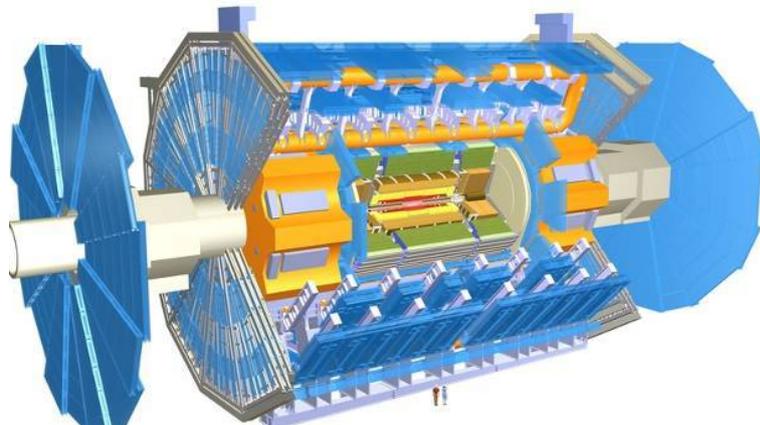


Technische Universität Dresden
Laborpraktikum

Der ATLAS-Versuch
Analyse von Proton-Proton-Kollisionen

The ATLAS Experiment
Analysis of Proton-Proton Collisions

Higgs Boson Decay
and
Search for new Physics



Authors: Johann Dittmer, Paul Moder, Manuel Gutsche

Contents

1	Introduction	4
2	The ATLAS Detector	4
2.1	Tracking Detectors	5
2.2	Calorimeter	5
2.3	Muon Spectrometer	6
2.4	Missing Transverse Momentum	6
3	The Higgs Boson	7
4	The Z' Boson	9
5	Statistics	10
5.1	The Likelihood Function	10
5.2	p-value	11
5.3	Application in Particle Physics	12
5.4	Upper Bound	13
5.5	Cuts and Statistical Significance	13
6	Tasks	15
6.1	Graphic Representation of Particle Reactions	15
6.2	$H \rightarrow WW$ Analysis	16
6.2.1	Analysis	16
6.2.2	p-Value Calculation	20
6.3	Search for Physics Beyond the Standard Model	20

1 Introduction

Since the launch of the Large Hadron Collider (LHC), a great amount of data has been generated from the collision of protons. The most important measurement as of today was arguably the detection of the Higgs boson in 2012, supporting the Higgs mechanism of the Standard Model. This lab course experiment will give a good overview about measurement and evaluation of data at CERN: It will specifically be shown what methods can be applied when searching for the Higgs boson and, on the other hand, the search for a particle that does not belong to the Standard Model will be attempted. For this purpose, data from proton-proton events, which were measured with the ATLAS detector, will be used.

2 The ATLAS Detector

For the following tasks, data recorded at the LHC with the ATLAS detector will be evaluated. In order to do so, you should have at least a rough idea of how the data is recorded. The ATLAS detector consists of several layers of subdetectors that are required for various particle measurements, see figure (1). In addition to the particles

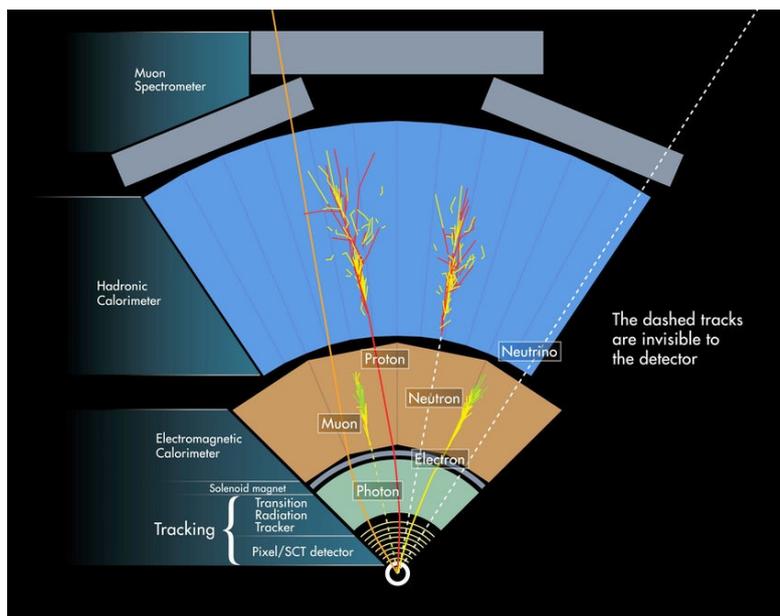


Figure 1: Schematic structure of the ATLAS detector [1].

that can be seen in this figure, there are of course many more that are of interest as well. However, the detector is designed in such a way that it can only measure certain final states. A large part of the resulting particles has an average lifetime of less than 10^{-10} s, which means that they usually cannot reach the detector before they decay. It therefore

makes sense to limit oneself to the decay products and from their kinematics reconstruct properties of the original particles.

2.1 Tracking Detectors

As can be seen in figure (1), the innermost part of the ATLAS detector consists of a tracking detector. As the name suggests, it records the track of the particles that are created in the center of the detector. For that, a magnetic field is applied in order to be able to deduce the charge of the particle from the curvature of the path. The tracking detector itself consists of three individual detectors. The innermost is a pixel detector consisting of 80 million pixels. Each pixel corresponds to a possible spatial point in the track of the particle. Since several particles often pass through the detector at the same time, it is necessary that all pixels can be processed in parallel. The second detector is a silicon semiconductor detector, which consists of over 6 million strips and can therefore also resolve the track very well. Last but not least, there is a transition radiation detector, which uses the transition radiation to determine the mass and, with the help of the curvature, the particle identity.

2.2 Calorimeter

The next layer of the detector consists of calorimeters, that are used to measure the energy of specific particles. Since different particles interact with each other via different interactions, different types of calorimeters are required as well. The first calorimeter is called an electromagnetic calorimeter. Here, all particles that interact via the electromagnetic interaction deposit energy. Its main task is to determine the energy of photons and electrons. Electrons interact with matter and emit a photon in the process, that, if it is sufficiently energetic, can decay into an electron-positron pair near a nuclear potential, which in turn releases energy in the form of a photon. A so-called shower develops, which can also be seen in figure (2).

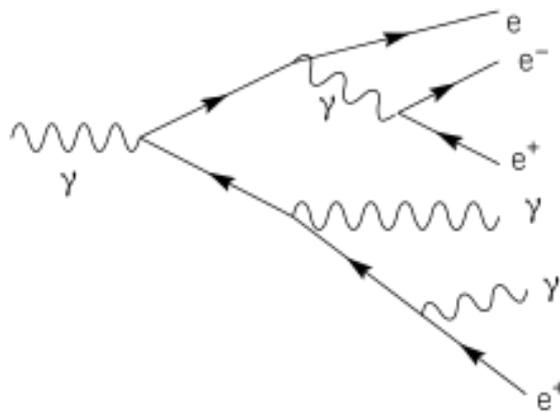


Figure 2: Electromagnetic shower [2].

If at some point the energy of the individual particles is small enough such that no more new particles can be generated, they transfer their remaining energy directly to the material of the electromagnetic calorimeter. The original energy of the electron or photon can then be determined from the sum of all these particle energies.

The remaining particles, such as muons or hadrons, also deposit energy when passing through the electromagnetic calorimeter, but in contrast to electrons and photons lose only part of their total energy. Therefore, additional detectors are required for these particles. In the case of hadrons, the fact that they consist of quarks and therefore interact via the strong interaction is exploited. Therefore, the electromagnetic calorimeter is surrounded by a hadronic calorimeter, in which hadrons trigger hadronic showers and thus deposit all their energy.

2.3 Muon Spectrometer

The last directly measurable particle that we want to consider is the muon. Although it interacts electromagnetically, it does not deposit all of its energy in the electromagnetic calorimeter, since the cross section for interaction with matter is significantly smaller than that of an electron. The muon is therefore one of the few particles that reaches the outer layer of the detector — the muon spectrometer. The muon spectrometer is the largest part of the ATLAS detector and consists of 4000 individual muon chambers. In contrast to the previous parts of the detector, it is not a calorimeter, but again a tracking detector, which tracks the muons with the help of argon gas. From these traces and the ones from the inner tracking detectors, one can then determine the momentum and other properties of the muon from the curvature, known mass and charge.

2.4 Missing Transverse Momentum

The last known final state particles that are of interest are neutrinos, but they have no charge and only weakly interact with matter. Therefore a direct measurement is not possible in case of the ATLAS detector — their presence can only be proven indirectly. Calculating the transverse momentum of all visible particles should actually yield 0, since the protons themselves have no momentum in the transverse direction. Typically, this is not the case and the excess transverse momentum is then compensated by the so-called missing transverse momentum, which then corresponds to the vectorial sum of the transverse momenta of all invisible particles (usually neutrinos).

It is important to note that this missing transverse momentum does not make it possible to describe exactly which neutrino carries how much energy. Furthermore, it is even possible that part of the transverse momentum is contributed by other, not directly measurable particles (e.g. dark matter candidates like LSPs). Calculations using it are therefore definitely associated with great uncertainties.

3 The Higgs Boson

The Higgs boson is an elementary part of the Standard Model. The necessity for its existence lies in the mass of the particles, because in the original formulation of the Standard Model the particles' mass terms would violate gauge invariance, meaning that all particles had to be massless, contradicting measurements. A solution to this problem was the Higgs mechanism, introducing a new field, the Higgs field. One of its symmetries gets broken in a process called electroweak symmetry breaking, which gives rise to a non-vanishing vacuum expectation value of the Higgs field that ultimately leads to the particle masses. The theory of this mechanism predicts a new particle — the Higgs boson. The discovery of which would provide strong evidence for the mechanism, therefore motivating Cern to look for the corresponding particle. Finally, in 2012, the particle could be detected and its mass was determined to be around 125 GeV.

As already mentioned, the Higgs field gives particles their mass by coupling to them, whereby a larger coupling parameter also means a higher mass. Accordingly, one can conclude from this, that it does not couple to massless particles such as photons, and in particular, to gluons. However, it is precisely the latter channel that is of interest in the generation of the Higgs boson through the collision of protons. The problem is solved by a fermion loop, as shown in figure 3. The fermions in the loop are usually top quarks, since they have the highest mass and therefore the strongest coupling to the Higgs boson. The Higgs boson cannot be detected directly due to its short lifetime

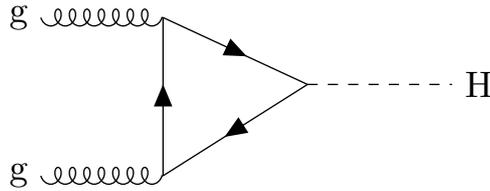


Figure 3: Generation of a Higgs boson via Gluon fusion (ggF).

of about 10^{-22} s as it decays before it reaches the detector. There are different decay channels of which one would like to observe a channel that has the highest possible effective cross section. A possible decay is the one into two gauge bosons (Z or W). In this analysis, we are primarily interested in the decay into two W bosons (see figure 4), as it has a higher cross section. However, W bosons do not reach the detector either and instead can decay hadronically or leptonically, as can be seen in figure 5. For a Higgs boson, that decays into two W bosons, there are either 4 quarks (hadronic decay), 2 quarks and 1 lepton (semileptonic decay) or 2 leptons (leptonic decay) detectable in the final state. The neutrinos are not further considered here, since they can only be measured as a missing transverse momentum. Since quarks cluster into jets, they are much more complicated to measure than leptons, which is why we focus on the leptonic decay. The possible end products are electrons or muons (τ leptons are ignored here as well for the time being, since they themselves decay into leptons or quarks, which

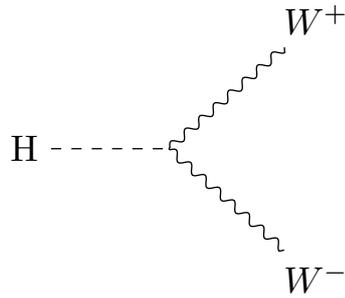


Figure 4: Higgs bosons decay into two W bosons.

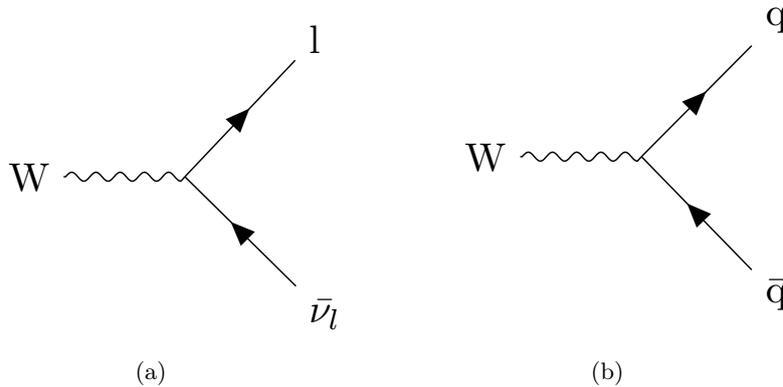


Figure 5: Possible decay channels of W bosons in (a) a lepton and neutrino or (b) a quark and anti-quark.

makes their reconstruction challenging). So there are three channels for us to consider: an electron-electron channel, a muon-electron channel and a muon-muon channel. (Note that, due to the conservation of charge, there is of course always a positively charged and a negatively charged lepton. For the sake of simplicity, however, we refer to both types of leptons in the same way. Actually, the first channel would be an electron-positron channel.)

Of course, there are different backgrounds to consider for both channels. In the case of final states with leptons of one generation, one of the largest backgrounds is the decay of a single Z boson into two leptons (see figure 6). The biggest difference between Higgs and Z decays is the missing transverse momentum, which should be significantly smaller in the decay of the Z boson, since no neutrinos are produced here. Of course, all channels have the simple WW decay as background, where the W bosons did not originate from a Higgs boson. A difference to this case is, for example, the transverse invariant mass of the two W bosons (unfortunately we can only consider the transverse invariant mass here, since only the missing transverse momentum is known, but not the

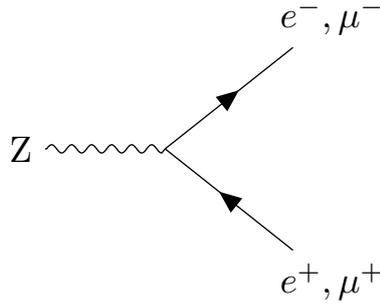


Figure 6: decay of the Z boson.

entire missing momentum, which means that a momentum component of the neutrinos in the calculation of the invariant mass is missing). In normal WW decay, the transverse invariant mass should be close to twice the W mass. Things are a bit different with the decay of the Higgs boson. Since the Higgs boson itself only has a mass of 125 GeV, at least one of the W bosons needs to be off-shell — meaning it does not satisfy the energy-momentum relation $p_\mu p^\mu = m^2$ and only appears in a virtual, intermediate state. The sum of the transverse invariant masses of the two W bosons is then close to the Higgs mass of 125 GeV. There are of course other differences in the two processes, including the angular distribution between the two leptons. Finding good selection criteria (cuts) will be an essential part of this analysis (see also section 5.5).

4 The Z' Boson

With the discovery of the Higgs boson (previous chapter) in 2012, the Standard Model of particle physics was extended by an elementary building block. However, not all observable phenomena can be explained within this theory, therefore, beyond the Standard Model (BSM) physics are necessary. Usually, new particles are predicted in such theories, one of which is the Z' (pronounced Z prime) boson. The Z' boson is a neutral heavy particle with spin 1, which appears, for example, in theories of grand unification (GUT), extending the electroweak sector by another $SU(2)$ symmetry. Depending on the theory, its mass ranges from the electroweak scale (100 GeV) to the Planck scale (10^{19} GeV). The production of the Z' at the LHC is similar to the production of the well-known Z boson via the annihilation of a quark-antiquark pair, but since the cross section decreases with increasing mass, significantly fewer Z' bosons are predicted to be produced. Due to the higher mass, the Z' can couple to all fermions, such that decay into all channels of the Z boson and into a $t\bar{t}$ pair is possible. In this experiment's analysis the decay over two top quarks is investigated. More information can be found in Dirk Duschinger's diploma thesis [3].

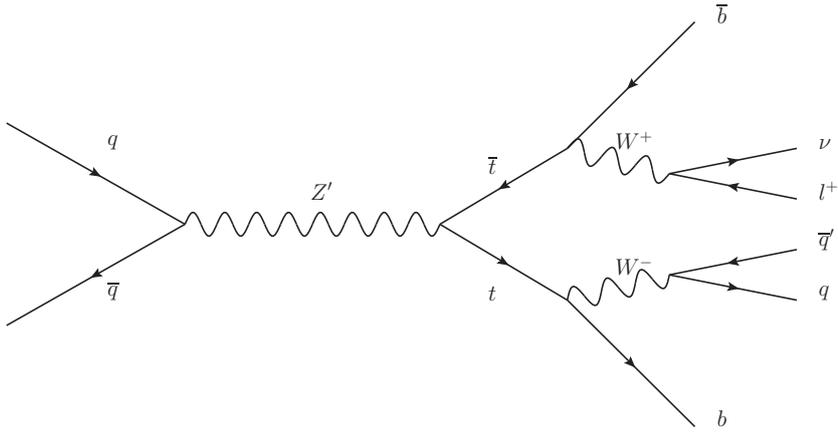


Figure 7: Production and decay of the Z' boson.

5 Statistics

5.1 The Likelihood Function

Physics measurements depend on various parameters, some of which can be unknown properties of particles, such as their mass. This dependency can be expressed in a density function of the form

$$f = f(x_1, \dots, x_N | \mu_1, \dots, \mu_M). \quad (1)$$

This describes the probability of measuring the values x_1 to x_N under the conditions μ_1 to μ_M . If one knows the parameters μ_1 to μ_M , one can create a probability density for them and calculate which measured values are more probable than others. In reality, however, exactly the opposite occurs. The mass of a particle, for example, is unknown to us and we want to find the most probable prediction, called best estimator, for one or more parameters from various measured values x_1 to x_N . This best estimator can be calculated from the likelihood function, which is defined as

$$l(\mu_1, \dots, \mu_M) = c \cdot \prod_{i=1}^N f_i(x_i | \mu_1, \dots, \mu_M), \quad (2)$$

where c is just a normalization factor. The best estimators $\hat{\mu}_1, \dots, \hat{\mu}_M$ are found for maximum l . So, assuming that the measured values only depend on one parameter, one can simply write

$$\left. \frac{dl}{dp} \right|_{\hat{\mu}} = 0. \quad (3)$$

Often, instead of the normal likelihood function, the log-likelihood is used, by simply taking the logarithm. As a result, the product of functions ultimately becomes a sum

$$L(\mu_1, \dots, \mu_M) = \ln l(\mu_1, \dots, \mu_M) = \sum_{i=1}^N \ln f(x_i | \mu_1, \dots, \mu_M). \quad (4)$$

5.2 p-value

In this chapter we want to take a closer look at the so-called p-value — how it is defined, what it is used for and how it can be determined. In order to be able to calculate the p-value, one first needs a prediction about the expected behavior of the data, that will then be tested with a measurement. The p-value of the measurement then indicates the probability, assuming that the null hypothesis is correct, of obtaining results at least as extreme as the ones observed. This is shown in figure 8. The rule of thumb is that the null hypothesis (e.g. the Standard Model without the Higgs boson) is rejected if the p-value is less than 0.05 (corresponds to 2σ) and that a new particle has been discovered if the p-value is less than $5.7 \cdot 10^{-7}$ (corresponds to 5σ). Figure 8 also shows directly

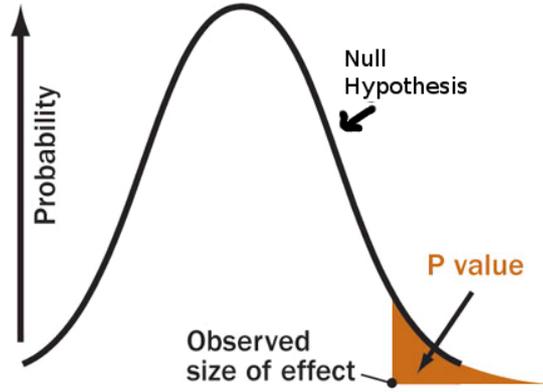


Figure 8: Illustration of the p-value [4].

how to calculate the p-value of a measurement. In principle, one only has to know the probability density of the null hypothesis and then integrate from the measured value. However, this measured value is usually not a specific value of a measurement, but a so-called test statistic, which is itself determined from measurements. We use the following test statistic in this experiment:

$$t_\mu = -2 \ln \frac{l(\mu)}{l(\hat{\mu})}. \quad (5)$$

Here $l(\mu)$ is of course the likelihood function and $l(\hat{\mu})$ is the likelihood function of the best estimator. The p-value can then be calculated as

$$p = \int_{t_\mu}^{\infty} f(t_\mu | \mu) dt_\mu. \quad (6)$$

5.3 Application in Particle Physics

When searching for a particle, we first assume that the known model is correct, this represents our null hypothesis. Monte Carlo simulations can then be used to calculate which result is expected from a specific analysis with specific cuts. We call this prediction background b . The same applies, of course, to the signal of the new particle, which is called s . These Monte Carlo results are then compared to the measurements n . Then, if the particle really exists, in the measurement there should be $b + s$ events in all bins of a histogram. So for our density function we choose a parameter μ , called signal strength, which, ideally, becomes 0 if the background alone is more compatible with the data, and becomes 1 if the combination of background and predicted signal describe the data better.

We now assume that the measured values in the individual bins are Poisson distributed, which results in

$$g_i(n_i|\mu) = \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \quad (7)$$

The likelihood function then is the product over all bins

$$l(\mu) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \quad (8)$$

and the test statistic becomes

$$t_\mu = -2 \ln \frac{l(\mu)}{l(\hat{\mu})}. \quad (9)$$

However, we will adjust this test statistic a bit. Since in our case it makes no sense to get a μ that is less than 0, the best estimator is then automatically $\hat{\mu} = 0$, resulting in

$$\tilde{t}_\mu = -2 \ln \begin{cases} \frac{l(\mu)}{l(\hat{\mu})} & , \hat{\mu} \geq 0 \\ \frac{l(\mu)}{l(0)} & , \hat{\mu} < 0 \end{cases}. \quad (10)$$

We want to calculate the p-value for the null hypothesis, i.e. we want to calculate the p-value for $\mu = 0$, which changes the test statistic to

$$\tilde{t}_0 = \begin{cases} -2 \ln \frac{l(0)}{l(\hat{\mu})} & , \hat{\mu} \geq 0 \\ 0 & , \hat{\mu} < 0 \end{cases} \quad (11)$$

and leads to the p-value for the null hypothesis:

$$p_0 = \int_{\tilde{t}_0}^{\infty} f(\tilde{t}_0|\mu = 0) d\tilde{t}_0. \quad (12)$$

However, one problem remains. In order to calculate the p-value, we have to integrate over the density function of the null hypothesis $f(t_\mu|\mu)$, which is not directly known. We

avoid this by creating pseudo-data, assuming that the null hypothesis is true, i.e. $\mu = 0$, and again that the measured values of the individual bins are Poisson distributed. This results in the density function

$$g_i(n_i|\mu = 0) = \frac{b_i^{n_i}}{n_i!} e^{-b_i}, \quad (13)$$

from which pseudo-data can be created. With this pseudo-data we then calculate the test statistic t_0 again and repeat this process several 1000 times. We basically create a density function ourselves, over which we do not integrate, but add up, and only over values that are greater than \tilde{t}_0 .

$$p_0 = \sum_{t_0 > \tilde{t}_0} f(t_0|\mu = 0) \quad (14)$$

5.4 Upper Bound

The p-value can give us an indication of whether the measured data would reject the null hypothesis. However, it does not tell us whether this happened due to our signal. Therefore, the so-called upper bound for the signal strength μ is usually calculated as well. For this we introduce a new test statistic

$$q_\mu = \begin{cases} -2 \ln \frac{l(\mu)}{l(\hat{\mu})} & , \mu \geq \hat{\mu} \\ 0 & , \mu < \hat{\mu} \end{cases}, \quad (15)$$

which differs in that the test statistic is now automatically 0 if μ is less than $\hat{\mu}$. With the help of this new test statistic, a p-value is calculated in analogy to the formula described above, i.e. first a \tilde{q}_μ is determined from real data and then other q_μ that are created from pseudo-data, such that

$$p_\mu = \int_{\tilde{q}_\mu}^{\infty} f(q_\mu|\mu) dq_\mu \quad (16)$$

can be converted into a sum. These p-values are calculated for all possible $\mu > \hat{\mu}$ and the μ at which the p-value drops below 0.05 for the first time is the upper bound.

5.5 Cuts and Statistical Significance

Cuts are of great importance for the analysis of particle collisions — what exactly are they and why are they so essential? In order to be able to answer these questions, let us take a closer look at a particle collision. When protons collide, the gluons or quarks interact with each other, possibly generating particles that we want to investigate (e.g. the production of Higgs bosons in chapter 6.2). However, we cannot be certain beforehand which particles will result from a specific collision. One can only predict the probability of a particular particle being formed. For example, the production of a

Higgs boson is much less likely than that of a Z Boson. So if one looks at all events, most processes of interest will be included, but some occur much more frequently. Accordingly, rare events such as the generation of a Higgs boson would be virtually unobservable by the vast background of other processes. Therefore, cuts are applied to the events, which means that not all events are included in the evaluation, but only those that meet certain selection criteria. For example, one could only consider events with exactly one lepton in the final state. Since a Z Boson can only decay into two leptons, a large number of events based on a Z Boson would be sorted out and not considered further. The fact that not all events with a Z Boson can be sorted out like this is, on one hand, because there are always secondary processes that lead to further particles in the final state, and on the other hand, particles can be simply misinterpreted.

Cuts are therefore a good way of filtering out unwanted underground processes. However, one must also note that all selections can of course also remove events that are of interest to us. We always have to weigh up which cuts we want to use so that we remove as much background as possible while retaining as much signal as possible. This is represented by the so-called statistical discovery significance \mathcal{Z} — the higher it is, the better the selection criteria are for a specific signal. We will briefly derive its formula here, assuming a Poisson distributed likelihood function

$$l(\mu) = \frac{(\mu s + b)^n}{n!} e^{-\mu s - b} \quad (17)$$

with the best estimator

$$\hat{\mu} = \frac{n - b}{s}. \quad (18)$$

We now consider the test statistic from equation 11 again and obtain the significance simply by taking the square root of it

$$\mathcal{Z}^2 = \begin{cases} -2 \ln \frac{l(0)}{l(\frac{n-b}{s})} & , \hat{\mu} \geq 0 \\ 0 & , \hat{\mu} < 0 \end{cases} \quad (19)$$

$$= \begin{cases} -2 \ln \left(\left(\frac{b}{n} e \right)^n e^{-b} \right) & , \hat{\mu} \geq 0 \\ 0 & , \hat{\mu} < 0 \end{cases}. \quad (20)$$

Using $s + b = n$, this results in

$$\mathcal{Z} = \begin{cases} \sqrt{2 \cdot [(s + b) \cdot \ln(1 + \frac{s}{b}) - s]} & , \hat{\mu} \geq 0 \\ 0 & , \hat{\mu} < 0 \end{cases}. \quad (21)$$

For a large number of background events, an useful approximation is

$$\mathcal{Z} = \begin{cases} \frac{s}{\sqrt{b}} & , \hat{\mu} \geq 0 \\ 0 & , \hat{\mu} < 0 \end{cases}. \quad (22)$$

6 Tasks

6.1 Graphic Representation of Particle Reactions

In this part of the experiment, the ATLANTIS program is used, which graphically displays particle collisions. The program can be started via

```
module load atlantis
atlantis
```

First, get familiar with the program — learn how to navigate, select particle tracks, apply criteria, etc. Load the learning data sets for electrons, muons, photons, τ leptons and jets and use them to understand how these particles can be recognized in the detector. After that, the supervisor will give you two of the following tasks:

1. What is — roughly estimated — the mean duration of a particle shower in the electromagnetic calorimeter of the ATLAS detector? Use a rough estimate to determine the mean transverse area covered. Is the area for electrons and photons similar? Use the first 15 electron and the first 15 photon events from the respective learning data sets.
2. Load the electron training data set and determine the radius of curvature of the electron tracks in the inner detector for 20 electron events. From this, determine the momentum of the electrons. Further determine the energy of the ECAL cluster and make a table of the ratio of cluster energy to electron momentum. Create a histogram from it.
3. Load the muon training data set and compare the momentum of the muons as measured in the inner detector with the momentum measurement in the muon detector system. Use the first 20 events for this. Determine the mean energy loss of the muons in the calorimeter. What quantity does the loss depend on the most?
4. Use the photon learning data set to determine the conversion probability for photons in the innermost layer of the detector. Think about up to which point you can still make meaningful statements about the conversion probability.
5. Determine the number of additional jets in events in which a W boson is produced.
6. Determine the invariant electron-positron mass for three $Z^0 \rightarrow e^+e^-$ decays
 - a) exactly, i.e. considering the electron mass
 - b) for $m_e \approx 0$
7. Analyse the data set Mystery-xxx.dat. Using information such as the number of leptons, jets, or E_T^{miss} , try to determine the Standard Model physics process that might be responsible for the events in the mystery dataset. Please note that the mystery dataset has already been preselected to contain at least two leptons (e, μ). Then justify your decision, even if several physics processes are possible at the end.

8. In 2007, cosmic radiation events were recorded with the ATLAS detector. At sea level, the cosmic radiation consists almost exclusively of muons, which originate from the decay of charged pions. Due to the large mean free path in matter, muons can be detected even at a depth of 100 meters. Use the cosmic radiation data set `ATLASData-Cosmics-M5.zip` to determine the zenith angle distribution of the muons. Consider how the zenith angle is related to the angles η and ϕ_0 obtained with the `Pick` command. Does this angular distribution match the known distribution at sea level?

6.2 $H \rightarrow WW$ Analysis

6.2.1 Analysis

In this part of the experiment we want to test the theory of the Higgs boson. However, as already mentioned, the generation of a Higgs boson is rare compared to other processes. We thus have to remove as much background as possible from our measurements through cuts in order to be able to discover something. So your task in preparation for the experiment is to think about how the $H \rightarrow WW$ process differs from standard model processes (Z , W , $t\bar{t}$) and what selection criteria should therefore be applied.

However, this alone may not be sufficient. In order to compare processes accurately, so-called Monte Carlo simulations are carried out. After simulating the relevant signal and background processes, one can compare them with each other as well as with actual data and check how certain cuts affect the final states. The ATLAS open data portal provides an analyser for this (<http://opendata.atlas.cern/visualisations/analyser-js.php>) in which you can implement cuts directly and see their effects. You can find the analyser by scrolling down the page.

Here you can find different histograms — with only the Monte Carlo simulations on one tab and with Monte Carlo and real data on another. We are less interested in the experimental data there (we want to examine our own), but in both analysers there are — partially different — observables available for which cuts can be tested. First, play around a bit to see the effects of possible cuts. Also observe how the behaviour changes when examining the data in different channels (electron-electron, electron-muon and muon-muon). Once you have gotten a feeling for it, your task is to increase the significance of the Higgs signal, which is automatically calculated and shown by the analyser. Once you have found cuts that you are satisfied with, your next task is to transfer them to the analysis. To do this, open `HWWAnalysis.py`. This document is your analysis and it not only determines which observables you plot, but also which cuts to apply to the events. The analysis consists of 3 parts.

The first section is called `initialize`. In this function, the histograms, which are to be filled later, are defined. In the example shown in figure 9, histograms of the total transverse momentum of two leptons, the angle between these leptons, and the transverse momentum of the so-called leading lepton (the lepton that has the highest energy) would be created. As you can already see, there are two ways to define histograms. The `addHistogram` function gives you the freedom to define your own histogram. You

```

def initialize(self):
    # defining the histograms
    self.hist_ptLL = self.addHistogram("ptll", ROOT.TH1D("ptll", "Transverse Momentum of Dilepton System: p_{T,ll}:Events", 20, 0, 200))
    self.hist_deltaPhiLL = self.addHistogram("deltaphill", ROOT.TH1D("deltaphill", "Azimuthal Opening Angle between Leptons: #Delta#phi_{ll}:Events", 16, 0, 3.2))
    self.hist_leadlep_pt = self.addStandardHistogram("leadlep_pt")

```

Figure 9: initialize part.

give the histogram a name, a title, a designation for the x and y axes and finally the number of bins, the start of the lowest bin and the end of the highest bin. In principle, we recommend this definition. However, there is also the possibility of using the function `addStandardHistogram`. You use a histogram that has already been defined in the file `StandardHistograms.py`. If you want to use this function, you must ensure that the name of your histogram must exactly match the name of the histogram in `StandardHistograms.py`.

The second part of the analysis is the **analyze** part, which is carried out for each event individually. First you have to get the objects you want to work with, then apply various operations on these objects (including selection criteria, but definitions of new observables are also possible) and finally fill the corresponding observables' histograms with the events. We'll go through each of these steps. First of all, the objects have to be fetched. The individual objects such as leptons or jets are already combined in containers, so we only have to get the containers themselves. You can see this in figure 10 as an example. `GoodLeptons` is the name of our container in which our leptons can be

```

# lepton Requirements
# Number of leptons? Charge?
GoodLeptons = AH.selectAndSortContainer(self.Store.getLeptons(), AH.isGoodLepton, lambda p: p.pt())

# other Requirements
# Number of Jets? Missing transverse momentum? Differences between channels?
jets = AH.selectAndSortContainer(self.Store.getJets(), AH.isGoodJet, lambda p: p.pt())
etmiss = self.Store.getEtMiss()

```

Figure 10: Fetching the containers.

found and `jets` the container with the jets. It is important to note that these are not all leptons or jets, but have already been preselected here (which can be seen in the argument `AH.isGoodLepton`). The reason for the preselection is that we only want to consider those objects that have a high probability of coming from the main process of the event. In addition to the main process, there are many secondary processes that can also produce leptons or jets in the final state. If we were to include all of these, it would be more difficult to find cuts with which we can separate signal and background from one another. Accordingly, these should not even be considered. We take advantage of the fact that the secondary processes usually have significantly less energy than the main process. Therefore, only leptons (or jets) above a certain energy threshold are considered in the analysis. You can read about the preselections made in this analysis in the file `AnalysisHelpers.py`.

You can easily determine how many objects are in a container using the `len` function you are familiar with. The missing transversal momentum is special in a way, as it is not contained in any container, but is itself an object that has already been reconstructed. After having fetched all containers, we can access the objects in them, which already have been sorted according to their energy, such that `GoodLeptons[0]` is always our leading lepton (same for the jets). Now, to apply cuts on our events, all we really have to do is set up if-conditions and return the Boolean `False` if the cut isn't met, as shown in figure 11. The line `self.countEvent(...)` has no effect on the event selection, but adding it will

```

leadLepton = GoodLeptons[0]
trailLepton = GoodLeptons[1]

if not leadLepton.pt() > 22: return False
if not trailLepton.pt() > 10: return False
self.countEvent("Lepton requirements", weight)

```

Figure 11: Example for applying cuts.

display the number of remaining events at this point for each sample. This way you can check the efficiency of each cut.

An useful class for the analysis is `TLorentzVector`. `TLorentz` vectors are exactly what they sound like: they are Lorentz or four-vectors, with which you can more easily obtain certain properties of objects. First you have to define an object as a `TLorentz` vector and then you can set its values with the function `SetPxPyPzE(...)` (as shown in figure 12). But there are also other possible functions such as `SetPtEtaPhiE(...)` (see also the

```

t11NoEta = ROOT.TLorentzVector()
t11NoEta.SetPxPyPzE(leadLepton.pt()*math.cos(leadLepton.phi()), leadLepton.pt()*math.sin(leadLepton.phi()), 0, leadLepton.pt())

t12NoEta = ROOT.TLorentzVector()
t12NoEta.SetPxPyPzE(trailLepton.pt()*math.cos(trailLepton.phi()), trailLepton.pt()*math.sin(trailLepton.phi()), 0, trailLepton.pt())

```

Figure 12: Example for the usage of Lorentz vectors.

`ROOT` documentation for `TLorentzVector`). You can then access individual elements of the Lorentz vector just as easily. For example, the function `Pt()` gives you the transverse momentum of the object directly. Objects that have already been defined by the analysis framework, such as leptons, are also Lorentz vectors. Here however, you must ensure that the functions are always written in lower case, for example `pt()`. If you want to use the Lorentz vectors of leptons or the missing transverse momentum, you must first attach `.tlv()` to the objects.

So after all cuts have been inserted, the histograms that we defined in the initialize part must be filled at the end. This works simply by adding a line

```

self.<Histogram-Name>.Fill(<Observable>, weight)

```

as shown in figure 13. As `<Observable>` you must pass the name of the variable which corresponds to the observable in your analysis. You also have to enter the `weight`. The reason for this is that the Monte Carlo sets were produced with a certain amount of

```

# Missing Et histograms
self.hist_etmiss.Fill(etmiss.et(),weight)

# Histogram for the higgs mass
self.hist_invMass.Fill(higgs.M(), weight)

# Histograms for deltaphi
self.hist_deltaPhiLM.Fill(abs(etmiss.tlv().DeltaPhi(leadLepton.tlv())), weight)
self.hist_deltaPhiML.Fill(abs(etmiss.tlv().DeltaPhi(trailLepton.tlv())), weight)
self.hist_deltaPhiLLM.Fill(abs(etmiss.tlv().DeltaPhi(combTLV)), weight)

```

Figure 13: Example for filling the histograms.

events that does not necessarily correspond to the amount of events one would expect for the given data, such that the Monte Carlo sets have to be reweighted to fit the amount of data. This happens at the beginning of the `analyze` part, as shown in figure 14.

The last part of the analysis is not important for you — the `finalize` part, which, as

```

# retrieving objects
eventinfo = self.Store.getEventInfo()
weight = eventinfo.scalefactor()*eventinfo.eventWeight() if not self.getIsData() else 1
self.countEvent("no cut", weight)

```

Figure 14: Definition of weight for Monte Carlo samples.

shown in figure 15, no longer requires any additions.

```

def finalize(self):
    pass

```

Figure 15: `finalize` part.

After you have implemented all the cuts, you can run the analysis. For this, type

```
python RunScript.py -a HWWAnalysis -p -n 25
```

in the terminal. But before you run over all the data, you should start with 10% to 30% of the data to test your script. To do this, open the `Configuration.py` file in the `Configurations` folder. At the top you will find a dictionary with an entry called `Fraction`, as also shown in figure 16. If this value is set to 1.0, the analysis runs over 100% of the data, so set a lower value here for now. If the analysis runs without errors, check whether all histograms have been filled correctly. You can either examine the root files in the results directory or plot the histograms. To do this, you must first check whether all histograms are already implemented in the file `PlotConf_HWWAnalysis.py` in `Configurations`. If so, one can plot histograms with the command

```
python PlotResults.py Configurations/PlotConf_HWWAnalysis.py
```

```

Job = {
  "Batch"           : True,
  "Analysis"       : "TTbarAnalysis",
  "Fraction"       : 1.0,
  "MaxEvents"      : 1234567890,
  "OutputDirectory" : "results/"
}

```

Figure 16: Job definition in the configuration file.

and find them afterwards in the directory Output. For upcoming calculations, sort your results into different folders for data, signal Monte Carlo and background Monte Carlo. You now have another opportunity to revise your analysis. If you want, you can also test cuts that you couldn't try out with the analyser (for example, the angle between the leading lepton and the missing transverse momentum). To see if this has increased their significance, you can use the Significance.py program, which can be executed by entering the following command in the console:

```
python Signifikanz.py Configurations/PlotConf_HWWAnalysis.py
```

Then enter the paths you created. If the significance has increased, the new cut should be kept, otherwise you should adjust it or remove it completely. If you are satisfied with all cuts, start a complete analysis with 100 % of the events and note down the significance of some observables.

6.2.2 p-Value Calculation

With the created histograms we now want to calculate the p-value and conclude whether a Higgs boson might exist or whether the Standard Model without Higgs boson explains the data better. For this, look at the files in the pValue directory. Before you can run the Analysis.py program, you need to take a look at and edit the following scripts:

1. the file that calculates the likelihood function,
2. the file that calculates the test statistic and the upper bound test statistic,
3. the file that creates each pseudo-experiment and the file that creates the upper bound pseudo-experiments,
4. Analysis.py to calculate the variables.

As a test, you can set several 100 pseudo-experiments before you start the actual calculation with around 10000 pseudo-experiments. Calculate p-value and upper bound for 3 observables and evaluate your results.

6.3 Search for Physics Beyond the Standard Model

In the last part of the experiment we want to search for new physics with the ATLAS detector in the semileptonic $t\bar{t}$ channel of the decay of a Z' boson (Fig. 7).

Several Monte Carlo samples with different masses of the Z' boson between 400 and 3000 GeV are available for this analysis.

The analysis is implemented with the file `ZPrimeAnanalysis.py` in the Analysis directory. It also consists of the 3 parts described in exercise 6.2. As you can see, a lot of observables are already included, but you may want to consider more and implement them as done before. In any case, the mass of the Z' boson should be displayed. When you are happy with your code, test the analysis on 10 % of the data; if everything runs without errors, you can run over all data. Start the analysis in the same way as in the previous task:

```
python RunScript.py -a ZPrimeAnalysis
```

Now plot your results for two mass hypotheses — you need to change the file `PlotConf_ZPrimeAnalysis.py` accordingly — with the following command:

```
python PlotResults.py Configurations/PlotConf_ZPrimeAnalysis.py
```

Here, the p-value needs to be calculated as well, using the transversal mass of Z' . However, after running the analysis you should get a histogram as shown in figure 17. It can be seen that the transverse momentum of the leptons and jets in the Monte Carlo samples must be corrected due to detector uncertainties. You still have to add these correction to the files `ZPrimeAnalysis.py` and `AnalysisHelpers.py` by changing the values for `LepScale` and `JetScale`.

Once you have found correction factors where all the data bins agree with or are above the Monte Carlo ones, run the analysis on the full data and plot the results. For the upcoming calculations, it again makes sense to sort the results into different folders for data, signal Monte Carlo and background Monte Carlo.

In the previous task you have already performed a p-value calculation. However, the calculation should now not cover the entire spectrum, but only the area with the best ratio of signal and background $\frac{S}{\sqrt{S+B}}$.

Find a mass interval with the best ratio for each mass hypothesis, perform the p-value calculation on that interval and evaluate the results. In the file `ZPrimeAna.py`, specify the paths to your sorted result directories and an output path. You may also need to change other lines of code. The calculation is started as follows:

```
python ZPrimeAna.py Configurations/PlotConf_ZPrimeAnalysis.py
```

You receive 2 output files:

Signifikanz.txt contains the interval boundaries and the value for the best ratio $\frac{S}{\sqrt{S+B}}$.

p-Value.txt contains the results of the p-value calculation.

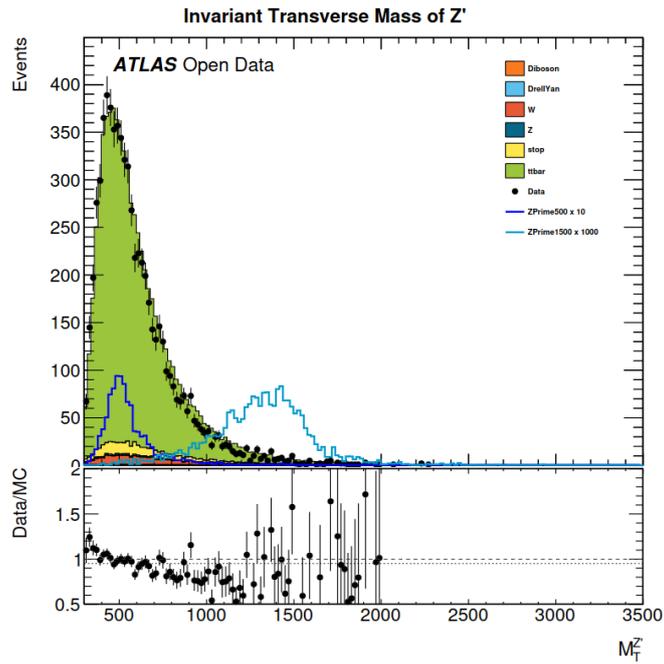


Figure 17: Transverse mass of the Z' boson. Only the two jets with the highest b-tag and the two jets with the highest transverse momentum are used for the calculation.

References

- [1] Michael Kobel et al. *Der ATLAS-Versuch — Eigenschaften von W-Bosonen und die Suche nach neuer Physik*. 2014. URL: https://tu-dresden.de/mn/physik/iktp/ressourcen/dateien/studium/praktikum/Versuchsbeschreibung_ATLAS_e214.pdf?lang=de (visited on 04/18/2023).
- [2] Wikipedia. *Particle shower*. 2022. URL: https://en.wikipedia.org/wiki/Particle_shower (visited on 04/18/2023).
- [3] Dirk Duschinger. *Search for New Physics in Fully Hadronic Di-tau Final States with the ATLAS Detector*. Diplomarbeit. 2013.
- [4] Corey Chivers. *P-value fallacy on More or Less*. 2011. URL: <https://bayesianbiologist.com/2011/08/21/p-value-fallacy-on-more-or-less> (visited on 04/18/2023).