FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

ASSIGNMENT OF BACHELOR'S THESIS

Title:	Optimization of the Matching Criteria Between the ATLAS and AFP Detectors at CERN
Student:	Petr Dostál
Supervisor:	doc. Dr. André Sopczak
Study Programme:	Informatics
Study Branch:	Web and Software Engineering
Department:	Department of Software Engineering
Validity:	Until the end of summer semester 2020/21

Instructions

The ATLAS Forward Proton (AFP) detector took large scale data in 2017 together with the ATLAS central detector at CERN. The goal of this project is to optimize the matching requirement between the two detectors.
The first task is to familiarize with the existing analysis software and the literature regarding two-photon interactions. A software development plan should be created with the focus to improve the current basic matching criteria. The uncertainties on the ATLAS central and AFP measurements should be taken into account in the matching optimization. The performance increase of the optimized matching should be compared with the performance of an existing simple matching criteria. The developed code should first be tested on a small event sample, and then applied to the whole data-set which requires the use of an user interface to grid computing. As a bonus, the method should also be tested on simulated data.

References

Will be provided by the supervisor.

Ing. Michal Valenta, Ph.D.
Head of Department

doc. RNDr. Ing. Marcel Jiřina, Ph.D.
Dean

Prague November 6, 2019

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Bachelor's thesis

Optimization of the Matching Criteria Between the ATLAS and AFP Detectors at CERN

Petr Dostál

Department of Software Engineering
Supervisor: doc. Dr. André Sopczak

June 4, 2020

Acknowledgements

I am very thankful to my supervisor, doc. Dr André Sopczak, for the patience he had with me and the opportunity to be a part of a real research. Also I would like to thank my friends and my family for the continuous support in my studies.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60 (1) of the Act.

In Prague on June 4, 2020 . . . .

Citation of this thesis

Dostál, Petr. *Optimization of the Matching Criteria Between the ATLAS and AFP Detectors at CERN*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2020.

Abstract

This thesis is a part of CERN's ongoing research of the search for axion-like-particle (ALP). It consists of a study on matching criteria for proton-proton interactions with two detected photons and a supporting software (compiled ROOT macros) to create ROOT ntuples that can be used to analyze the data with focus on matching. Included is also the development of three codes that are used to analyze ATLAS 2017 data – ntuple production, statistic uncertainty analysis, systematic uncertainty analysis. The study discusses the performance of three different matching criteria regarding the kinematics of the di-photon system and at least one detected proton. The main discovery is that the interactions studied in the data are caused by pile-up background, thus they are random interactions with no indication of a specific physics process.

Keywords CERN, LHC, ATLAS, AFP, axion-like-particle, ROOT, data analysis, matching criteria

Abstrakt

Tato bakalářská práce je součástí výzkumu CERNu, který se snaží najít axion-like-particle (ALP). Skládá se ze studie přiřazovacích kritérií pro proton-proton interakce se dvěmi detekovanými fotony a podporujícího softwaru (zkompilované ROOT makra), který vytváří ROOT ntuple, pomocí kterých se mohou analyzovat data se zaměřením na přiřazování. Obsahuje také vývoj tří kódů, pomocí kterých se analyzují ATLAS data z roku 2017 – vytvoření ntuple, statistická analýza chyby, systematická analýza chyby. Studie pojednává výkon tří různých přiřazovacích kritérií ohledně kinematiky dvoj-fotonového systému s alespoň jedním detekovaným protonem. Hlavním objevem práce je zjištění, že pozorované interakce vznikají díky pile-up backgroundu, tedy se jedná o náhodné interakce bez indikace fyzikálního procesu.

Klíčová slova CERN, LHC, ATLAS, AFP, axion-like-particle, ROOT, analýza dat, přiřazovací kritéria

Contents

List of Figures

List of Tables

Introduction

In 2017, CERN's ATLAS Central and ATLAS Forward Proton (AFP) detectors took large scale data including proton-proton interactions data. Since then, in 2019 ATLAS observed light-by-light (LbyL) scattering [1, 2, 3, 4] (a rare interaction where two highly energetic photons produce another pair of photons) by performing an analysis focused on lead-lead collisions with exactly two photons present in the system. Light-by-light scattering could also be observed in high energy proton-proton collisions, but the occurrence is very rare.

However there is a theory of existence of an axion-like-particle (ALP) [5] which could increase the occurrence rate over the expected amount coming from the Standard Model (SM) of particle physics. This thesis is a part of ongoing research that explores this ALP mediated LbyL scattering. The specific investigation is the matching of kinematic properties of the di-photon system with at least one proton detected.

There has already been an analysis [6] that serves as a proof of concept that the AFP can be used together with ATLAS Central Detector data. As a former Experimental Nuclear and Particle Physics student, the opportunity to be a part of this research has been great.

The main goal of this thesis is to analyze and optimize the matching criteria in proton-proton interactions with exactly two outgoing photons observed. This analysis will provide clarity to the randomness coming from unwanted reactions to enhance the clarity of the signal coming from LbyL scattering. In order to achieve this goal several scripts have to be written to analyze recorded and simulated data.

Outline The thesis is organized into several chapters starting with the theoretical background to this thesis. In Chapter 2 is a study on the matching criteria. The software development is discussed in Chapter 3 and the observed results are discussed in Chapter 4.

CHAPTER 1

Theoretical background

This chapter is separated into three parts – the introduction to CERN, the ATLAS experiment and the physics motivation behind this thesis. The section CERN briefly talks about the organization. The next section, The ATLAS Experiment, goes into detail about the detectors present at the ATLAS experiment. And lastly, in the section Physics motivation, the important introduction to the physics part of this thesis is discussed.

1.1 CERN

CERN [7], also known as the European Organization for Nuclear Research is one of the most important contributors to particle research. The acronym CERN (Conseil Européen pour la Recherche Nucléaire) is also used for the largest particle physics laboratory in the world that is a part of the organization. Another notable information is that CERN is the birthplace of World Wide Web [8].

A very important part of the laboratory is the Large Hadron Collider (LHC, Figure 1.1) [9, 10]. As of now it is the largest particle collider in the world. It also holds the world record for the highest total collision energy at 13 TeV. The LHC consists of four major experiments – the ATLAS experiment, the ALICE experiment, the LHCb experiment and the CMS experiment.

1.2 The ATLAS experiment

A Toroidal LHC ApparatuS (ATLAS) is the largest experiment of LHC at CERN [12, 13, 14]. It is also the biggest set of general purpose particle detectors.

The ATLAS central detector (Figure 1.2) is a layered multipurpose particle detector. The layers are designed to increase the spectrum of particles that can be identified. Each layer detects a specific region of pseudorapidity η of the

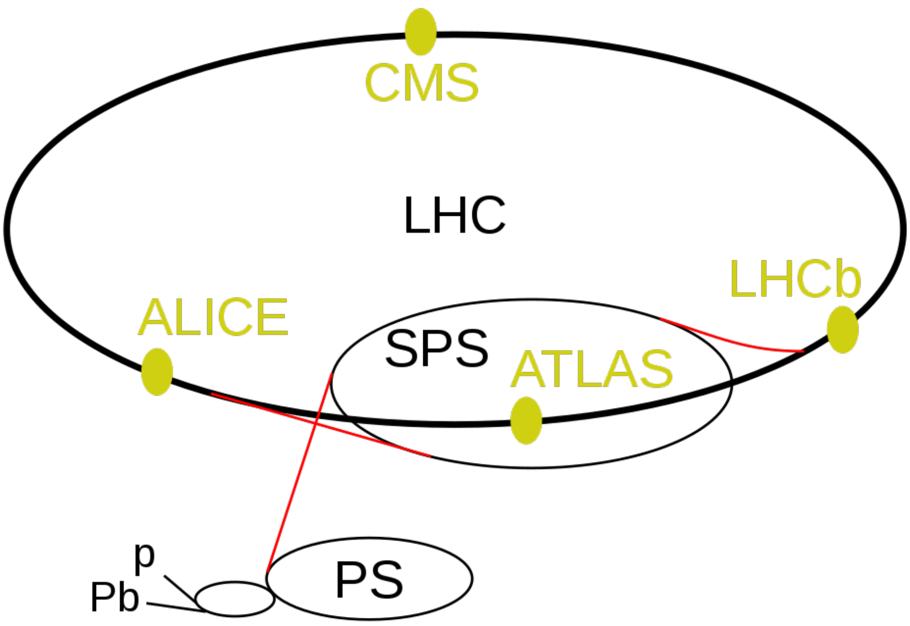

Figure 1.1: The LHC experiments and the preaccelerators. Source: [11]

particles. The main layers are the inner detector, the calorimeters, the muon spectrometer and the magnet system. Also a part of the ATLAS detector is the ATLAS Forward Proton (AFP) detector.

The AFP [15] is a set of pixel sensors located at around 210 m on each side of the beam from the interaction point (IP). The main goal is to detect protons that lose a fraction of their energy during an interaction in the IP (for example by photon emission). The internal structure are two stations, referred to as NEAR and FAR stations (distance from IP $\approx$ 205 m and $\approx$ 217 m, respectively). Each station consists of four 3D silicon pixel sensors that measure the trajectory of the passing protons.

1.3 Physics motivation

In theory, when the two very energetic protons get very close together, three options can occur. Both of them remain intact, one of them remains intact (the other one is destroyed) or neither of them remain intact (both protons are destroyed). During standard proton-proton collisions the protons are destroyed. In this research the detected protons are expected to come from the specific LbyL scattering interaction. Even if they stay intact, the expectation is that during the interaction the protons lose some of their energy, which alters their direction according to the energy lost.

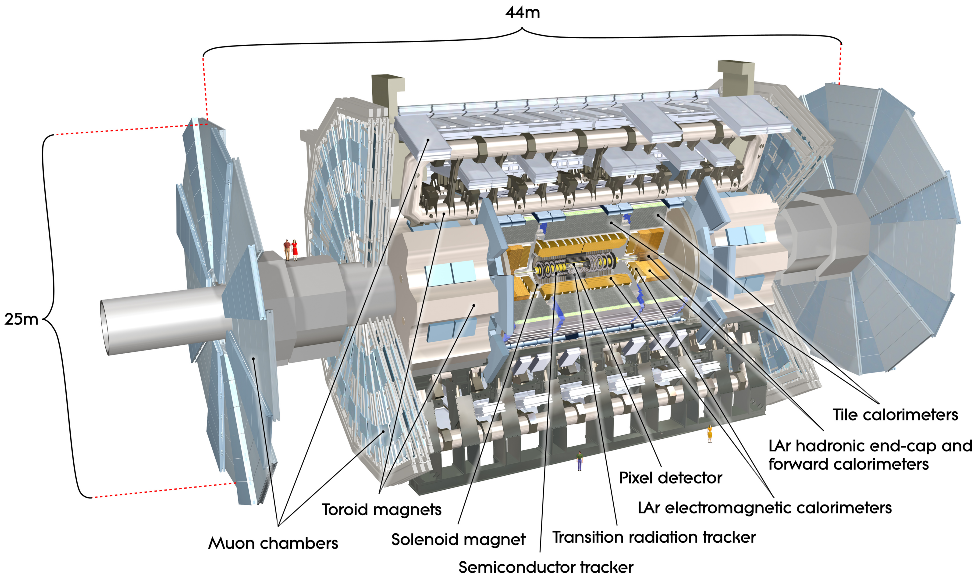

Figure 1.2: Computer generated image of the whole ATLAS detector. Source: [16]

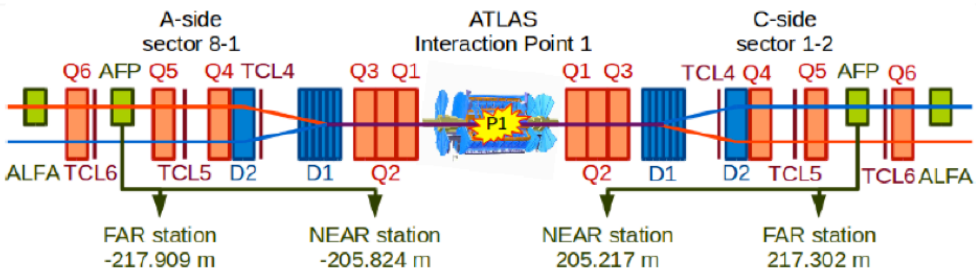

Figure 1.3: ATLAS central and AFP detectors. Q1–Q6 are magnets that focus the beam, D1–D2 are dipoles that bend the beam. Source: [17]

The direction change is caused by a set of magnets that affects particles with different electromagnetic fields differently. This electromagnetic field is directly related to the energy of the particle, thus by extension the energy is directly related to the path the protons take.

According to the law of conservation of energy the energy must be transferred or transformed. This analysis expects that during the interactions exactly two photons are created from the energy lost by the protons. In case a different amount of photons is detected in the central detector, the event is filtered out. This allows to reduce the unwanted background in the search for an axion-like-particle (ALP). The Feynman diagram (Figure 1.4) illustrates that when the protons get very close to each other, two photons are emitted.

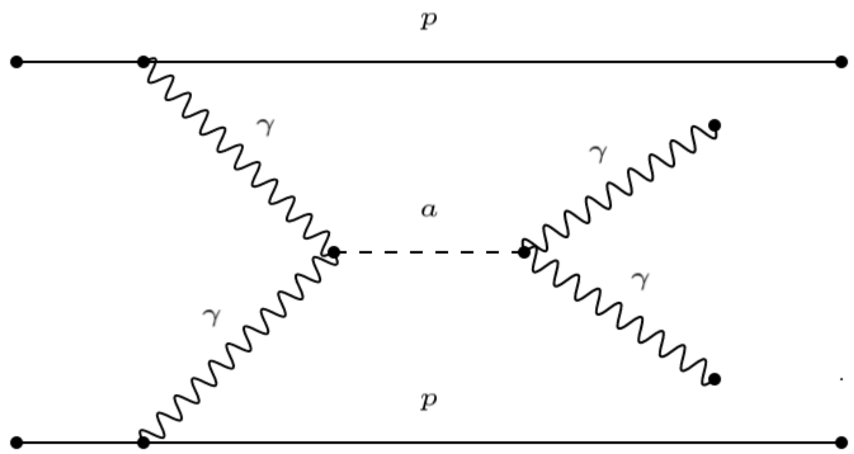

Figure 1.4: Feynman diagram illustrating light-by-light scattering mediated by an ALP (a). Source: [17]

These photons fuse to create an ALP that shortly after decays into a pair of photons. These photons could be detected in the ATLAS central detector.

This thesis is focused on pairing the detected kinematic properties (their energy) from the ATLAS central detector (detected photons) to the data from the AFP (detected protons) to see if they match. The analysis does not include only measured data, it also includes simulated data. First, a particle generator (in this case SuperChic3 [18]) is used to generate truth particles. Next, this generated data goes through a simulation process to obtain the simulated data which models the ATLAS detector.

There are many measured variables in the data, in order to reduce it there are so called "derivations". There are numerous derivations, each containing a different list of variables that suites the needs of the groups using them. In this analysis, STDM2 (Standard Model 2) derivation is used. The notable part about STDM2 is the inclusion of proton data measured by the AFP detector.

Chapter 2

Study on matching

As discussed in section physics motivation, the matching is the pairing between the kinematics between detected protons and detected photons. To be precise, we are looking at relative proton energy loss ξ. The definitions for ξ_{AFP} (protons) and $\xi_{\gamma\gamma}$ (photons) are different (Equations 2.1, 2.2, respectively [17]). By definition both the ξ_{AFP} and $\xi_{\gamma\gamma}$ values are in interval the $\langle 0; 1 \rangle$. The upper limit for $\xi_{\gamma\gamma}$ is a result of additional constraints on $m_{\gamma\gamma}$ and η coming from the law of conservation of energy.

$$\xi_{AFP} = 1 - \frac{E}{E_{beam}} \tag{2.1}$$

$$\xi_{\gamma\gamma} = m_{\gamma\gamma} \cdot \frac{e^{\pm\eta}}{2E_{beam}} \tag{2.2}$$

A short explanation of the symbols used in Equations 2.1, 2.2:

- E – Energy of the proton measured by the AFP detector
- $m_{\gamma\gamma}$ – Di-photon invariant mass
- η – Di-photon pseudorapidity, defined as $\eta = -\ln\tan\frac{\theta}{2}$, where θ is the polar angle in relation to the beam
- E_{beam} – Energy of the beam (a constant value, 6.5 TeV) [19]

Since there are two sides of the beam (side A, side C), the matching is done for both sides separately. Without the loss of generality it has been decided that in Equation 2.2, side A corresponds to positive η, side C corresponds to negative η.

2.1 Selected matching criteria

It is natural that there can not be an expectation that the compared energy loss would be exactly the same for protons and photons. This is where the matching criteria comes in. The goal is to get the highest possible amount of matches with the lowest possible amount of false positives.

The first matching criteria, also later referred to as standard (Equation 2.3), is a naive approximation that the relative difference of energy losses should be less than 10%. $\Delta\xi$ is defined as the absolute difference between proton and photon ξ ($\Delta\xi = \xi_{AFP} - \xi_{\gamma\gamma}$). Left side of the equation is divided by $\xi_{\gamma\gamma}$, it is more precise when compared to ξ_{AFP}.

$$\frac{|\Delta\xi|}{\xi_{\gamma\gamma}} < 10\% \tag{2.3}$$

The next matching criteria, labeled as 1σ matching (Equation 2.4), takes into account uncertainties on the measurements. $\sigma_{\xi_{AFP}}, \sigma_{\xi_{\gamma\gamma}}$ represent the absolute uncertainty on ξ_{AFP} and $\xi_{\gamma\gamma}$, respectively. From the theoretical point of view the values should be the same, by extension the difference between the detected numbers should be lower than the uncertainties on the measurements.

$$\Delta\xi < \sigma_{\xi_{AFP}} + \sigma_{\xi_{\gamma\gamma}} \tag{2.4}$$

There is one more matching criteria that will be taken into account, labeled as 2σ matching (Equation 2.5). It is included because of the fact that in reality there will be some energy lost from other sources than the interaction itself. But the amount of energy lost this way should be fractional compared to the total energy. To account for this additional energy loss, the threshold for match when taking uncertainties into account is doubled. By definition, $\approx 68.3\%$ of the matches should be detected by 1σ constraint and $\approx 95.5\%$ of the matches by the 2σ constraint.

$$\Delta\xi < 2\left(\sigma_{\xi_{AFP}} + \sigma_{\xi_{\gamma\gamma}}\right) \tag{2.5}$$

2.2 Photon uncertainties

The determination process of $\sigma_{\xi_{\gamma\gamma}}$ is to partially derive Equation 2.2 by its variables (in this case $m_{\gamma\gamma}$ and η). It is known that the angle is measured very precisely (the uncertainty on the angle is very small), we can approximate that there is no uncertainty on the angle. With this assumption we are left with Equation 2.6.

$$\sigma_{\xi_{\gamma\gamma}} = \sigma_{m_{\gamma\gamma}} \cdot \frac{e^{\pm\eta}}{2E_{beam}} \tag{2.6}$$

In order to determine $\sigma_{m_{\gamma\gamma}}$ (the absolute uncertainty on di-photon invariant mass), because $m_{\gamma\gamma}$ is calculated and not measured, we must mention how it is calculated (Equation 2.7).

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1-\cos\alpha)} \tag{2.7}$$

In this case α is the angle between the two photons. And again it is measured very precisely so we neglect the uncertainty on it. Again we partially derive by variables and add the results in a quadrature sum in order to determine the equation for $\sigma_{m_{\gamma\gamma}}$ (Equation 2.8).

$$\sigma_{m_{\gamma\gamma}} = \frac{m_{\gamma\gamma}}{2}\sqrt{\left(\frac{\sigma_{E_1}}{E_1}\right)^2 + \left(\frac{\sigma_{E_2}}{E_2}\right)^2} = \frac{m_{\gamma\gamma}}{2}\cdot\left(\frac{\sigma_{E_1}}{E_1} \oplus \frac{\sigma_{E_2}}{E_2}\right) \tag{2.8}$$

There has already been a study that mentions photon energy resolution uncertainty [20], the Equation 2.9 is taken directly from it.

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.9}$$

In Equation 2.9, the listed variables are:

- a – Sampling term - $9-10\%$ (GeV) (10% is used)
- b – Noise term - $350\cosh\eta$ (MeV)
- c – Constant term - 0.7%
- E – Measured energy of the photon
- η – Single photon pseudorapidity

2.3 Systematic uncertainties of the proton measurement

The already mentioned di-lepton study [6] includes a chapter with AFP systematic uncertainties.

Because proton energy is measured using the trajectory of the protons, the uncertainties are related to the position on the x-axis on the detector (with the exception of beam optics).

- Global alignment – $300\,\mu$m
- Beam optics – $50\,\mu$rad
- Local alignment – $20\,\mu$m
- Proton transportation – 2%

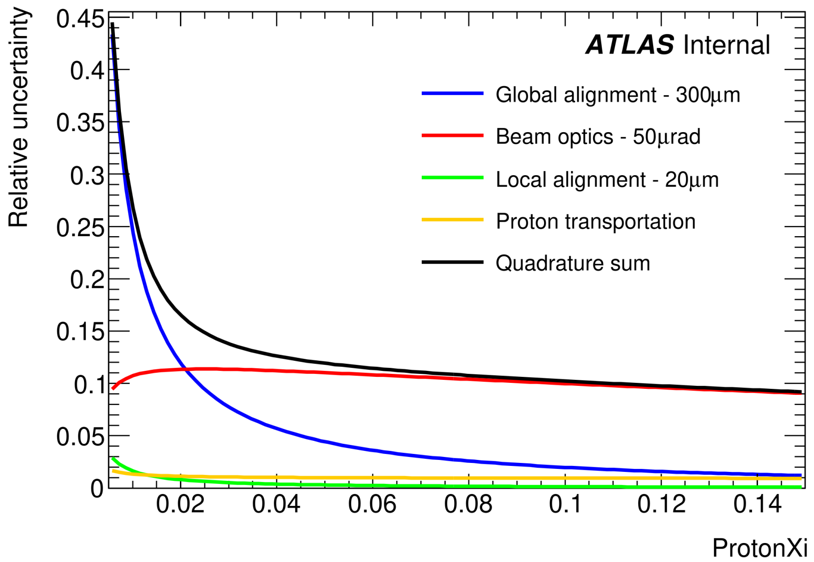

Figure 2.1: Relative proton uncertainties that are taken into account with their quadrature sum (total proton ξ uncertainty).

In order to calculate the $\sigma_{\xi_{AFP}}$ there is an approximation that calculates ξ from the x-axis position on the detector (Equation 2.10) [21]. σ_x is defined as the uncertainty for the x-axis position. For the beam optics and the proton transportation the $\sigma_x(\xi)$ is shown in Equations 2.11, 2.12, respectively [6].

$$\sigma_\xi = \frac{\sigma_x}{-119 - 328\xi} \tag{2.10}$$

$$\sigma_{x_{beam}} = -0.02 + 15.38\xi \tag{2.11}$$

$$\sigma_{x_{transport}} = 0.00508 + 1.104\xi + 2.834\xi^2 \tag{2.12}$$

The total uncertainty $\sigma_{\xi_{AFP}}$ is calculated by quadrature sum of all its parts (Equation 2.13, Figure 2.1).

$$\sigma_{\xi_{AFP}} = \sigma_{\xi_{global}} \oplus \sigma_{\xi_{beam}} \oplus \sigma_{\xi_{local}} \oplus \sigma_{\xi_{transport}} \tag{2.13}$$

2.4 Statistical uncertainty of the proton reconstruction

The statistical uncertainty of the proton resolutions can be obtained from the simulated data. While the ATLAS central detector has a dedicated group to-

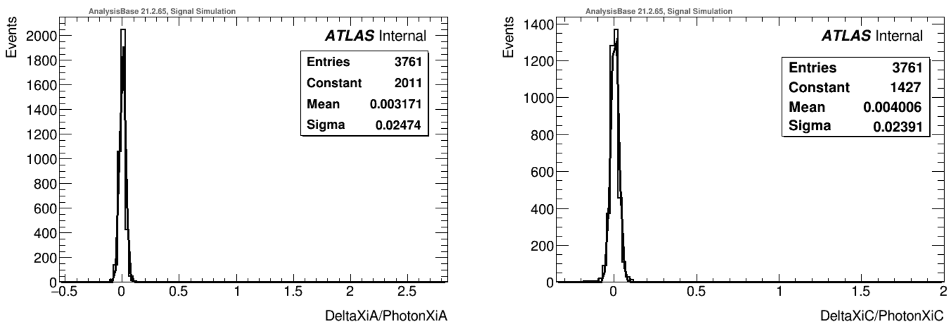

Figure 2.2: Relative $\Delta\xi$ uncertainty coming from signal simulation for side A (left) and side C (right).

wards simulation, the AFP uses `AFPFastSimTool` inside `AFPToolbox` to simulate the proton data. In order to simulate the proton data, because the STDM2 derivation of the ATLAS simulation did not include so called "truth protons", a local modification of the `AFPFastSimTool` had to be done in the group production ntuple code. The statistical uncertainty is determined by the width of the gaussian fit of the $\Delta\xi$ coming from the two simulations. The determined uncertainty for the $\Delta\xi$ is around 2.5% for both sides A and C (Figure 2.2), thus the proton statistical uncertainty is lower than 2.5%.

2.5 Proton selection

There is a requirement for the analyzed events to have exactly two photons. For protons there is a requirement that at least one proton has to be detected in the system. This leads to a large amount of events where there might not be a single proton or there might be more than one proton detected on either side of the beam. More than one proton on one side is possible, because the interactions are not one proton to one proton, but a bunch of protons into a bunch of protons.

The case of zero protons on either side is solved by not-matching the side to a photon. But where there are more protons only one can be chosen as a candidate to be matched.

There are 3 main strategies – choosing the proton with the lowest measured energy, the highest measured energy or the proton, which has the most similar energy loss to the compared photon. Early in this analysis the proton with highest energy loss was chosen, but since then it was found out that the highest efficiency is when a proton with the closest energy loss to the photon was selected.

2.6 Matching background

The matching background is the randomness of the matching. Because the events are uncorrelated, there are two main strategies – one is to take proton data from a different event and try to match it to the photons from current event (later referred to as "mixed case"), the other is to switch sides when matching (matching side A proton to side C photon and vice versa, later referred to as "switched sides").

After matching, the similarities or differences between the number of matches of nominal matching and mixed matching can serve as a background model. This model shows how many matches from the number of nominal matches are a random coincidence.

In order to eliminate any unwanted correlation the proton selection must be done right before matching (after mixing events or switching sides).

For the mixed case there is a statistics that can be easily made. The results should, by definition, be similar when taking proton information from event $n-1$, $n-2$, ...The number of matches in these can be fitted using a Gaussian function. The width of the Gaussian fit results in the uncertainty on the number of random matches.

In the implementation the statistics is not done using the number of matches, but using the difference between the number of nominal matches and the number of mixed matches. The width of the Gaussian is still the same in both cases.

At this time, the number of matches for all three matching criteria (Table 4.1) was known. As a result, the matching background was analyzed only for 2σ matching.

Chapter 3

Software development

This chapter is dedicated to the software part of the thesis. It starts with the introduction to the software used, ROOT, which is followed by the development plan. Next, there are sections dedicated to each part of the developed software – the shared resources, the script that creates an ntuple focused on the analysis, the script that analyses the variation between random and nominal matches and a script that visualises the systematic uncertainties.

3.1 ROOT

ROOT [22, 23] is a multi-platform software developed by CERN with intended use of big-data storage, analysis and visualisation. The main applications are plotting and storing complex or high amounts of data.

The storage files are self-descriptive and compressed binary files. One of the main benefits of ROOT's storage solution is that one file can be split into several smaller files that are chained and accessed as a single object. A simple file containing several entries for variables can also be called a ntuple.

The main way to control ROOT is using a command line interface (CLI). By default it uses a C++ interpreter Cling, but there is also an option to have an interactive session using a Python interpreter. This allows the creation of very complex scripts with custom classes and structures.

Creating scripts and interpreting them using the CLI is not the only way to interact with ROOT files. Because interpreting scripts can be time ineffective, there is also an option to compile the scripts or compile them as separate applications [24].

For visualization there is also a simple graphic user interface (GUI). The basic window is the `TBrowser` (Figure 3.1), which serves as a tool to view the structure of a ROOT file. It also includes a part with `TCanvas`, a canvas that is used for plotting graphs.

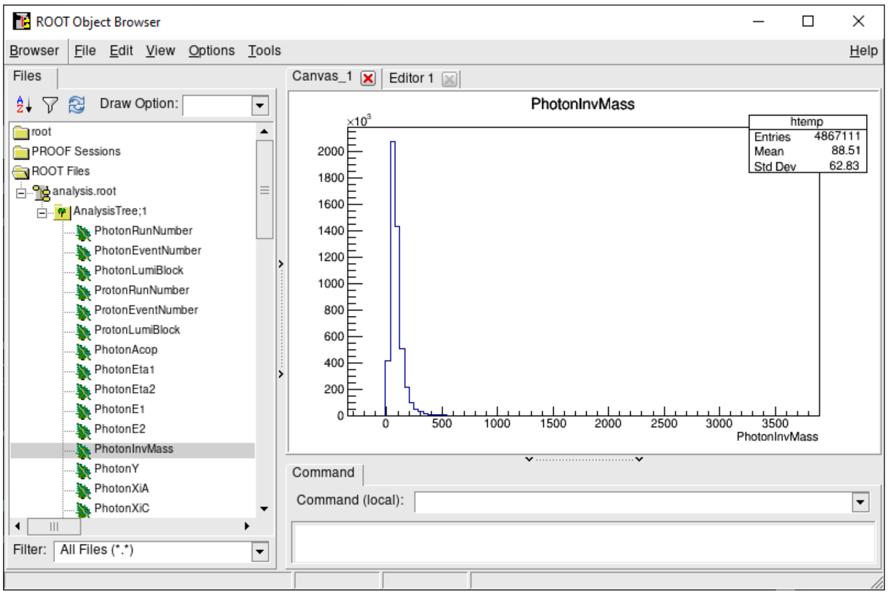

Figure 3.1: An example window of ROOT's `TBrowser` GUI.

Both the CLI and GUI include tools to modify the visualisation of data. The visualised data can be also modified while visualising.

3.2 Development plan

The complexity of the physics resulted in my decision to choose Iterative development over Waterfall or Agile. The main advantage of Iterative development is that the software is developed in incremental iterations where each iteration includes additional functionality. In contrast, when using Agile development strategy, the iterations contain all functionality. And lastly the Waterfall strategy does not use iterations and is slow to adopt changes.

The first part was identifying which variables are available in the files that serve as input files for my developed code. In case any needed non-calculable variable was missing, it was important to do a local modification of the group ntuple production code and include that modification later as well. In the simple approach of matching, there are no calculated variables needed, but when taking the uncertainties into account, some variables had to be added. In order to focus the information in the main group ntuple for matching, a need for modifying scrips arises. The main goal if the scripts is to read, modify and write data in ROOT files.

The first iterations of developed code included only functions that calculated the uncertainty on photon energy and di-photon invariant mass. Next iterations extended the functionality to include all photon uncertainties. After that proton uncertainties were added. Lastly the random background functionality was needed, at that point the software got split into two parts – matching and random matching. The final iteration includes the change of structure (unifying shared resources).

As discussed in the previous section, ROOT's scripts can be interpreted, compiled or compiled as standalone. Because this developed code can be still used in the future and changes in the code might be wanted, the option of compiling the code as standalone application is the least preferable option. The lack of speed of interpreting the code leaves compiling the code, but still running it as a script, as the best option, which will be endorsed in this case.

3.3 Shared resources – `shared.cxx`

In order to reduce the possibilities of errors coming from inattention, having one file with functions needed by two separate scripts is an option. In these two files are functions and their definitions that are used by both `matching.cxx` and `random.cxx` scripts (Listing 3.2).

There is one function in particular in this file that I would like to mention – `quadratureSum` (Listing 3.1). There is a need to calculate a quadrature sum of a particular list of variables more than once. Instead of hard-coding the quadrature sum, I decided it is a good idea to write a function that takes beforehand an unknown number of values and calculates the quadrature sum. The easiest way to do this is to have a function with one parameter, a vector of values. Next you have an accumulator and for each value inside the vector you add the squared value into the accumulator. In the end it returns the square root of the accumulator. This function provides more readable code with less possibilities to make a mistake from lack of attention.

```
Double_t quadratureSum ( vector<Double_t> list ) {
  Double_t acc = 0;
  for ( auto & i : list ) {
    acc += pow( i, 2 );
  }
  return sqrt( acc );
}
```

Listing 3.1: Implementation of the function `quadratureSum`.

```
Double_t quadratureSum ( vector<Double_t> list );

Double_t calcPhotonSigmaE ( Double_t & PhotonE, Double_t &
    PhotonEta );

Double_t calcProtonSigmaXi ( Double_t & ProtonXi );

Double_t calcDeltaXi ( Double_t & ProtonXi, Double_t & PhotonXi );

Double_t minimizeDeltaXi ( vector<Double_T> & Protons, Double_t &
    PhotonXi );
```

Listing 3.2: Definitions of the functions in `shared.cxx`.

3.4 Matching script – `matching.cxx`

The main point of this script is to create a ntuple with variables that are specific to matching. This means it reduces the amount of unnecessary data and calculates a few matching specific variables that are missing in the ROOT ntuple that is produced using the group analysis code.

There is no specifically required time complexity, but the code should run with linear time with respect with the number of events in the ROOT ntuple. The final version of this script has execution time of around two minutes on CERN's lxplus servers for both cases – interpreting and compiling the code.

The main problem with the input ntuple is the structure of the data. In the input ntuple there are two `TTree`s, one with Photon data and the other with proton data. Thankfully in both `TTree`s are two variables that are specific to the event – `EventNumber` and `RunNumber`. Since ROOT links `TBranch`es to C++ variables using references, it is important to have the data from same event in the variables. In order to achieve that we can use ROOT's `TTree:AddFriend` method to access both `TTree`s at the same time. But in order to use that method we need to set unique identifiers for specific event. `TTree:BuildIndex` method that can set primary and secondary index of `TTree`. In this case `RunNumber` is the primary index, `EventNumber` the secondary index.

In order to analyze the mixed-case and switched sides cases for the random matching, there are commented parts of the code that represent the different proton to photon matching styles. The reason for not including all three styles is the execution time. In order to eliminate any unwanted correlation coming from the proton selection, the pairing of the proton to photon had to be done separately for each event for each matching style. In case of switched sides matching the increase would be coming only from proton selection, but in mixed case there would be either increase in memory usage by copying the data (with a chance of a failure to allocate memory) or the method `TTree:GetEntry` would be called three times as much, which would result in a significant time increase.

The rest of the script is straight-forward. Iterate over all events that have at least one proton in the system (this condition is important for matching) – we can iterate through the TTree with proton data. All calculations are done for each event separately (Listing 3.3).

There is one more comment about this script that has to be discussed. One iteration of this script included time measurement for each part of the script – script initialization, the calculation and saving the file. For these time measurements std::chrono was used. This inclusion increased execution time from around 2 minutes to over 10 minutes. Because of this, there are only printed lines in the standard output about the completion of each part of the script.

```
//Conversion to GeV
PhotonE1 /= 1000;
PhotonE2 /= 1000;

//Calculating PhotonSigmas
PhotonSigmaE1 = calcPhotonSigmaE( PhotonE1, PhotonEta1 );
PhotonSigmaE2 = calcPhotonSigmaE( PhotonE2, PhotonEta2 );
PhotonSigmaInvMass = PhotonInvMass / 2 * quadratureSum( {
    PhotonSigmaE1 / PhotonE1, PhotonSigmaE2 / PhotonE2 } );
PhotonSigmaXiA = PhotonXiA / PhotonInvMass * PhotonSigmaInvMass;
PhotonSigmaXiC = PhotonXiC / PhotonInvMass * PhotonSigmaInvMass;

//Minimizing DeltaXi -- standard case
nProtonsA = ProtonCandsXiA->size();
nProtonsC = ProtonCandsXiC->size();
nProtons = nProtonsA + nProtonsC;
ProtonXiA = minimizeDeltaXi ( *ProtonCandsXiA, PhotonXiA );
ProtonXiC = minimizeDeltaXi ( *ProtonCandsXiC, PhotonXiC );
DeltaXiA = calcDeltaXi ( ProtonXiA, PhotonXiA );
ProtonSigmaXiA = calcProtonSigmaXi ( ProtonXiA );
DeltaXiC = calcDeltaXi ( ProtonXiC, PhotonXiC );
ProtonSigmaXiC = calcProtonSigmaXi ( ProtonXiC );
```

Listing 3.3: Calculations done for each event for standard photon-proton matching.

3.5 Random matching script – random.cxx

Another script that is important is a script that gathers the variation in number of events between nominal matching and mixed matching, where proton data is used from another event. The input for this script is the same as for the previous one, a ROOT ntuple that is created using the group analysis code. This means the small problem of having data in two TTrees is resolved the same way. The output is a ROOT ntuple containing one TTree and inside it four TBranches. Each branch contains the information on variation of random matches compared to nominal matches. The branches are `A`, `C`, `AorC`, `AandC` – representing the side requirement for the matches.

At first this functionality of mixed case photon-proton matching was included in the previous script. But in this case the goal is to get the statistic variation between random matching and nominal matching. In order to do that, there have to be statistical iterations of the mixed selection case. The first iteration is the proton information from directly previous event, in the next iteration proton information from event "n-2" is taken. In order to remove unwanted correlation, the proton selection must be done separately for each statistical iteration for each event.

This change introduced a huge priority into speed optimization of the script. The decision was made that $n = 1000$ should be enough statistical data. That results that this script has to iterate through 4.8 million events a thousand times. This implies the importance of constant iteration time, which equals total execution linear time with respect on the number of iterations. Anything asymptotically slower would result in very long execution times.

The first iteration of this script was a naive one. The main point of it was to get any data without a focus on speed. In each statistical iteration for each event the script accessed the photon and proton data in the ROOT ntuple separately. This is where the biggest downside of this script comes in. Accessing data in ntuples has linear time complexity based on the location of previously accessed event.

This resulted in very long and increasing execution time per iteration that can be estimated using a linear function $t(n) = 120 + n \cdot 80$ seconds. The Equation approximates that iteration n will take $120 + n \cdot 80$ seconds to finish. Time complexity was not satisfying enough for this iteration of the code.

Realising this aspect of accessing data in the ntuple multiple times has lead to a more satisfying iteration of this script. It was necessary to have constant access times for both photon and proton data for each iteration. To achieve that, the data was copied into more appropriate data structures. The main concern is the constant access time. There are many structures satisfy that condition, but for the ease of implementation `vector` from std was chosen. Another candidate was standard C++ array, but the ease of use of `vector` and the readability of the code when using it were the main decision arguments.

This change resulted in constant time per iteration as was needed. Even though it introduced a constant time increase before first iteration – copying the events into `vector` – the iteration execution time was around 7.8 seconds. Worth mentioning is that this is when the script was interpreted, not compiled. By compiling the iteration time was lowered to around 1.5 seconds.

There is one problem that can unexpectedly occur. There can be a failure to allocate memory while copying data into the before-mentioned `vector`. Therefore, a try-catch block was added to exit the application gracefully in case of a failure. During testing, this has not happened once, but if there is a larger amount of data to process it is possible this limitation might occur.

For $n = 1000$ the execution time is around 30 minutes (when compiled), which is not unbearably long, but can be still improved. Another way of optimizing it would be using more than one thread. ROOT has options to use multi-threading [25], but I believe the problems this option brings (mainly thread-safety) outweigh the increase of speed.

3.6 Visualisation of systematic uncertainties – systematic.cxx

The goal of this script is to visualize the share of each part of systematic uncertainties on the total uncertainty. The reason to include it into the thesis is to highlight how simple it is to visualize data using ROOT. Figure 2.1 was created using this script.

```
auto f1 = new TF1("global", "abs(0.3/(-119-328*x))/x", 0.005,
    0.15);
auto f2 = new TF1("beam", "abs((-0.0227+15.381*x)/(-119-328*x))/x
    ");
auto f3 = new TF1("local", "abs((0.02)/(-119-328*x))/x");
auto f4 = new TF1("protonTransport", "abs((0.00508+1.104*x+2.834*x
    *x)/(-119-328*x))/x");
auto ff = new TF1("quadrature sum", "sqrt(((-0.0227+15.381*x)
    /(-119-328*x)/x)^2 + ((0.3/(-119-328*x))/x)^2 + (abs((0.02)
    /(-119-328*x))/x)^2 + (abs((0.00508+1.104*x+2.834*x*x)
    /(-119-328*x))/x)^2)");
```

Listing 3.4: Definitions of the functions that are displayed in the Figure generated by systematic.cxx.

Chapter **4**

Results

In this chapter are discussed all the results from the data created using the developed code. First, the cuts on the data are mentioned. Next, the final number of detected matches is presented. Following is a discussion on the determined uncertainties. Lastly, the randomness of the matching and the ALP signal matching efficiencies are discussed.

4.1 Initial data selection

Important conditions that have not been listed before – the 2017 data (not simulations) needs to be so called blinded – the photon acoplanarity must be greater than 0.01. The acoplanarity is the angle from being back-to back. The reason for blinding is to avoid any bias in the analysis which could influence the detection or exclusion of an ALP signal. It is a standard procedure for any search for a new particle.

Independent of the blinding there are also cleanup cuts $\xi_{AFP} > 0.01$ and $\xi_{\gamma\gamma} > 0.01$. The reason for these cuts is the detection range of the AFP detectors.

4.2 Determined uncertainties

The main uncertainties taken into account are the photon energy resolution uncertainty, systematic uncertainties on the proton resolution and statistic uncertainties on the proton resolution coming from the simulation. In the 2017 data, the photon energy resolution uncertainty contributes to roughly 2.5% uncertainty on $\xi_{\gamma\gamma}$, as seen in Figure 4.1 for $\xi_{\gamma\gamma} \in \langle 0.02; 0.1 \rangle$. The proton systematic uncertainties are by far the biggest with around 10% uncertainty as seen in Figure 2.1. And the statistical uncertainty on $\Delta\xi$ is around 2.5% (Figure 2.2). This 2.5% uncertainty coming from the simulation is dominated by proton uncertainty, as the photons in the signal simulation have higher

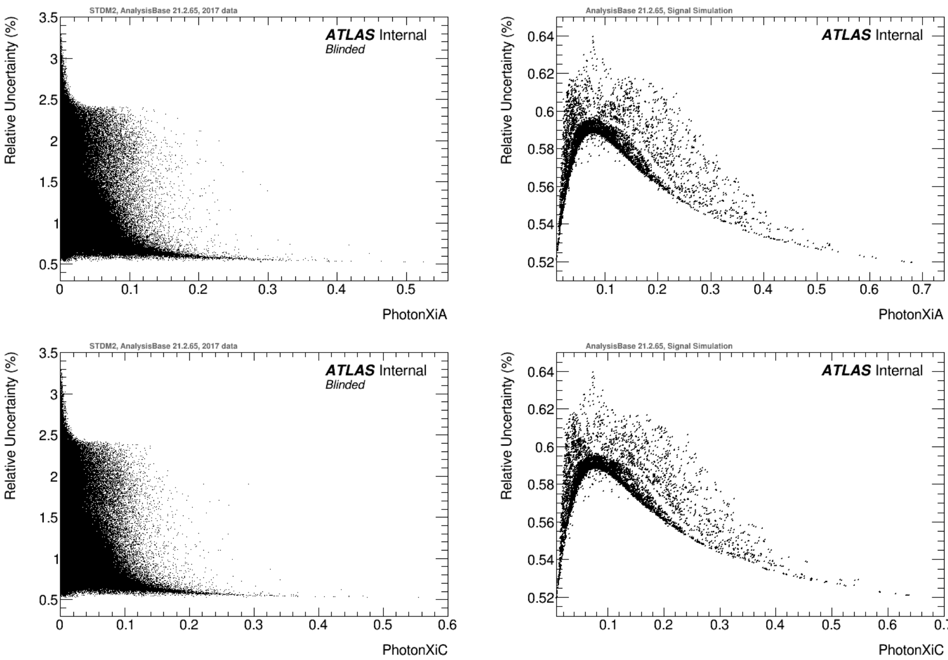

Figure 4.1: Relative uncertainty on $\xi_{\gamma\gamma}$ as a function of $\xi_{\gamma\gamma}$ for side A (top) and side C (bottom), for 2017 data (left) and signal simulation (right).

energy compared to the photons in the 2017 data, which leads to a photon uncertainty of around 0.6%, as seen in Figure 4.1.

4.3 Number of matches

The number of matches in 2017 data for each side and each matching condition is shown in Table 4.1. The low amount of matches for 10% matching can be explained by the high systematic uncertainty discussed in section systematic uncertainties.

The di-photon invariant mass distributions for A and C matching for all three matching criteria is shown in Figures 4.2, 4.3, 4.4. The figures show that the di-photon invariant masses look fairly similar among all matching criteria with a peak at around 500 GeV. Clearly, the 1σ and 2σ matching conditions increased the amount of matches.

	A	C	A or C	A and C
10%	34 442	47 849	82 224	67
1σ	47 421	66 642	113 943	120
2σ	95 971	135 386	230 848	509

Table 4.1: Number of matches across side A, side C, side A or side C, and side A and side C for 10%, 1σ and 2σ matching (2017 data).

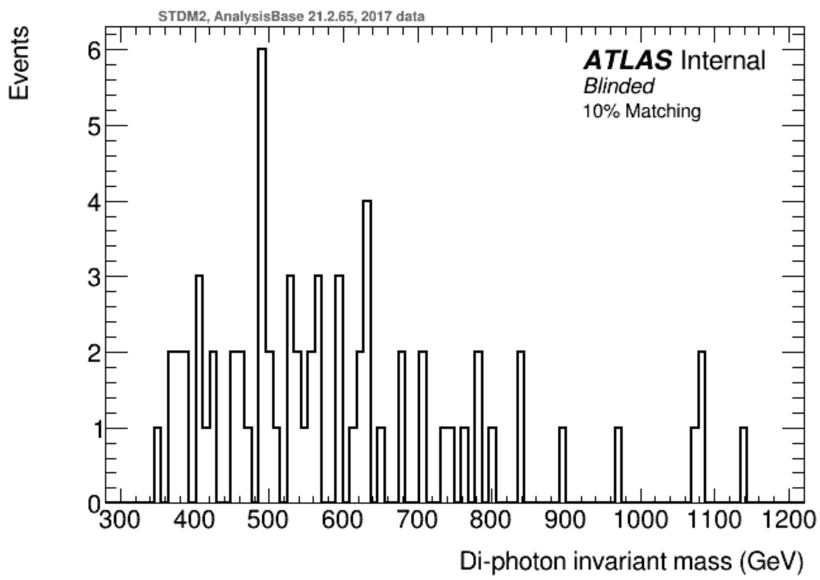

Figure 4.2: Di-photon invariant mass of the matched events on sides A and C (at the same time) for 10% matching.

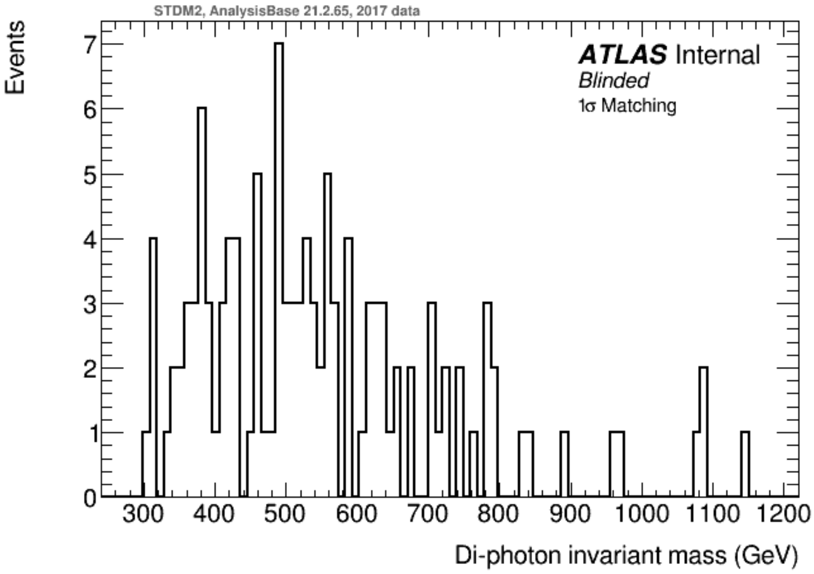

Figure 4.3: Di-photon invariant mass of the matched events on sides A and C (at the same time) for 1σ matching.

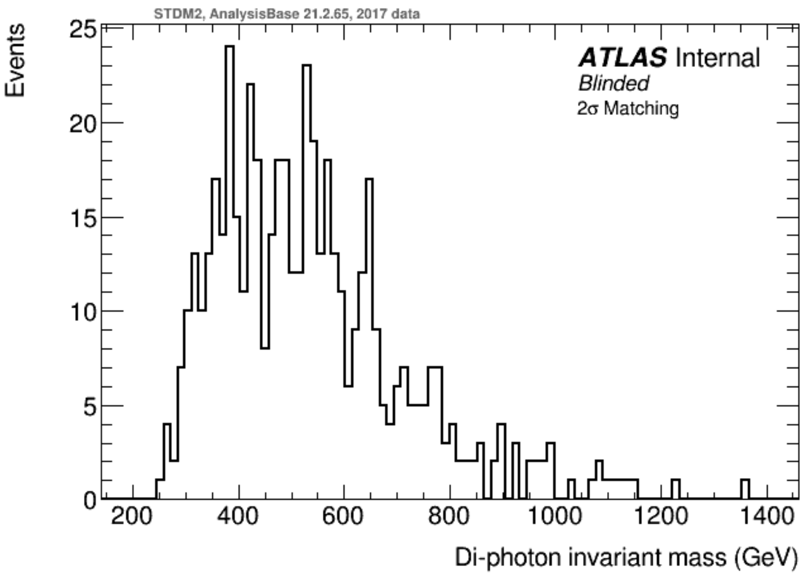

Figure 4.4: Di-photon invariant mass of the matched events on sides A and C (at the same time) for 2σ matching.

	A	C	A or C	A and C
10%	34 534	47 870	82 338	66
1σ	47 338	67 000	114 209	129
2σ	95 469	135 636	230 580	525

Table 4.2: Number of matches for "mixed" case (di-photon information taken from event "n" and AFP information taken from event "n-1").

	A	C	A or C	A and C
10%	34 340	48 130	82 399	71
1σ	47 296	67 068	114 224	140
2σ	95 830	136 466	231 745	551

Table 4.3: Number of matches for "switched sides" case.

	Nominal (blinded)		Random matching	Uncertainty
A	95 971		95 469	$\pm$ 271
C	135 386		135 636	$\pm$ 307
A or C	230 848		230 580	$\pm$ 415
A and C	509		525	$\pm$ 22

Table 4.4: Comparison between the number of nominal matches and random matches with uncertainty (2σ matching).

4.4 Random matches (pile-up background)

The number of matches for "mixed" and "switched sides" cases are shown in Tables 4.2 and 4.3, respectively. The direct comparison between nominal and mixed matches with the respective uncertainty gained from $n = 1000$ statistical iterations are shown in Table 4.4.

The main conclusion is that the number of nominal matches is compatible within the determined statistical uncertainties with the number of random matches. This concludes that the matching done in this analysis was a series of random coincidences.

4.5 ALP signal matching efficiencies

The ALP signal matching efficiencies are analyzed by performing the matching criteria on the simulated data (Table 4.5). In this case the ξ data cleaning cut is $\xi \in \langle 0.02; 0.1\rangle$ for both ξ_{AFP} and $\xi_{\gamma\gamma}$. As expected, the efficiencies are lower than at the generator level. These efficiencies are important to determine the sensitivity to detect an axion-like-particle with the ATLAS central and AFP detectors.

	A	C	A or C	A and C
10%	3736	3725	5752	1709
	50.6%	50.5%	78.0%	23.2%
1σ	3745	3739	5770	1714
	50.8%	50.7%	78.2%	23.2%
2σ	3757	3760	5790	1727
	50.9%	51.0%	78.5%	23.4%
Generator – 10%	66.9%	66.9%	89.1%	25.0%

Table 4.5: The number of matches and their percentages related to the total number of di-photon events with at least one proton present (7377 events) for all three matching criteria. The table also lists the percentages on the generator level for 10% matching.

Conclusions

The main goal of this thesis was to optimize the matching criteria in proton-proton interactions with exactly two photons observed. The important extension is the that a simple matching criteria was extended to include the statistical and systematic uncertainties in the photon and proton measurements. The optimization was used to study the number of events in the recorded data and the efficiencies of a simulated ALP signal. Also another goal was to provide clarity to the randomness of the matching. The large data set provided by the ATLAS central and AFP detectors is a challenge from the point of view of Information Technology, and the successful contribution to the physics data analysis fulfills the thesis tasks in all aspects.

An important result is that the analyzed similarity between ξ_{AFP} and $\xi_{\gamma\gamma}$ is a random coincidence instead of a physics process with low randomness. The three matching criteria were compared by the number of matches and the di-photon invariant mass distribution. Also the uncertainties on proton and photon reconstruction were discussed. It was determined that the proton systematic uncertainties are much larger than the statistical proton uncertainty and the di-photon reconstruction uncertainty. Lastly, the ALP signal matching efficiencies were compared against efficiencies determined on the generator level.

This thesis provides an important step in the analysis of light-by-light scattering mediated by an Axion-Like-Particle (ALP). The written software can serve for similar analysis and for reproduction of these results. The code is written in a way that can be changed in case of an update of the selection of the di-photons or changes in the proton detection system. Another use for the software is to serve as an example of how to work with correlated data in two different `TTree`s, or how to visualize functions using ROOT.

Bibliography

[1] ATLAS Collaboration. Observation of light-by-light scattering in ultra-peripheral Pb+Pb collisions with the ATLAS detector. *Phys. Rev. Lett.*, volume 123, no. 5, 2019: p. 052001, doi:10.1103/PhysRevLett.123.052001, 1904.03536.

[2] d'Enterria, D.; da Silveira, G. G. Observing light-by-light scattering at the Large Hadron Collider. *Phys. Rev. Lett.*, volume 111, 2013: p. 080405, doi:10.1103/PhysRevLett.111.080405,10.1103/PhysRevLett.116.129901, [Erratum: Phys. Rev. Lett.116,no.12,129901(2016)], 1305.7142.

[3] Euler, H. On the scattering of light by light according to Dirac's theory. *Annalen Phys.*, volume 26, no. 5, 1936: pp. 398–448, doi:10.1002/andp.19364180503, [Annalen Phys.418,no.5,398(1936)].

[4] ATLAS Collaboration. Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC. *Nature Phys.*, volume 13, no. 9, 2017: pp. 852–858, doi:10.1038/nphys4208, 1702.01625.

[5] Baldenegro, C.; Fichet, S.; et al. Searching for axion-like particles with proton tagging at the LHC. *JHEP*, volume 06, 2018: p. 131, doi:10.1007/JHEP06(2018)131, 1803.10835.

[6] Beresford, L.; Bussey, P.; et al. Measurement of proton-tagged lepton pairs in photon fusion using the ATLAS Forward Proton spectrometer. Technical report ATL-COM-PHYS-2020-205, CERN, Geneva, Mar 2020, [Online; accessed April 10, 2020]. Available from: https://cds.cern.ch/record/2712727

[7] About CERN. [Online; accessed May 28, 2020]. Available from: https://home.cern/about

[8] The birth of the Web. [Online; accessed May 28, 2020]. Available from: https://home.cern/science/computing/birth-web

[9] Evans, L.; Bryant, P. LHC Machine. *JINST*, volume 3, 2008: p. S08001, doi:10.1088/1748-0221/3/08/S08001.

[10] The Large Hadron Collider. [Online; accessed May 28, 2020]. Available from: `https://home.cern/science/accelerators/large-hadron-collider`

[11] Horvath, A. The LHC experiments and the preaccelerators. 2006, [Online image; accessed May 28, 2020]. Available from: `https://commons.wikimedia.org/wiki/File:LHC.svg`

[12] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, volume 3, 2008: p. S08003, doi:10.1088/1748-0221/3/08/S08003.

[13] ATLAS Collaboration. Performance of the ATLAS detector using first collision data. *JHEP*, volume 09, 2010: p. 056, doi:10.1007/JHEP09(2010)056, 1005.5254.

[14] About the ATLAS Experiment. [Online; accessed May 28, 2020]. Available from: `https://atlas.cern/discover/about`

[15] Adamczyk, L.; et al. Technical Design Report for the ATLAS Forward Proton Detector. 2015.

[16] Pequenao, J. Computer generated image of the whole ATLAS detector. 2008, [Online image; accessed May 28, 2020]. Available from: `https://cds.cern.ch/images/CERN-GE-0803012-01`

[17] Sopczak, A.; Bussey, P.; et al. Search for an Axion-Like Particle in Light-by-Light scattering using the ATLAS central detector and the ATLAS Forward Proton detector. Technical report ATL-COM-PHYS-2020-238, CERN, Geneva, Mar 2020, [Online; accessed June 1, 2020]. Available from: `https://cds.cern.ch/record/2714416`

[18] Harland-Lang, L. A.; Khoze, V. A.; et al. Photon-Photon Collisions with SuperChic. *CERN Proc.*, volume 1, 2018: p. 59, doi:10.23727/CERN-Proceedings-2018-001.59, 1709.00176.

[19] Todesco, E.; Wenninger, J. Large Hadron Collider momentum calibration and accuracy. *Phys. Rev. Accel. Beams*, volume 20, 2017: p. 081003, doi:10.1103/PhysRevAccelBeams.20.081003. Available from: `https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.081003`

[20] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. 2014, doi:10.1140/epjc/s10052-014-3071-4, 1407.5063.

[21] Staszewski, R. Towards optics uncertainty. 2019, [Private conversation].

[22] Brun, R.; Rademakers, F. ROOT – An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, volume 389, no. 1, 1997: pp. 81 – 86, ISSN 0168-9002, doi:10.1016/S0168-9002(97)00048-X.

[23] About ROOT. [Online; accessed May 28, 2020]. Available from: `https://root.cern.ch/about-root`

[24] ROOT/Getting Started/Many Ways to Use ROOT. [Online; accessed May 28, 2020]. Available from: `https://en.wikibooks.org/wiki/ROOT/Getting_Started/Many_Ways_to_Use_ROOT`

[25] ROOT User's Guide. [Online; accessed May 28, 2020]. Available from: `https://root.cern.ch/root/htmldoc/guides/users-guide/ROOTUsersGuide.html#threads`

Appendix A

Acronyms

AFP Atlas forward proton

ALP Axion-like-patricle

ATLAS A Toroidal LHC Apparatus

CERN Conseil Européen pour la recherche nucléaire

CLI Command line interface

GUI Graphical user interface

IP Interaction point

LbyL Light-by-light

SM Standard model

Appendix **B**

Contents of enclosed SD card

```
readme.txt ........................ the file with SD card contents description
thesis.pdf ..................................... the thesis text in PDF format
implementation ..................................... implementation sources
└── input ................................ example input files for the software
thesis ................... the directory of LaTeX source codes of the thesis
└── figures ...................... the directory with figures in .png format
```