878
	M_H =130 GeV	mc11_7TeV.125429.PythiaZH130_ZZ4lep	19999	0.116
	M_H =135 GeV	mc11_7TeV.125430.PythiaZH135_ZZ4lep	20000	0.14
	M_H =140 GeV	mc11_7TeV.125431.PythiaZH140_ZZ4lep	19950	0.155
	M_H =145 GeV	mc11_7TeV.125432.PythiaZH145_ZZ4lep	20000	0.158
	M_H =150 GeV	mc11_7TeV.125433.PythiaZH150_ZZ4lep	20000	0.146
	M_H =160 GeV	mc11_7TeV.125435.PythiaZH160_ZZ4lep	19950	0.0565

Table 28: List of Monte Carlo samples used in the analysis, cross sections are taken from [40, Summer 2012, Useful cross sections]