Analysis of the Track Reconstruction Algorithm in the ALFA detector of the ATLAS experiment

DESY Summer Student Programme, 2014

Krzysztof Janas

AGH University of Science and Technology, Cracow, Poland

Supervisors
Wolfgang Friebel, Karl-Heinz Hiller

31th of August 2014

Abstract

During the Summerstudent Program I dealt with the ALFA experiment, which is focused on elastic scattering in pp collisions. I was using the *ROOT* framework to write macros that analyzed the results of the Track Reconstruction Algorithm. Seven weeks is too short a time period to come up with a new algorithm or even check all its features. I dealt with several issues and focused on the influence of the detector geometry on the reconstruction of tracks in ALFA. I wrote macros for visualizing each event, I tested the imperfections in the detector geometry and checked the distribution of residuals for all fibers in the detectors. I also tested the efficiency of each fiber and each layer. With the implemented geometry, I could test a new method to calculate positions of each track.

My latest work was preparing data for the new algorithm, and also new methods for reconstructing or looking for showers. This part of the work is not finished yet, so it is only mentioned at the end of this paper.

Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie
AGH

DESY

Contents

Dedykuję tę pracę mojej Gwiazdeczce . . .

I dedicate this work to my girlfriend . . .

1 Introduction to the ALFA experiment

1.1 The ALFA subdetector

The following two subsections have been taken from [1].
ATLAS is a multi-purpose detector designed to study elementary processes in proton-proton interaction at the TeV energy scale. The main aim of one of its subdetectors - the ALFA detector (Absolute Luminosity For ATLAS) is to measure the total cross section for elastic scattering in pp collisions. Two tracking stations are placed on each side of the central ATLAS detector at distances of $238m$ and $241m$ from the interaction point. We can see

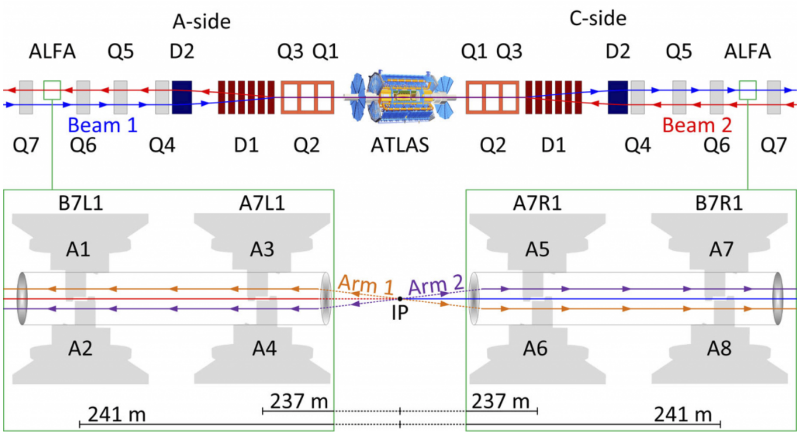


Figure 1: A sketch showing the positions of the ALFA tracking stations (not to scale) - from [1]

the name of each station in **Fig. 1.** Each station is comprised of two pots - lower and upper. The eight pots allow to test elastic-scattering event in two arms: arm 1 - pots A1, A3, A6, A8; and arm 2 - pots A2, A4, A5, A7. Theoretical basics of elastic scattering is not the essence of my work, so I won't explain it here.

The most important part of this introduction is the construction of a single pot. In **Fig. 2.** we can see one station, where the most significant parts are the two main detectors (MDs). Elastically scattered protons are detected there. Two dedicated overlap detectors (ODs) measure the distance between upper and lower MDs, but although they are an important part of the experiment, they are not needed for track reconstruction in the MDs.

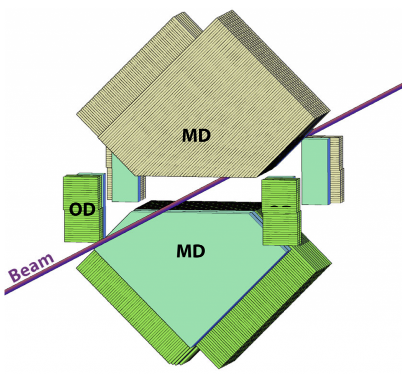


Figure 2: One station with two MDs and two ODs with highlighted beam - from [1]

Each MD consists of 2 times 10 layers of 64 square scintillating fibers each, with $0.5mm$ side length glued on titanium plates. The fibers on the front and back sides of each titanium plate are arranged at angles of $\pm 45°$ with respect to the y-axis. The projection perpendicular to the fiber axes define the u and v coordinates which are used in track reconstruction. The individual fiber layers are staggered by multiples of 1/10 of the fiber size to improve the position resolution[1].

1.2 The Track Reconstruction Algorithm

The local tracks in the MDs are reconstructed from the hit pattern of traversing protons in the scintillating fiber layers. The reconstruction assumes that the proton pass through the fiber detector perpendicularly. The first step of the reconstruction is to determine the u and v coordinates from the two sets of ten layers which have the same orientation. The best estimate of the track positions is given by the overlap region of the fiducial areas of all the fibers. As illustrated in **Fig. 3.**, the staggering of the fibers narrows the overlap region and thereby improves the resolution. The center of the overlap region gives the u and v coordinate, while the width determines the resolution. Pairs of u and v coordinates are transformed to spatial positions in the beam coordinate system.

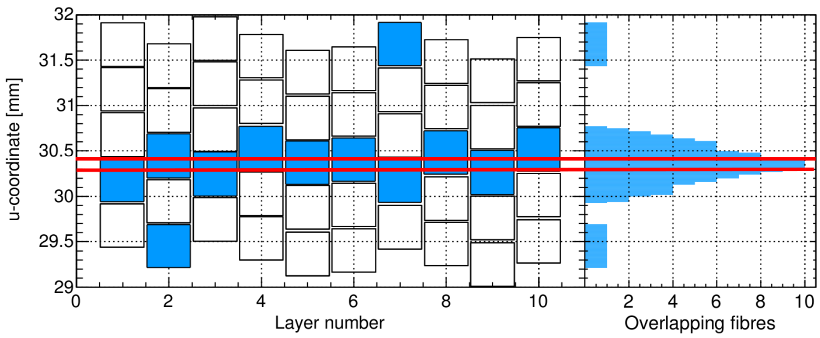


Figure 3: Sample hit pattern of a proton trajectory and histogram with overlap, red strips means corridor edges from section 6. The original picture comes from [1]

To exclude problematic events (such as events with hadronic showers or a high noise level), a few cuts are made to the data. Fiber layers with more than ten hits are discarded, and at least three layers are required to have between one and three hit fibers. Finally, the u and v coordinates must be formed from at least three overlapping hit fibers.

1.3 Cuts for the selection of elastic events

To analyze the track reconstruction algorithm in my work, it is enough to use simpler cuts than in [1] to enhance the fraction of elastic events - [2].

1. Number of tracks must to be equal 1. Multiple tracks can be due to non-elastic phenomena (showers, etc.)

2. A track must be reconstructed in each of the four detectors from at least one of two arms.

[1]This description comes from the experiment documentation. As we can see later, the layers are not staggered as regularly as shown.

3. The minimum number of fibers in a reconstructed track[2] is equal to 7.

4. Some cuts in the y coordinate are used to remove beam halo events. I used the cuts calculated in [1].

5. Similar track positions in adjacent pots. The cuts used for the difference in inner and outer pots are the following (in mm):

   - In x coordinate: $|1.2x_i - x_o| < 0.14$.
   - In y coordinate: $|0.954y_i - y_o| < 0.17$.

   The cuts are simpler than in [1], and I adopted them to remove most of the background (it is especially important to get rid of halo events). The index o stands for outer pot and i - inner pot.

6. Similar track positions in opposite inner pots. The cuts which I used are similar to the cuts in [1], especially for the y coordinate they are the same. For simplicity, I replaced an ellipse by a hexagon for the x coordinate. The following cuts are in mm:

   - In x coordinate[3]: $|x_A| < 1.25$, $|x_C| < 1.25$, $|x_A + x_B| < 1.0$.
   - In y coordinate: $|y_A + y_C| < 3$.

   The indices A and C denote the side of ALFA, as we can see in **Fig. 1.**

In **Fig. 4**. results of the cuts used are shown. After using those cuts I produced new *ntuple*[4] files with almost exclusively elastic data, which I could use sor further analysis.

2 Event visualization

To better see the results of the algorithm, I prepared my own visualization macro for events. The first version - *DrawEvent.h* - takes two parameters: event number and pot number, and generates a simple canvas in ROOT - see **Fig. 5.**, where all the fibers in each layer in u and v are drawn. The hit fibers are colored in yellow, and fibers selected for reconstruction are marked with green color. The axes show the u and v coordinates in the detector reference frame.

The fiber positions are taken from metrology files. They form straight lines with slope a (close to 1 or -1) and intercept b. Positions in u and v were calculated with the assumption that the fibers are parallel and slopes are exactly 1 or -1. The u, v coordinates depend on x, y as follows:

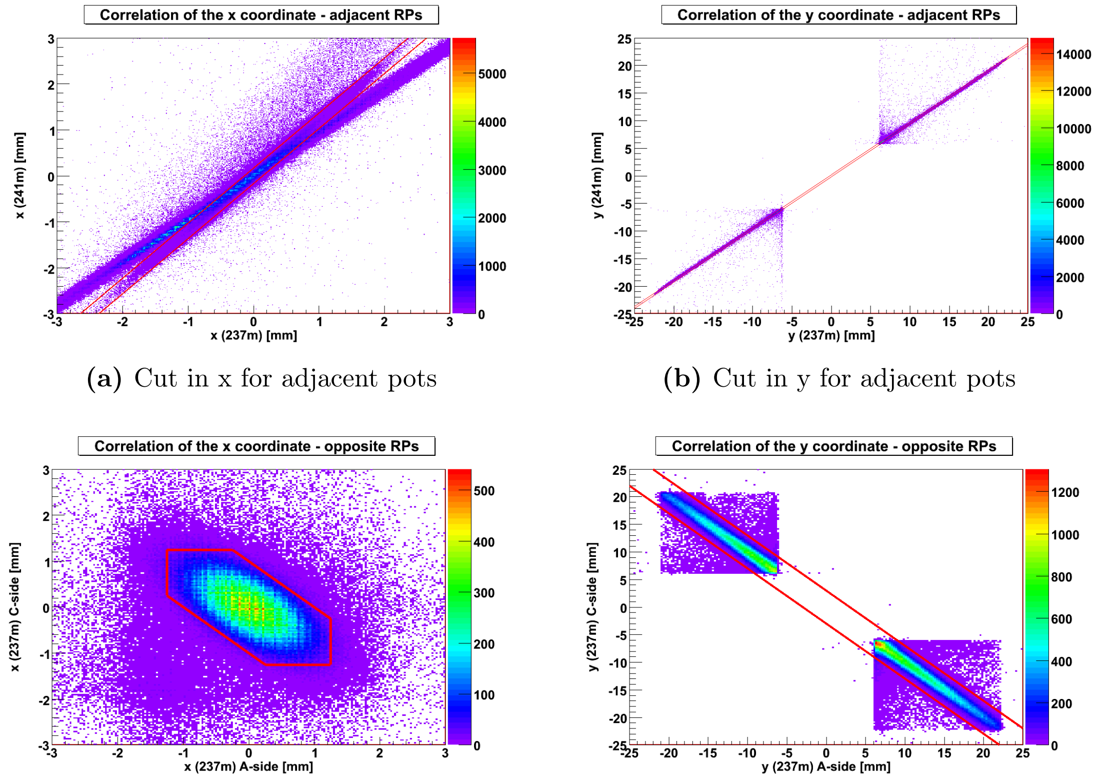
$$u = \tfrac{\sqrt{2}}{2}(-x - y) \qquad\qquad u = \tfrac{\sqrt{2}}{2}(-x + y) \tag{1}$$

after this assumption we get that the fiber positions in u and v are:

$$u = -\tfrac{b\sqrt{2}}{2} \qquad\qquad v = \tfrac{b\sqrt{2}}{2} \tag{2}$$

Next, I made *DrawBigEvent.h* - this version can handle all four detectors from one arm. In this case we can see the entire event (or rather half of it, as it's rare that there is

[2]This cut allows me to simplify a few tasks, for example P. Dereń does not use this cut in his work - [3].
[3]The second band in **Fig. 4a**. is due to beam halo
[4]*ROOT*'s container for storing data



(a) Cut in x for adjacent pots

(b) Cut in y for adjacent pots

(c) Cut in x for opposite pots

(d) Cut in y for opposite pots

Figure 4: Cuts in track positions for elastic events

something special in both arms at the same time), and compare tracks in inner and outer pots, for example looking for results of showers.

For elastic tracks I refined both of these macros. When we have one track, we can get its position in x_t, y_t, convert it to u_t, v_t coordinates with **Eq. 1.**, and use it in **Eq. 3.**, which is the corrected version of the fiber position coordinates, where we do not make any assumption about the slopes:

$$u = \frac{(1+a)v_t - b\sqrt{2}}{1-a} \qquad v = \frac{(1-a)v_t + b\sqrt{2}}{1+a} \tag{3}$$

After this we can visualize elastic events with better precision.

3 Detector geometry

3.1 Comparison between ideal and real geometry

The layers in detectors should be staggered by multiples of 1/10 of the fiber size to minimize the uncertainty of the track positions. In this way, if every fiber has a different position, one fiber width is subdivided into ten corridors, which particles can penetrate. A corridor is defined by the two nearest fibers edges, they set possible paths for proton, which is moving in parallel to the beam (i.e. perpendicularly to the pot). For this staggering, the resolution is equal to 1/10 of the fiber size, this is an ideal geometry for detectors.

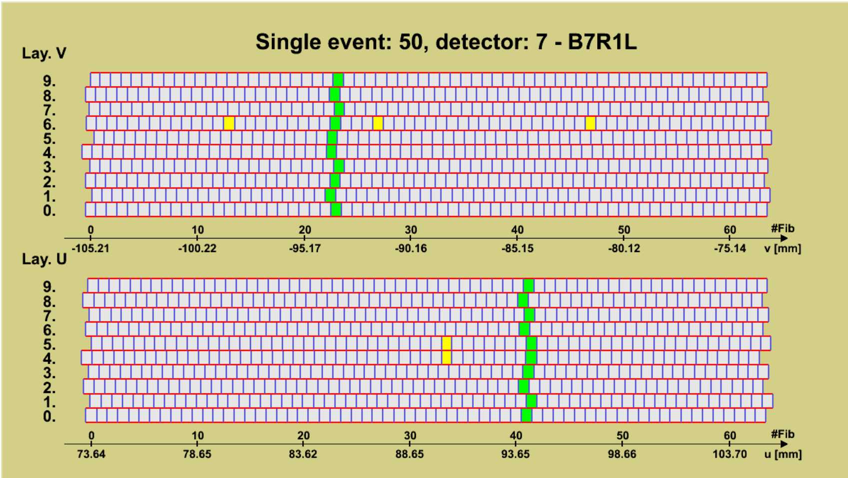
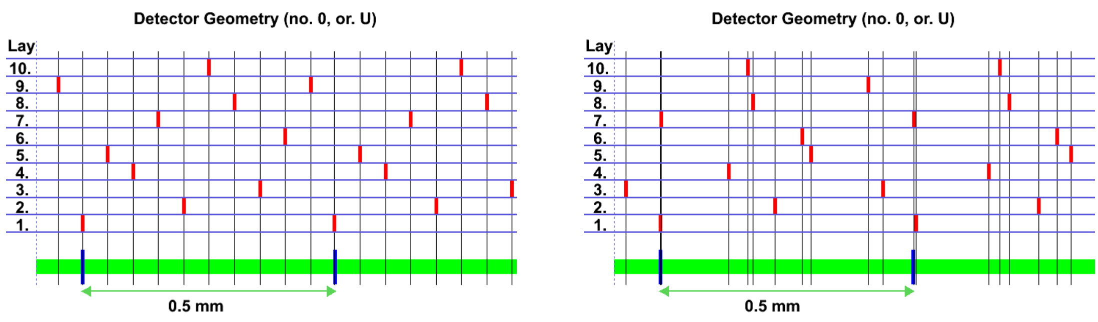


Figure 5: A sample visualization of a very clear event.

A first task was testing the real geometry and estimating the extent to which imperfections in the geometry can influence the results of track reconstruction. On **Fig. 5.**, apart from a clear event, we can see that the staggering is very imperfect (for example layer #4 has only a small shift with respect to layer #5). In **Fig. 6.** we can see the difference between real and ideal geometry. This is a close up of pot A1, near fiber #45 in the u coordinate. In **Fig. 6a.** - the ideal case - the fibers are staggered regularly, all corridors have the same width. In **Fig. 6b.** is very different. Instead of a uniform distribution of fibers, in some layers they lie in the same place. Corridors are distributed very irregularly. When we look into the marked area on the left, the first corridor is the widest of all, followed by five narrow corridors. The seventh corridor is wider too, and the remaining two have an average width. The tenth corridor is practically nonexistent.

The corridors' positions and widths are relevant in track reconstruction. It is more likely that a proton penetrates the wider corridor. On the other hand, the overlapping region (the error in position) is smaller when the corridor is narrow.



(a) Ideal geometry

(b) Real geometry

Figure 6: Ideal and real geometry (for pot A1, near the #45nd fiber, u orientation). The red strips mark the edges of fibers, and the black lines separate corridors. The fiber width of approximately $0.5mm$ is marked below.[5]

[5]One fiber is $480\mu m$ wide, not counting approximately $20\mu m$ of glue.

3.2 Detailed analysis

To look how imperfections of the geometry influence the results, I tested the distribution of reconstructed tracks. The most evident consequences are visible, when we look into the distribution of tracks near fiber #45, which is shown in **Fig. 7.**, and compare it with the geometry from **Fig. 6.**

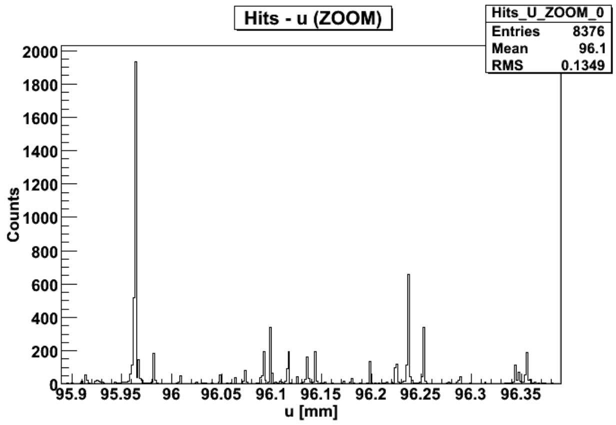


Figure 7: The distribution of tracks near fiber #45

As we could expect, this distribution is highly correlated with the detector geometry. The most significant is the first peak, we can link it with the first corridor from **Fig. 6.**. Next, we can see a small group of a few peaks, which correspond to five narrow corridors. Similarly, the subsequent peaks are correlated with the last corridor in **Fig. 6.**.

We didn't get one peak per one corridor. When one of the fibers in a track is missing, the corridor is wider and our algorithm calculates a different position of the hit than the middle of one of the ten nominal corridors. The content of the peaks are explained in section 4.

Though this section speaks a lot about the detector geometry imperfections, it is in fact not so bad. It is true that the layers are not staggered so well, but all fibers have very similar width. For this reason, the grid in the detector has something akin to discrete translational symmetry, which we can see at **Fig. 5.**. In the next fiber the corridors are distributed in a similar way. We can see it on next histogram in **Fig. 8.** too, where distributions of distances between tracks and the fiber centers are shown (for first layer in pot A1). We have negative distance too, because this takes into account the side. The most important feature of this histogram is the similarity with the distribution in **Fig. 7.** The peaks are smooth, as the discrete translational symmetry is not so ideal, but still we can see big peaks for wide corridors and short but wide peaks linked to the few narrow corridors.

When we know the slope a and intercept b for each fiber we can calculate the distance using the simple formula[6]:

$$d = \frac{ax_t - y_t + b}{\sqrt{a^2 + 1}} \tag{4}$$

The fiber with the smallest calculated absolute distance from the track, is taken to have been hit.

[6] I have skipped calculating the absolute value, to retain information about the direction

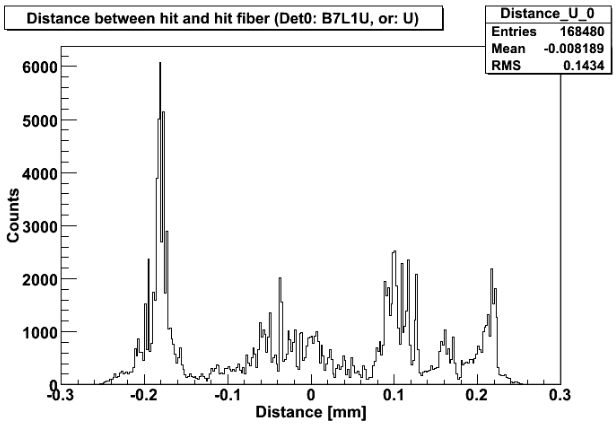


Figure 8: Distance between track and hit fiber in first layer in u orientation in pot A1. The peaks are very clearly visible.

3.3 Final results and conclusions

After this we can explain the distribution of all reconstructed tracks in pot A1. With ideal geometry, we could expect a very regular grid of reconstructed tracks, with resolution equal to 1/10 of the fiber size. For the real geometry, the grid is still regular, but not so clear, the distribution is more discrete. It is important to note that although a more discrete distribution lowers the precision, it does not mean that the results are wrong.

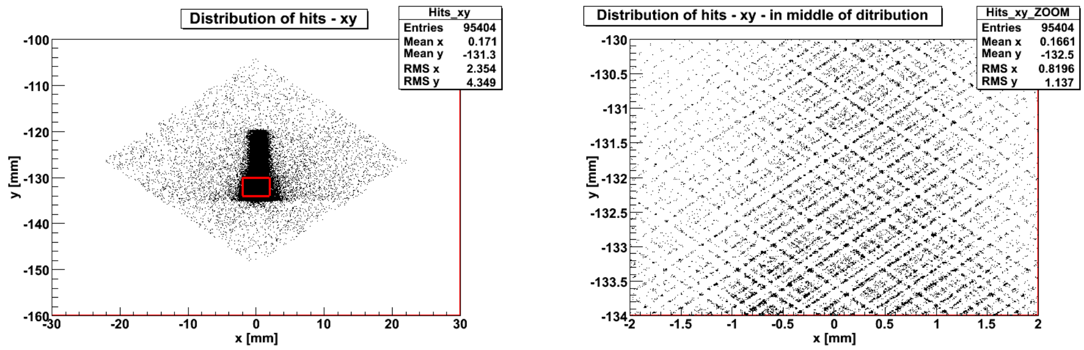


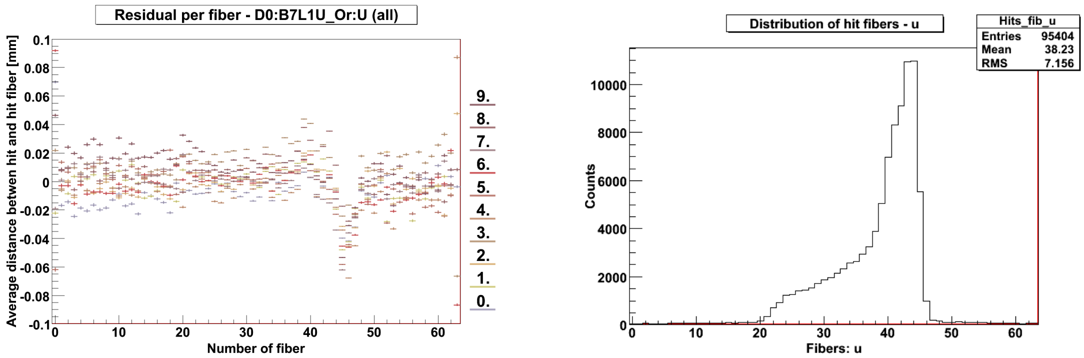
Figure 9: Distribution of reconstructed tracks in pot A1. The histogram on right is a close-up of the middle part of the distribution (histogram on left - close-up area is inside the red rectangle).

The analysis is not very rigorous, but as the pictures show, the precision of determining the track coordinates is determined by the detector geometry and is less than $0.05mm$ (width of one fiber divided by ten layers), because fibers in each layer do not produce exactly 10 corridors, but sometimes lie in the same place. It is possible to introduce more geometry tests and determine the nominal precision for each detector.

Finally, we can add that the v layers are slightly better staggered than the u layers (the discrepancy in corridor width is slightly smaller), and the precision in determining the v coordinate is a bit better than u.

4 Residuals

One of my analyses is focused on testing the residuals. By 'residual', the average distance[7] between track and hit fiber is meant. Histograms with residuals depend on the distribution of proton tracks. For example, residuals should be equally distributed around zero for a uniform distribution (in ideal geometry). Shown below are histograms of residuals per fiber from pot A1, for ten layers in u coordinate - **Fig. 10a.** Residuals from different layers are marked by other colors. The most important feature of this histogram is the characteristic "S" shape near fiber #43. The residuals average is a bit above zero near fiber #40 (which is not so clearly visible), and have a dip near fiber #46 .



(a) Histogram with residuals **(b)** Distribution of hit fibers

Figure 10: Histograms for pot A1 in u coordinate: residuals per fiber for each layer (see the legend with colors for each layer on the right) and distribution of hit fibers. The peaks in the second histogram lie in the same area as the "S" in the first histogram.

As mentioned before the explanation of this characteristic "S" shape lies in the tracks distribution. We can see, that the "S" lies in the same area as the peak in the histogram from **Fig. 10b.** When the density of tracks decreases very sharply, as for fibers #44-47, in the left side of the fiber there are more hits than on the right side[8]. The negative values of residuals for these fibers are the result - see negative peak in **Fig. 10a.** For fibers near #40, we can see positive average residuals, but the effect is not so clear, because the density change is not as sharp on the other side.

To compare, i prepared the same histograms after the cuts- **Fig. 11.** We can see an ideal "S" shape in **Fig. 11a.**, because the distribution is sharper than before.

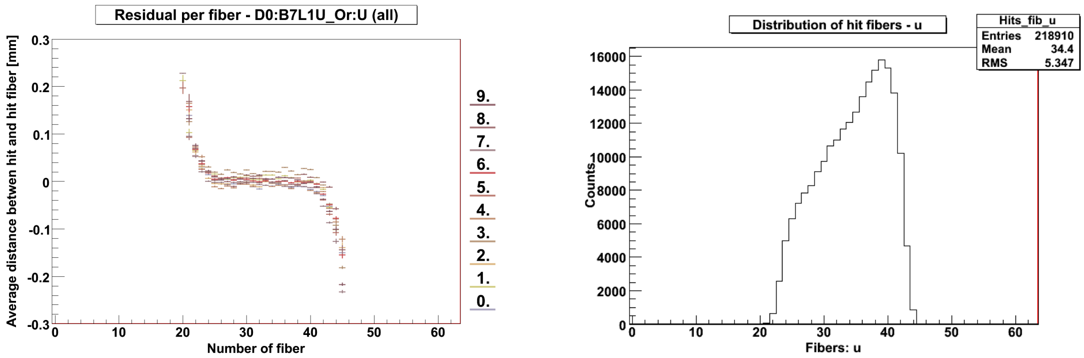
5 Efficiency of reconstruction

In this paragraph I show my results of testing the layers' efficiency. I did not deal with real efficiency, as that requires very complicated analysis. The average probability of fibers producing light when a particle passes through is about 95% - [1].

I dealt with something, which could be rather called 'efficiency of reconstruction for each layer'. It was calculated in the following way for each layer: for each reconstructed track, I counted the number of times when the algorithm did not choose any fiber in this layer, and after this, I scaled received numbers by number of all reconstructed tracks in the

[7] Calculated from **Eq. 4.**

[8] As in **Fig. 7.**, where the distribution of tracks near fiber #45 is shown.



(a) Histogram with residuals **(b)** Distribution of hit fiber

Figure 11: The same histograms as in **Fig. 10.** but for events after cuts. We can see an ideal "S" shape.

pot. This is not real efficiency, but it can provide some information. For example, it can tell us that layer #19 in pot A8 was broken - its efficiency is smaller than expected - see **Fig. 12.** This fact has been confirmed - this layer was damaged during the installation of pot A8.

One can also determine the orientation in the z-direction of each pot. The orientation of the lower pots in z is opposite to the orientation of the upper pots. The layer with the lowest efficiency in the pots (expect for the broken layer) is the first layer hit by elastically scattered proton.

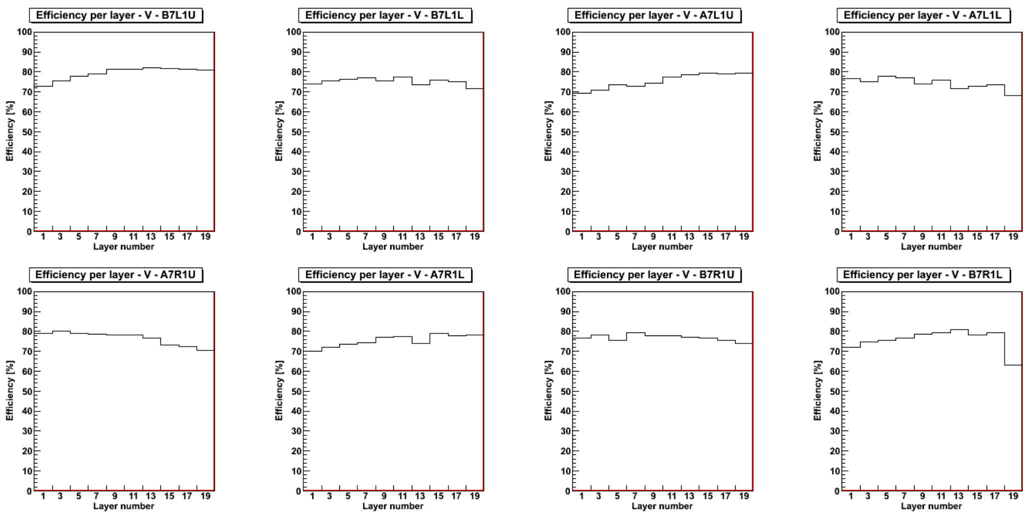


Figure 12: Efficiency of the reconstruction for each layer in u coordinate

These tests were done for all events. I can repeat them for elastic events only, in this way we can get more information about efficiency. I dealt with efficiency for each fiber separately, and we can find for example "dead fibers" - fibers with very low efficiency, and other interesting things.

6 Calculating the track position

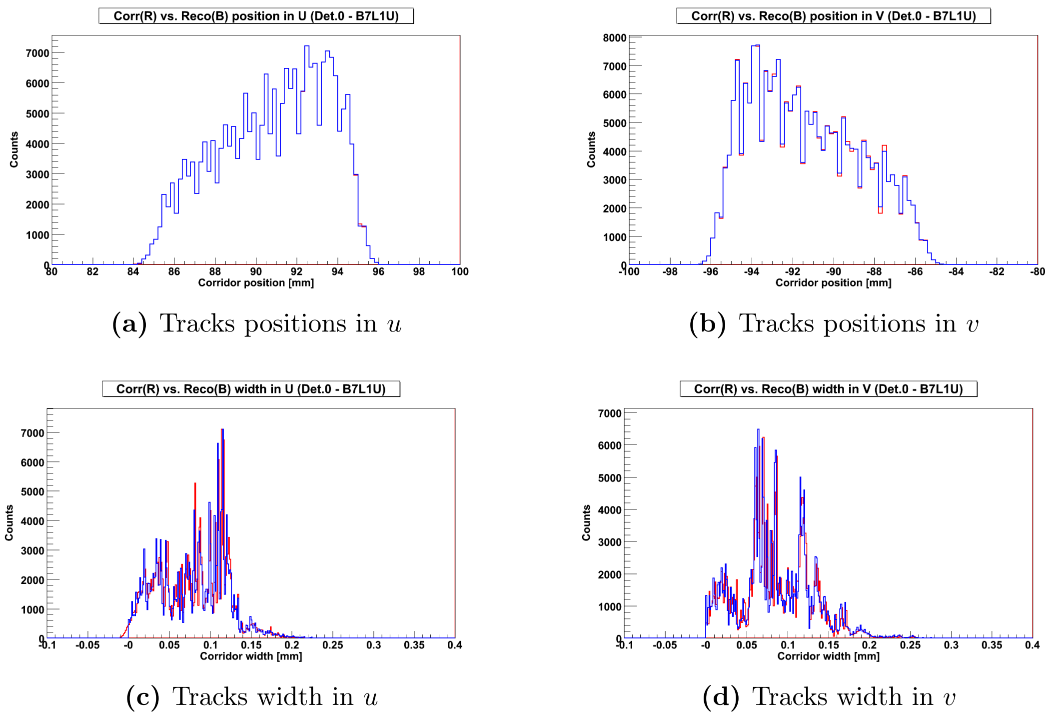
6.1 Simplified method of calculating the track position

The original method of calculating the track position was mentioned in the introduction - 1.2 - the entire range was divided into $1\mu m$ bins, and for each bin was the amount of fibers hit was counted. From the received histogram, the positions and the overlapping width were calculated.

The new method focuses on the corridors penetrated by protons, separately for both coordinates. At first, the corridor edges are calculated - the left edge is obtained by the maximum of all left edges of fibers selected for reconstruction, similarly for the right edge (see **Fig. 3.**). After this, we get the narrowest possible corridor, where the proton could penetrate the pot. The position of the track is assumed to be in the middle of a corridor, and its width is assumed to be the width of the overlapping region.

6.2 Comparison of both methods

In **Fig. 13.** we can see a comparison of both methods. Histograms **13a** and **13b** show positions of reconstructed tracks, separately for each coordinate. Histograms **13c** and **13d** show a comparison between the widths of the overlapping region. The blue line shows results for the old method, the red line - the new.



(a) Tracks positions in u

(b) Tracks positions in v

(c) Tracks width in u

(d) Tracks width in v

Figure 13: Comparison between new (blue) and old (red) method to calculate reconstructed track parameters.

All resulting histograms are very similar (it is especially difficult to see the differences in positions). This confirms that the new method is consistent with the old one. Slight differences in positions are due to rounding. Histograms for overlapping regions have more differences, the width is typically smaller for the new method than for the old one. We can accept this fact as an indication of better precision of the new method.

Both methods have small problems. In the old one, we get a significant peak for the width of $480\mu m$ for non-elastic events. In the new method, we sometimes get a negative value for the width of the overlapping region, but it is always less than $10\mu m$. The reason again lies in the non-ideal geometry. In calculations, a fiber width of $480\mu m$ was assumed, but sometimes it can vary slightly. We can assume that the negative values overlap of less than $10\mu m$ are insignificant - [2].

The analysis was conducted with only clear events (see: 1.3), and positions of fibers were calculated with the more accurate equations: **Eq. 3.**

7 Future tasks

One of the main future goals is designing an algorithm capable of identifying showers and getting rid of them. In his work - [3], P. Dereń has shown evidence suggesting that even for supposedly elastic events, there are still proton-detector interactions, which produce showers. One of the possible methods could be analyzing the variations of lit fibers in the neighborhood of a track, however getting rid of all the showers requires far subtler methods. This task is important for the further development of the algorithm. In **Fig. 14.** an example shower is shown.

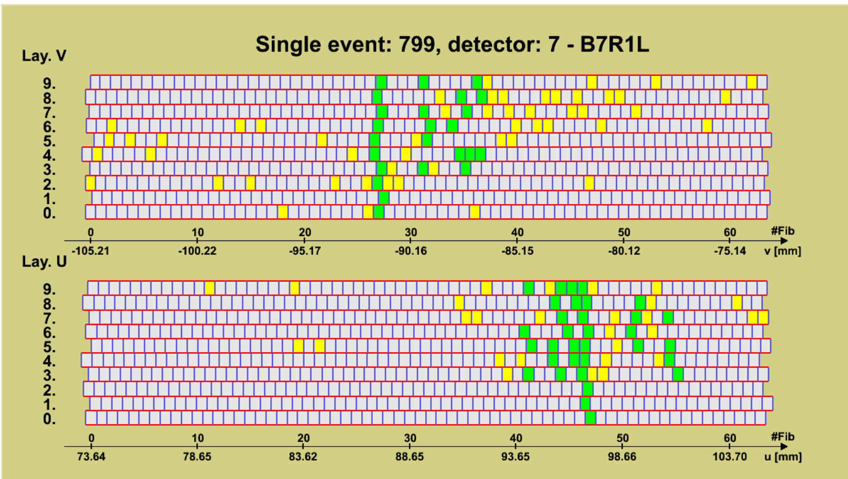


Figure 14: Sample visualization of a shower

Having developed methods for seeking showers, it would be pertinent to focus on improving the track reconstruction algorithm itself. One could consider writing the basis of a new algorithm, whose results would be comparable to the current one. One could use the fact that fibers produce light with a high percentage when hit by particles. Missing consecutive fiber hits or unusual activity near an identified track would indicate fake tracks or the start of showers. We would expect the new algorithm to be more optimized and more sensitive with respect to effects such as showers or background noise, which lower the quality of the data.

8 Summary

In the above paper I have presented the results of my work at the DESY Summerstudent Program. My project consisted of several separate tasks pertaining to the analysis of the track reconstruction algorithm. In section 3 I have quantified the extent to which the non-ideal detector geometry affects the results. I also focused on possible ways of improving the algorithm (sections 4 and 6), and in section 7 I presented a few ideas which could potentially be used in constructing a new algorithm. A useful tool is a new event visualization method, detailed in section 2.

I hope that my work here will prove useful. There also is a possibility that I will continue my work in the ALFA group in Cracow, basing my thesis on it. This would allow me to apply the knowledge and skills I gained at DESY, particularly with respect to using the ROOT framework.

Acknowledgments

My work would not be possible without the help of my supervisors: Wolfgang Friebel and Karl-Heinz Hiller. I hope to be able to work with them in the future. I would like to thank them for their assistance during the Summerstudent Program, and for everything they have taught me. Thanks to them I have familiarized myself with the ALFA experiment and with the geometrical reconstruction of tracks that is part of the data analysis.

I would also like to thank the administrative staff of the Summerstudent Program, for providing this wonderful opportunity of working and living in Zeuthen, where I was able to develop my skills and interests, as well as meet wonderful people.

Finally I would like to thank Piotr Dereń for offering to correct this text for me.

References

[1] Measurement of the total cross section in pp collisions at $\sqrt{s} = 7$ TeV from elastic scattering with the ATLAS detector, The ATLAS Collaboration, CERN, 1211 Geneva 23, Switzerland (July 24, 2014);

[2] W. Friebel and K.H. Hiller, private communication.

[3] Quality assurance of the Track Reconstruction Algorithm in the ALFA detector of the ATLAS experiment, P. Dereń, DESY, Zeuthen, Germany (September 5,2014)