Summer Student Report: Clustering-based Studies on the upgraded ITS of the Alice Experiment

Bofang Jiang
jiangbofangfang@gmail.com
CERN, European Organization for Nuclear Reasearch, Geneva, Switzerland

Abstract
ALICE (A Large Ion Collider Experiment) is a general-purpose, heavy-ion detector at the CERN LHC which focuses on QCD, the strong-interaction sector of the Standard Model.

The ALICE Collaboration is currently upgrading the entire detector in view of the third run of LHC (2021-2023). One of the key feature is the construction of a low-material budget Inner Tracking System (ITS) based on Monolithic Active Pixel Sensors (MAPS). Each sensor features about 520000 pixels with an adjustable discriminating threshold. Typically the threshold is uniform among the pixels of the sensor. Sensors which shows small areas with higher threshold in our experiment, are named "High Threshold Clusters". In this work, we study the possible effects produced by the high-threshold clusters on the track reconstruction. We first develop a clustering algorithm to find such High Threshold clusters on the chips, hence we excavate some observables of the clusters to do multivariate analysis. Take advantage of the simulation toolbox O2, we introduced the experimental data into simulation and thus certifying the correlation between the cluster parameters and the track reconstruction performance.

Keywords
clustering algorithm; Alice O2; silicon tracker, detector classification.

1 High threshold cluster problem in Alice ITS silicon chips

Implementing Monolithic Active Pixel Sensors (MAPS) in upgraded inner tracking system(ITS), Alice will achieve a higher impact parameter resolution. Updraged ITS is longitudinally segmented in Staves featuring 9 sensors in the innermost layers (Inner Barrel). Utilizing seven layers of silicon pixel detectors with 13.5 Giga-pixels covering 10 square meters, Alice will make the biggest amount of monolithic active silicon trackers around the world. The silicon chips have a simple data acquisition principle based on the charge collected within a time window and will be fired once the charge is above a threshold in the active volume of the sensor (epitaxial layer).A digital signal is given in the output is the collected amount of charge is above a given threshold (adjustable for each pixel). Typically the threshold is expected to be uniform with some fluctuation within few electrons, however, in the threshold measurement, some high threshold areas are discovered on a fraction of ITS chips. As shown in Fig. 1, the stave J006 (one of the produced staves) chip0 shows some areas with much higher thresholds than the other pixels in the matrix, especially the yellow region. The adjoined high threshold pixels forms a so called " high threshold cluster". A possible reason of such a phenomenon is the leakage current during the chip operation. Obviously, if some charges are fetched by pixels inside a cluster, it will lead to a lower spatial resolution and charge collection efficiency. In order to improve the charge collection efficiency, we can apply a voltage bias to the sensor but this shortage is it will aggravate the ununiform threshold. In this work, we attempt to find and analyze the high threshold clusters, analysing how they affect the track reconstruction in the silicon tracker.

2 A new Clustering algorithm

We first develop an algorithm to locate the clusters. As is shown above, we pick out the abnormal points set P, define a constant "cluster distance", which refers to the dispersion ratio in a cluster, then we compare the distance of each two points in P. The pixel inter-distances inside the chip are evaluated, and those which fall

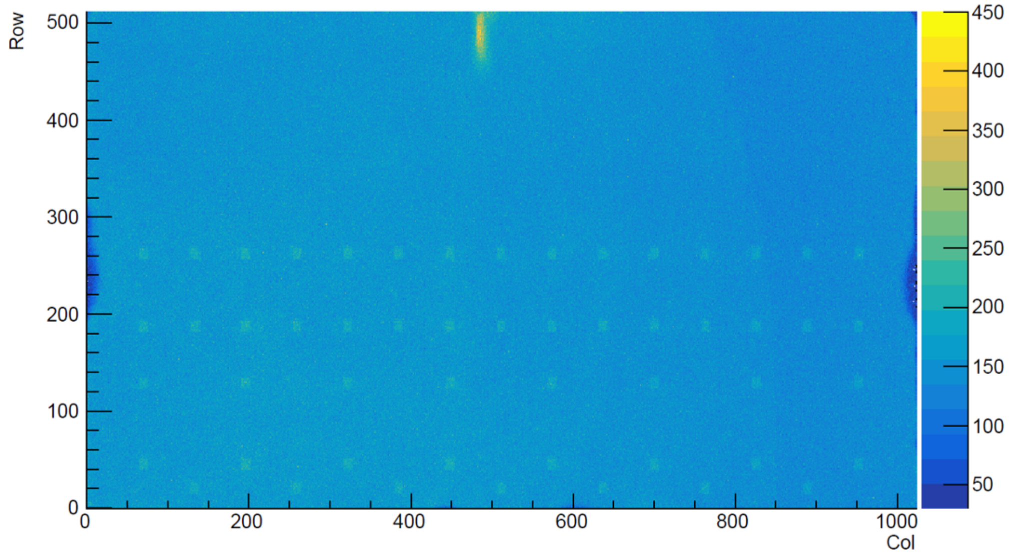

Figure 1: Threshold map of Alice ITS IBSTAVE J006 chip0, with yellow part of high threshold cluster,horizontal and vertical axes are the long and short edge of a rectangle chip and the color scale refers to the threshold in electrons

within a certain distance will be assigned to the same cluster. For each point in a certain cluster, it has at least one neighbor, whose distance is shorter than the cluster distance. After a cluster is created, we scan the point set again and again to add points . the cluster definition will only end when no more unassociated points are found to satisfy the inter-distance relation. Hence a new cluster is begun. Finally we distinguish all the clusters and collect the output cluster parameters such as number of clusters, Area of each cluster, average threshold, shape of each cluster, position of cluster centers, threshold distribution in clusters. It is noticeable that the isolated high threshold pixels due to statistical fluctuation will be abandoned since we set a lower bond of the cluster area. The reconstructed clusters (starting from the map shown in Fig. 1) are shown in Fig. 2. The parameters can be used to do the classification and to study the effect of high threshold clusters on track reconstruction.

Clustering algorithm

```
Abnormal points set P={p1, p2, p3 ...}
Create a first cluster C1
Adding an arbitrary point p1 to C1
while (point pi in P & p not in C1)
    if (pi has a neighbor in C1)
        Add pi to C1
        Remove pi from P
    end
    if (no new point added after one loop)
        break
    end
end
Create a cluster C2 ...
```

3 Modified O2 simulation

We present here the results obtained using a realistic detector response by inserting the experimental threshold distribution in a simulation developed in the ALICE Online and Offline project (O2). We attempt to prove how the high threshold problem affect the track reconstruction. The workflow of the simulation is displayed as follows. The first step of Monte Carlo simulation is to simulate the energy deposition of particles interacting in the active volume of the ITS. Then we add the detector response to obtain digitized hits on readout layers.

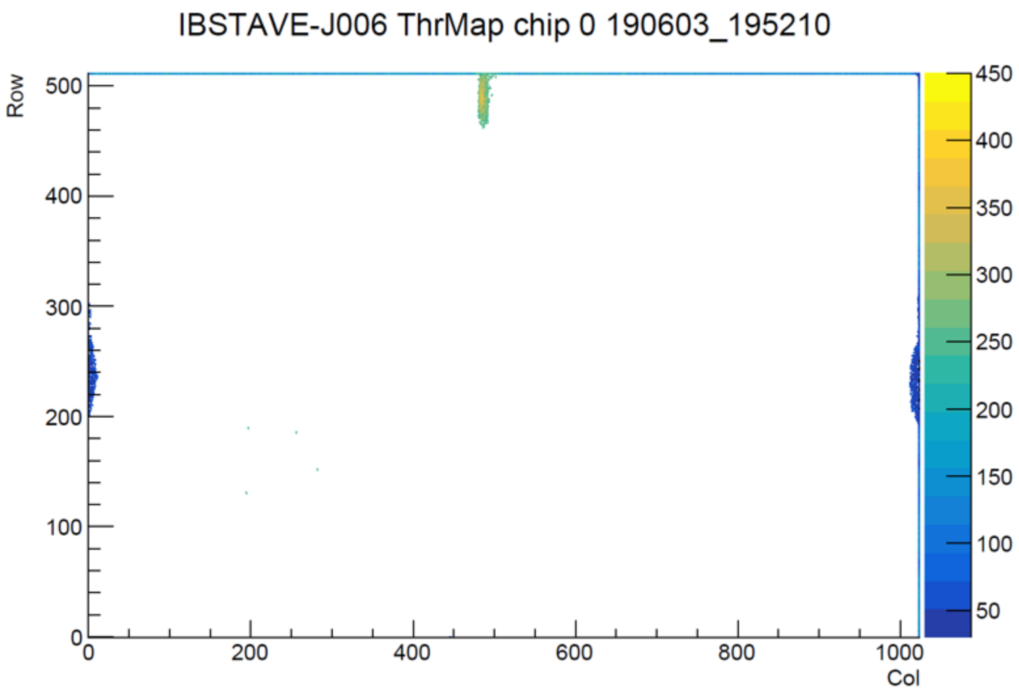

Figure 2: High threshold and low threshold clusters on J006 chip0

Afterwards we do space reconstruction by treating the center of triggered pixels as the charge center. Hence we can calculate the spatial resolution by comparing the distance between the real charge center and reconstructed cluster center. Previously, the threshold is regarded as constant,the real discriminating thresholds of each pixel are inserted in the simulation.

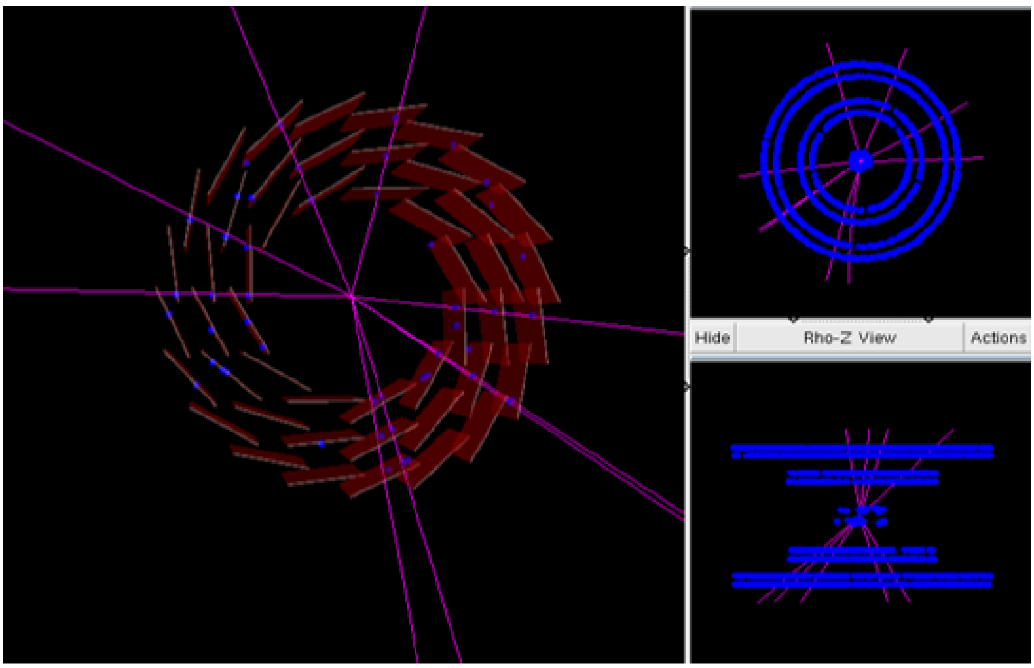

Figure 3: Geometry and trajectory in ITS Monte Carlo simulation

First we generate particles as the MC true events, the geometry of the ITS and initial tracks in a event are shown above in Fig. 3. In each run, we generate 10,000 events randomly, the random seed is defined internally in O2. In this process, detailed tracks of the ionizing particles in the ITS active volume are generated.

As a particle travels through the readout layer, it deposits energy along the trajectory, leaving some charge (electrons) inside the silicon sensors. . In each step, information such as timestamp, particle type, momentum, energy deposition, position, and physical process involved in the interaction are registered.

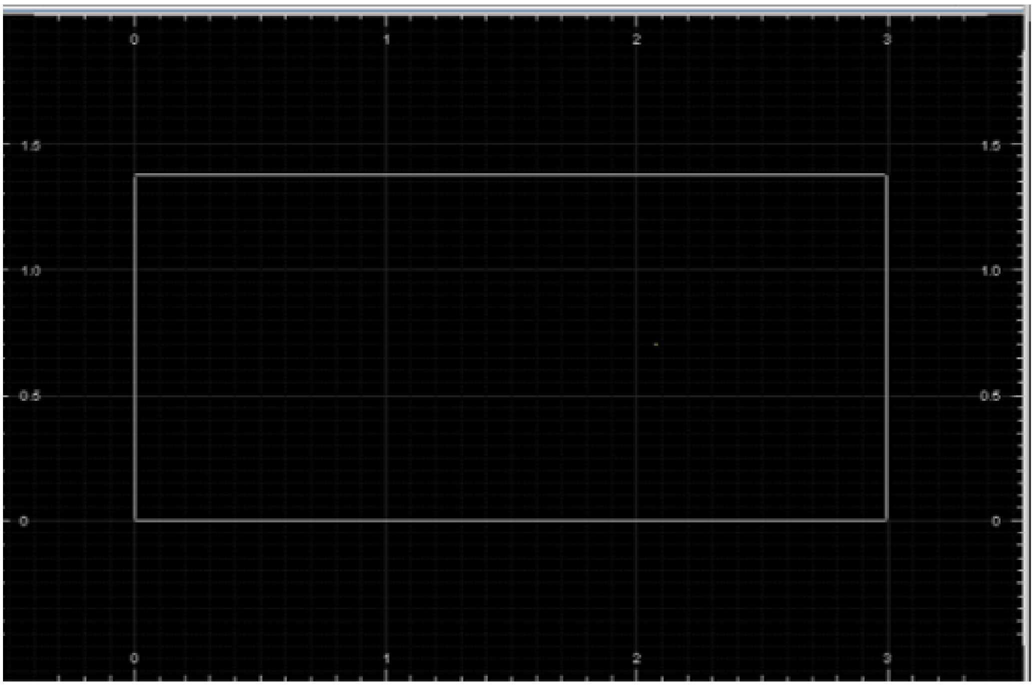

Figure 4: Chip view in O2 with two pixels fired

For the digitizing process, time and energy information will be utilized to judge if a pixel will be triggered. As displayed in a chip view, a set of pixels are triggered in Fig 4. If the charges collected in a time window is above the local threshold, only spatial information of the pixels will be recorded. In this work, we replace the uniform threshold by real thresholds (including the high-threshold clusters). Every time data of only one chip will be inserted in every imaginary chip in our simulation.

Afterwords it comes to the space reconstruction process. As is shown in Fig. 5, the two fired pixels are displayed. Such adjoined triggered pixels is named as a cluster. The reconstructed charge center is considered as the center of the cluster.

Combining the result of MC simulation and track reconstruction, we display the distribution of the position of charge centers from the two results in Fig. 6 in which x is the short edge of a chip and z is both the long edge and the beam direction. We thus define the space resolution as $dx^2 + dz^2$, which refers to the accuracy of hit space reconstruction and serves as a most important parameter to measure the performance of the detectors.

4 Detector classification based on two approaches

Here we come to a first result of the two ranking as is shown in Fig. 7. We analyze some chips in Alice ITS, both with and without high threshold clusters, select some chips whose average threshold are similar and compare the relation between the cluster areas and the spatial resolution. In this case we fail to see a clear correlation between the two parameter since the event number is 10 for each run . Another reason is that the cluster area is limited since only less than 1% pixels are problematic. However, when make the detector response closer to reality, we can prove the resolution is worse than using the previous constant threshold.

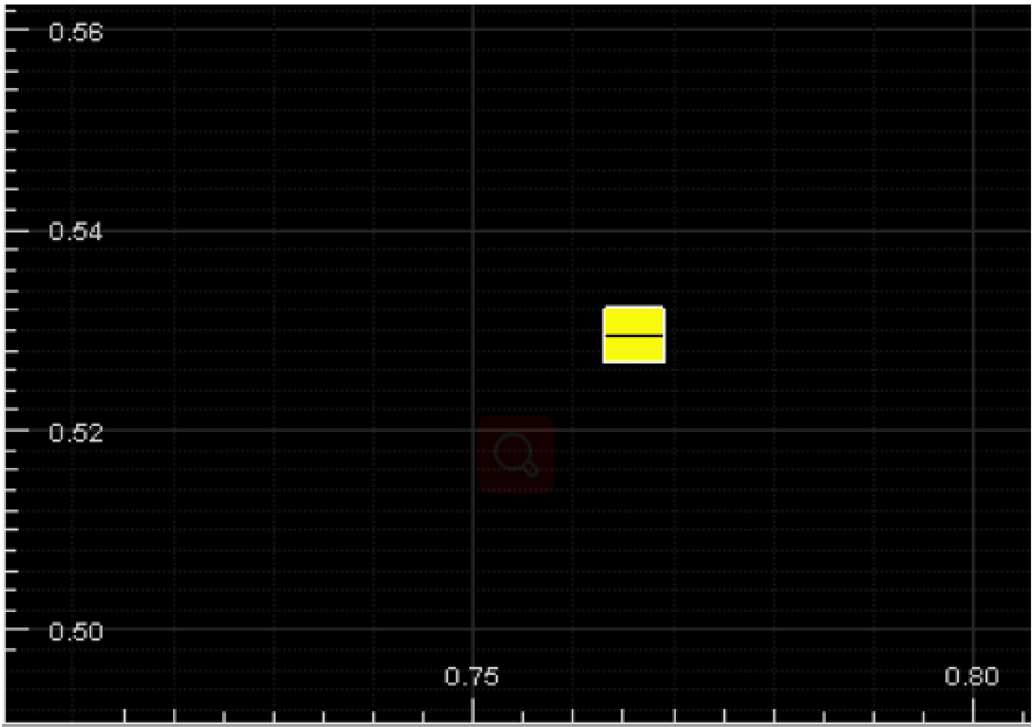

Figure 5: Detailed Chip view in O2 with two pixels fired

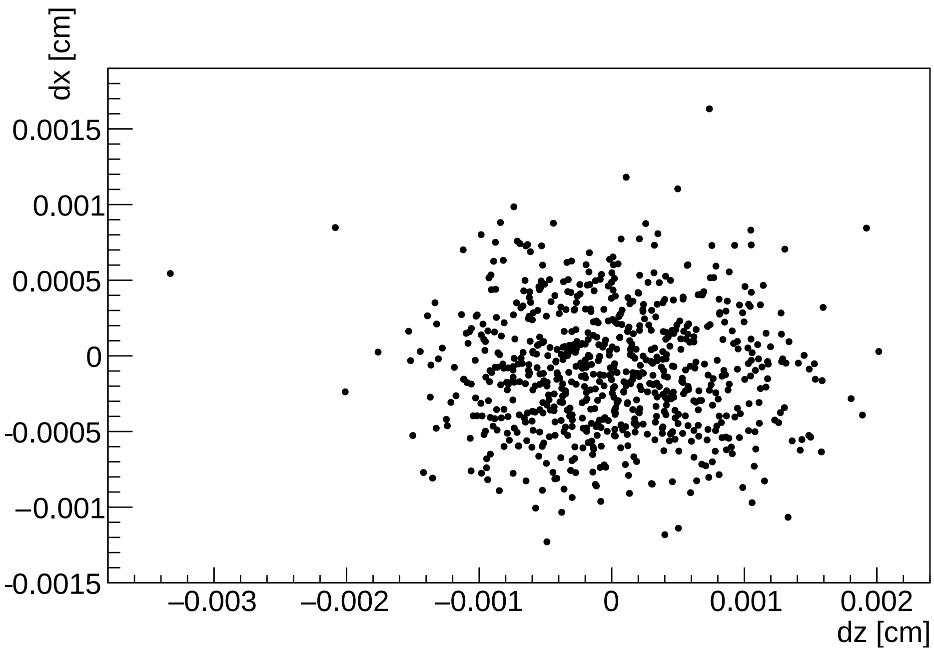

Figure 6: scatter diagram of the distance between MC charge center and reconstructed charge center, dx and dz refer to the distance of x and z (beam) directions

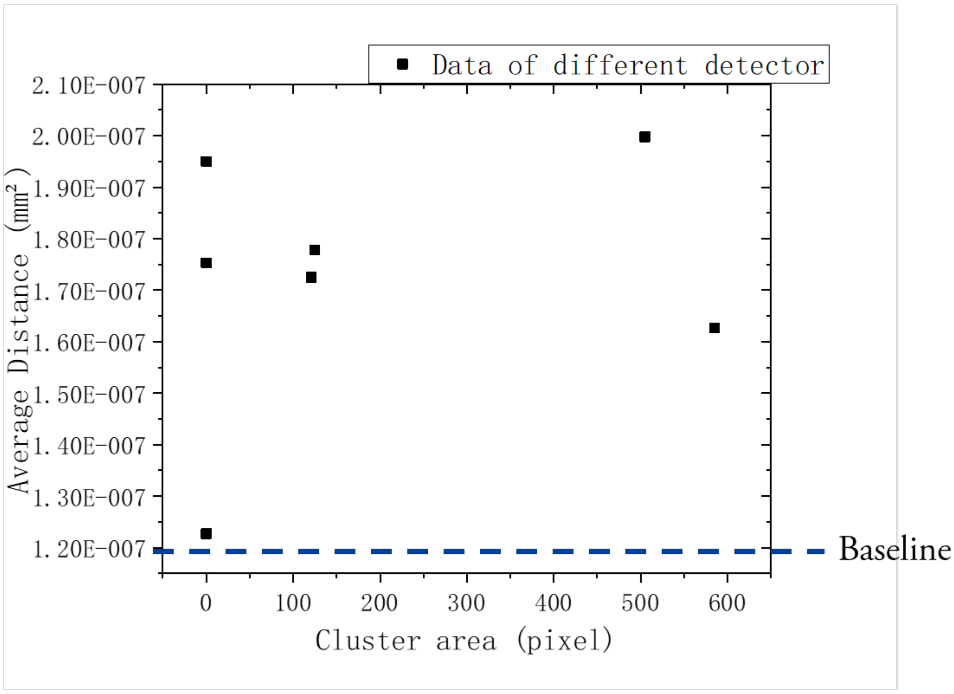

Figure 7: distributions of cluster area and space resolution of some different chips, data are select from chips with similar average thresholds. The baseline is come from the previous simulation which set the threshold as a constant.

5 Quantitative study on the effect of high threshold problem

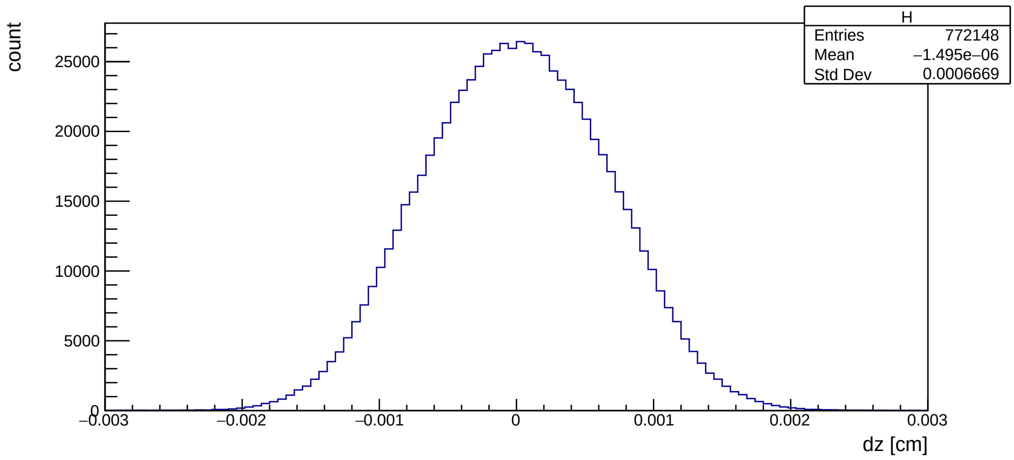

Figure 8: dz distribution of 10 times cluster area and 4 times average threshold in cluster .

In this section, we focus on chip 0 of stave J006 and extend the high threshold cluster area and thresholds inside the clusters in order to systematically study the effect of these parameters on the spatial resolution. We first use the algorithm described in section 2 to find cluster parameters, fit the high threshold cluster in the top center of Fig. 2 as a $30 * 65$ pixels rectangle and run the simulation. Then it continuously copy this rectangle to the left nearby region and run the simulation over and over to see if the resolution will decrease as the cluster area increases. The results are shows below in Fig. 9 and Fig. 10. There appears some negative correlation between the cluster area and hit collection efficiency, but the linear relationship is not clear since the R^2 is far from 1. The error bar is determined by the dx and dz distribution, for instance, the dz distribution in Fig. 8. In a similar way, we show the hits collection efficiency is decreasing when cluster area increase in Fig. 10 since only a very large amount of charges are able to trigger the pixels in a cluster.

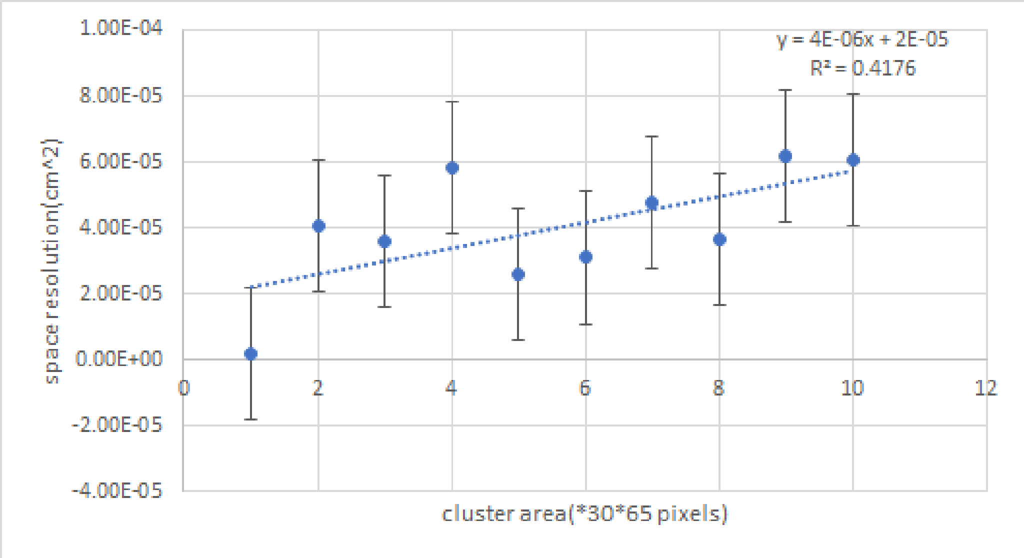

Figure 9: Area versus resolution and the linear fit, the error bar is calculated from the dx and dz distribution

In a similar way, we show the hits collection efficiency is decreasing when cluster area increase in Fig. 10 since only a very large amount of charges are able to trigger the pixels in a cluster.

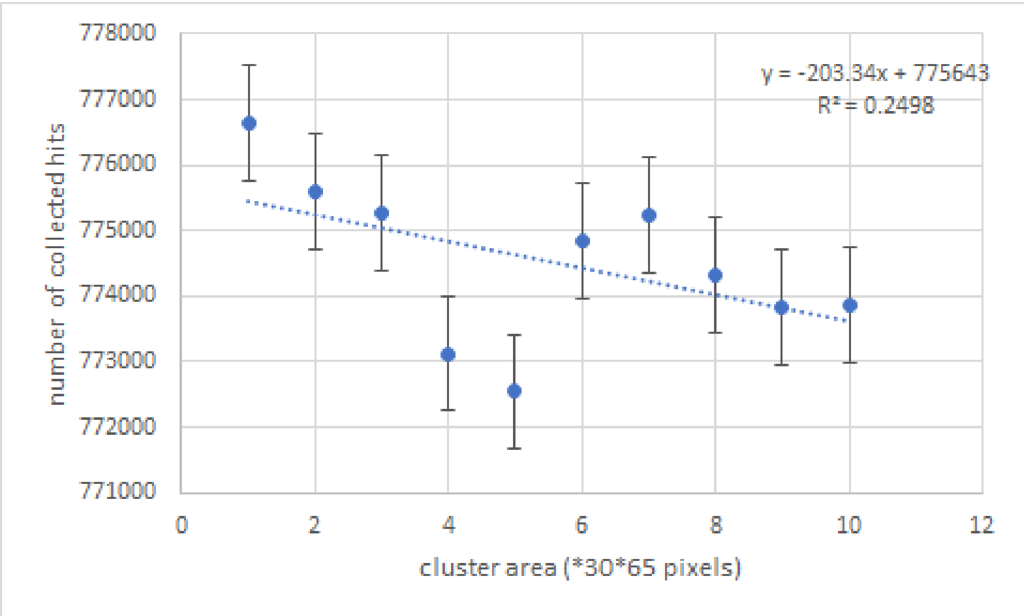

Figure 10: Area versus hits collection and the linear fit, the error bar is calculated from the Poisson distribution

Afterwards, we multiply the threshold in clusters by a multiply factor ranging from 1 to 4, and try to find the correlation between such a factor and detector performance. The result in Fig. 11 shows a similar negative relationship between the factor and the resolution as discussed. In Fig. 12 however, the relationship is not clear enough.

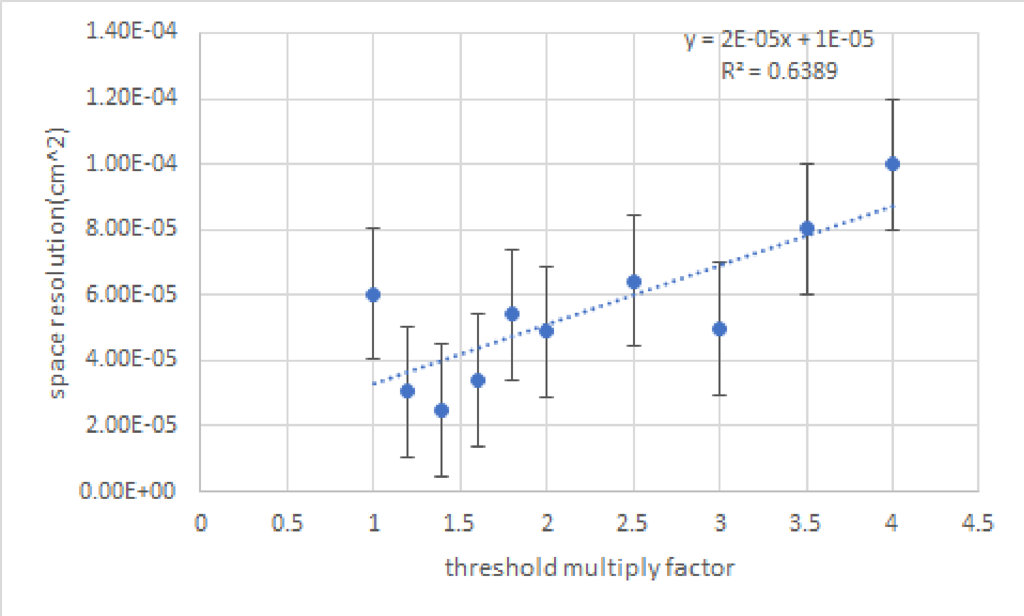

Figure 11: Average threshold in cluster versus resolution, and the linear fit, the error bar is calculated from the dx and dz distribution

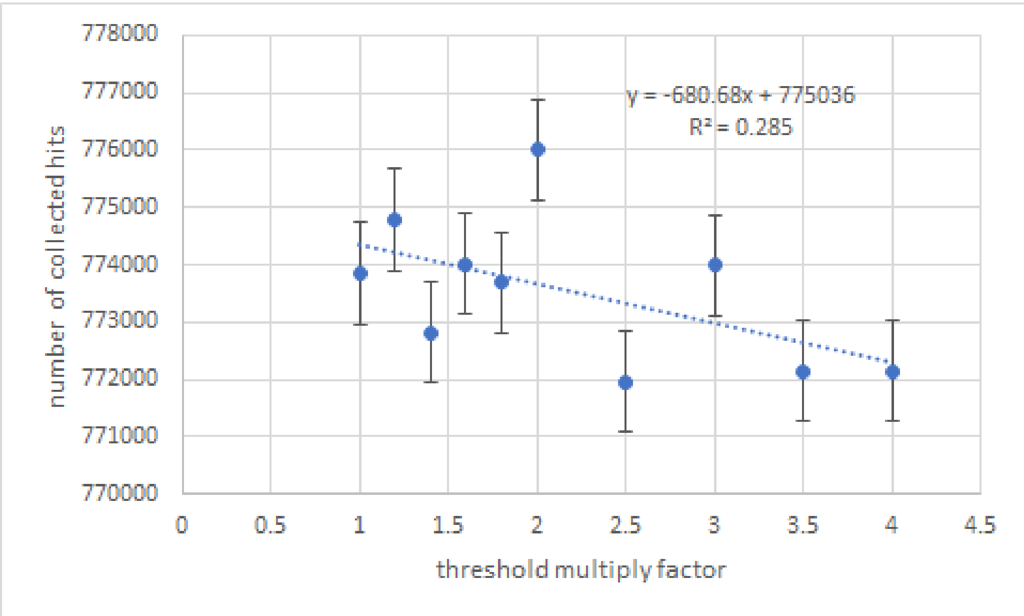

Figure 12: Average threshold in cluster versus number of hits, there is no clear relationship between the two parameters

One reason could be most charge tend to be small, with only a few amount of them beyond 200 in Fig. 13. Under the circumstances, a large threshold multiply factor may not affect a lot since it's dominated by statistical fluctuation.

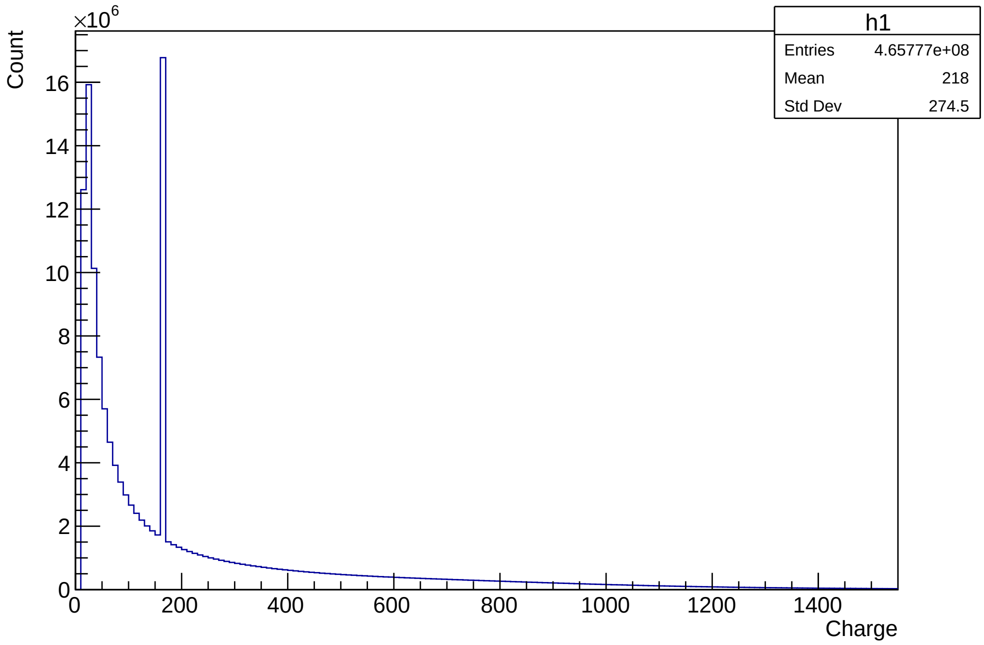

Figure 13: charge distribution in simulation

6 conclusion

In this work, an algorithm to find and analyze high threshold clusters on silicon detectors is developed. Besides, the detector response simulation is closer to the reality by inserting the experimental data in O2. Current high threshold clusters may not have a obvious negative correlation with spatial resolution. When modifying the high threshold cluster area and average threshold in clusters based on the experimental data, we can see both the space resolution and the charge collection efficiency decreasing with the cluster area increasing, and only the space resolution decreasing while the threshold in cluster increasing. However, these relationships are accompanied by a large fluctuation with the same magnitude, especially when the area and multiply factor is close to the experiment. A conclusion is that both the cluster-area and threshold do not affect the detector performance in the track-reconstruction process. Effects are measurable with much larger cluster than real one. A future work could be searching for the balance between "high threshold problem" and voltage bias to optimize the performance the detector performance and to improve the precision of the results by increasing the statistics in terms of number of events and chips analyzed in the O2 simulation. .

Acknowledgements

I wish to thank Antonello Di Mauro and Ivan Ravasenga for their warm help during the whole period of my Summer Student Program. I very much appreciate their concerning about my life, study and scientific research in CERN.

Bibliography

[1] Ravasenga. Ivan, Charged-hadrons production in pp and A–A collisions and characterization of the components for the upgrade of the Inner Tracking System of the ALICE experiment at the LHC (2019).