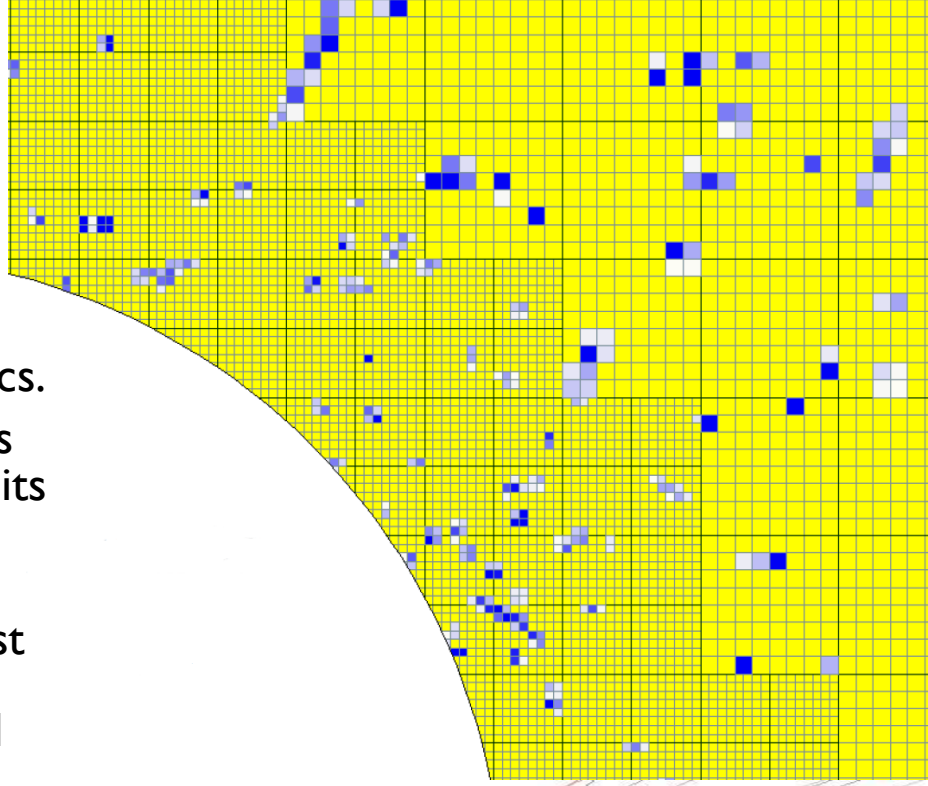
Development of clustering algorithms for the CBM experiment at FAIR

Grigory Kozlov

LIT JINR

Dubna, 2012

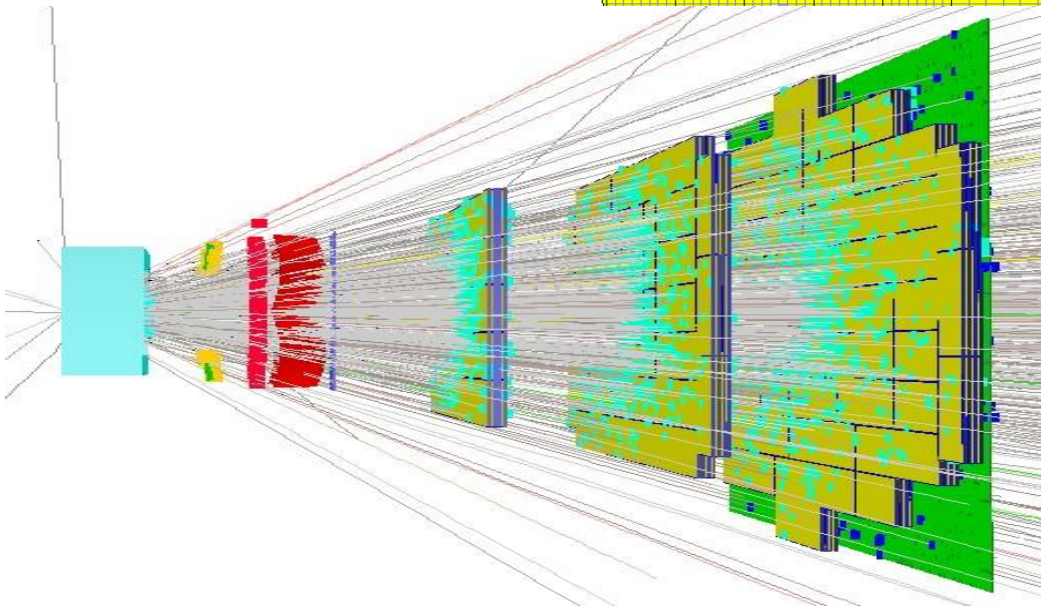
Motivation



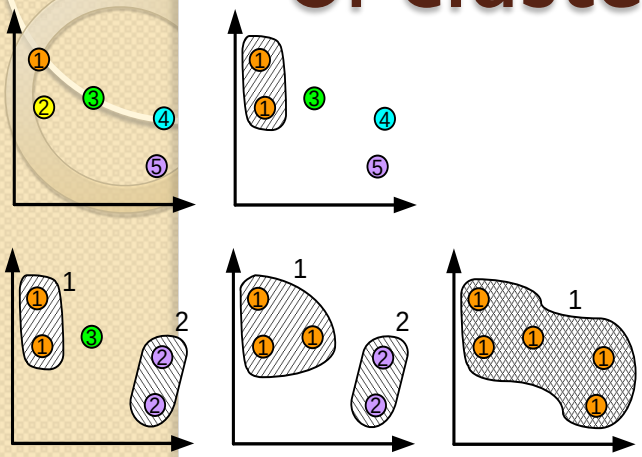
- Clustering algorithms are an important part of the event reconstruction in high energy physics.
- This algorithms translate fired strips and pads into space coordinates - hits which are later used in the tracking algorithms.
- Clustering algorithms have to be fast and effective and has to be able to deal with high track multiplicity and density.

This presentation discusses the following CBM detectors:

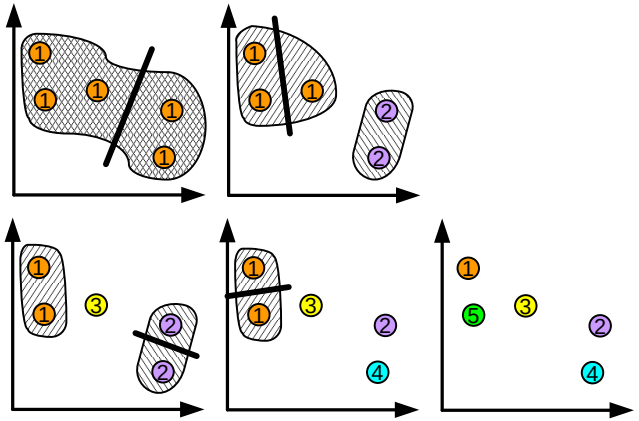
- MuCh – Muon Chamber;
- MVD – Micro Vertex Detector.



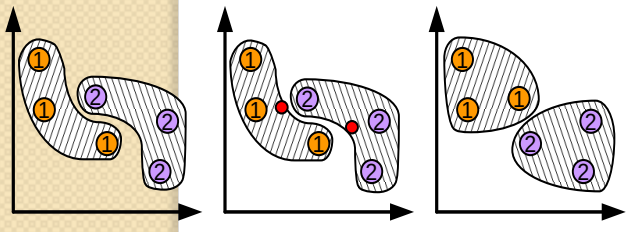
Overview of clustering methods



Hierarchical agglomerative methods combine the original elements and decrease the number of clusters.



Hierarchical Divisive methods divide original data set on different parts that represent a single cluster.



Iterative methods transform the original data without changing the number of clusters.

Ward's method

Ward's method combines objects with a minimum distance between their centers of mass.

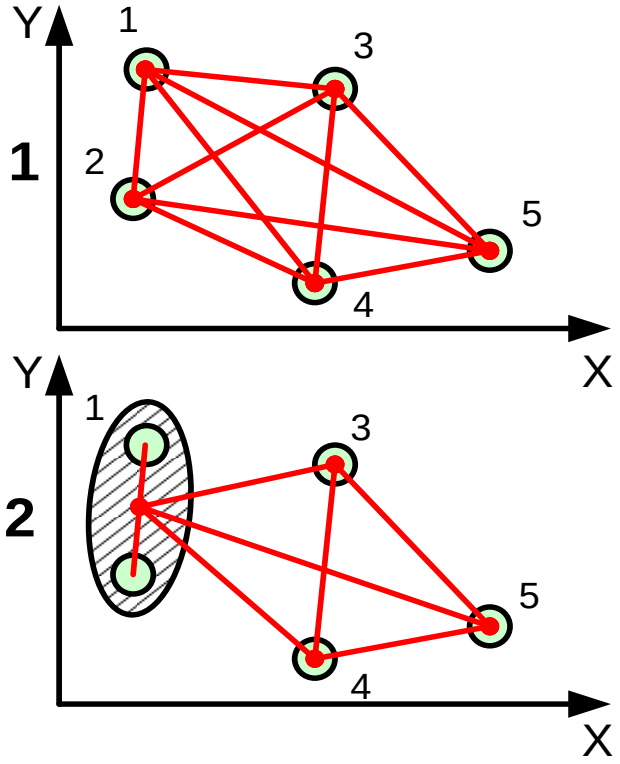
1

	1	2	3	4	5
1	-	1.15	1.33	1.86	2.88
2		-	1.71	1.67	2.47
3			-	1.63	2.11
4				-	1.25

	1	2	3	4	5
1	+	1.15	~~1.33~~	~~1.86~~	~~2.88~~
2		+	~~1.71~~	~~1.67~~	~~2.47~~
3			-	1.63	2.11
4				-	1.25

2

	1	2	3	4	5
1	-	-	1.37	1.58	2.41
2		-	-	-	-
3			-	1.63	2.11
4				-	1.25

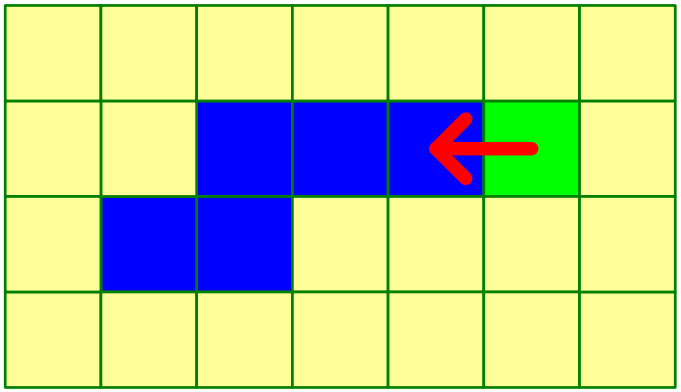


$$d_{RS} = \frac{a_R \bullet a_S}{a_R + a_S} \bullet \left\| \bar{X}_R - \bar{X}_S \right\|^2$$

$$\bar{X}_R^{(new)} = \frac{1}{a_R + a_S} \bullet (a_R \bullet \bar{X}_R + a_S \bullet \bar{X}_S)$$

Single linkage method

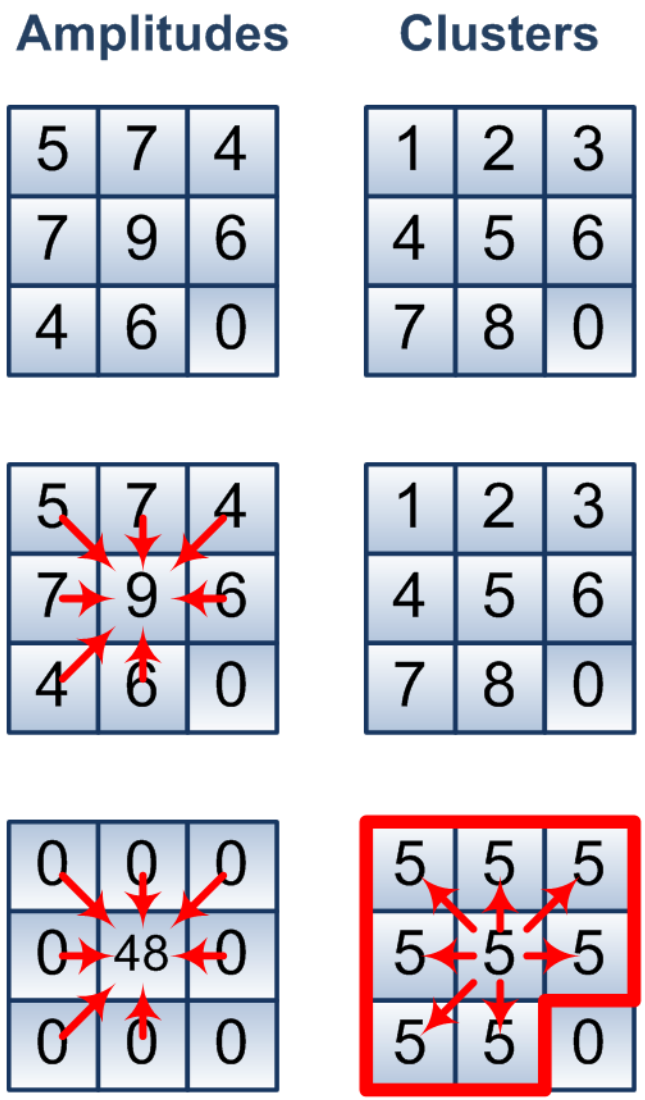
Single linkage method adds the object to the cluster, if this object is close to one of the objects already included in this cluster.



In order to make implementation fast and easy calculation of distance matrix was omitted.

Pads were merged on the basis of their neighborhood.

Developed clustering algorithm

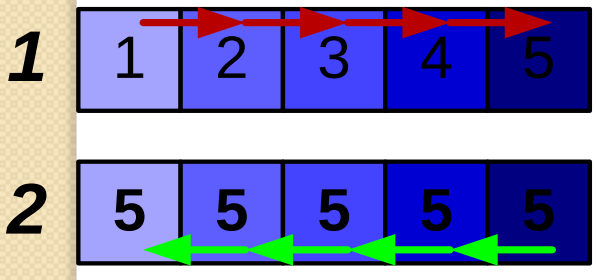


Algorithm:

1. each pad with a non-zero amplitude is considered as a separate cluster;
2. if the pad is not a local maximum, then it is joined the neighbor with the largest amplitude, else it is a separate object;
3. algorithm is executed until all pads are reviewed.

Attachment of pads

Two approaches for combination of group of pads were developed:

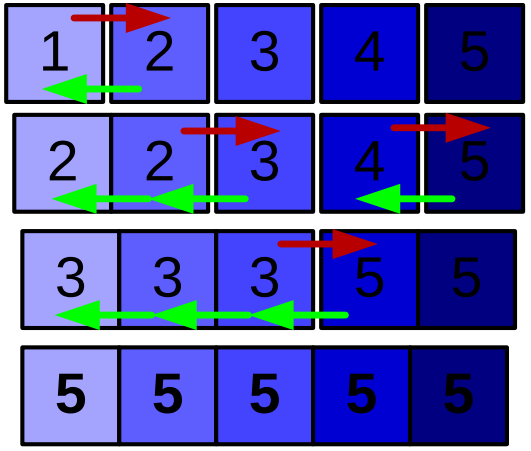


Method 1:

- Create a list of adjacent pads.
- Override pad numbers according to the list of adjacent pads.

Method 2:

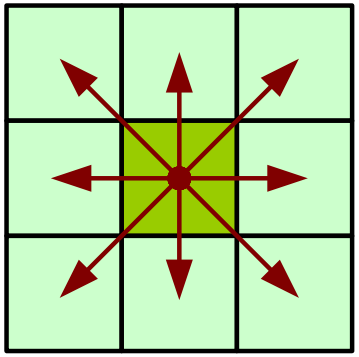
- Merge pads immediately after the definition of communication.
- Override pad numbers by a recursive function.



Neighboring pads

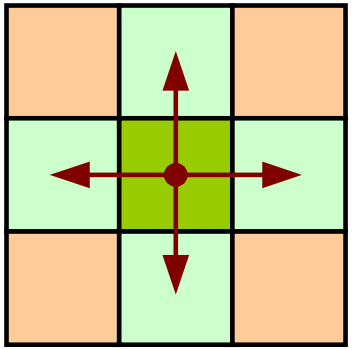
Algorithm use two different ways to determine neighbors :

1



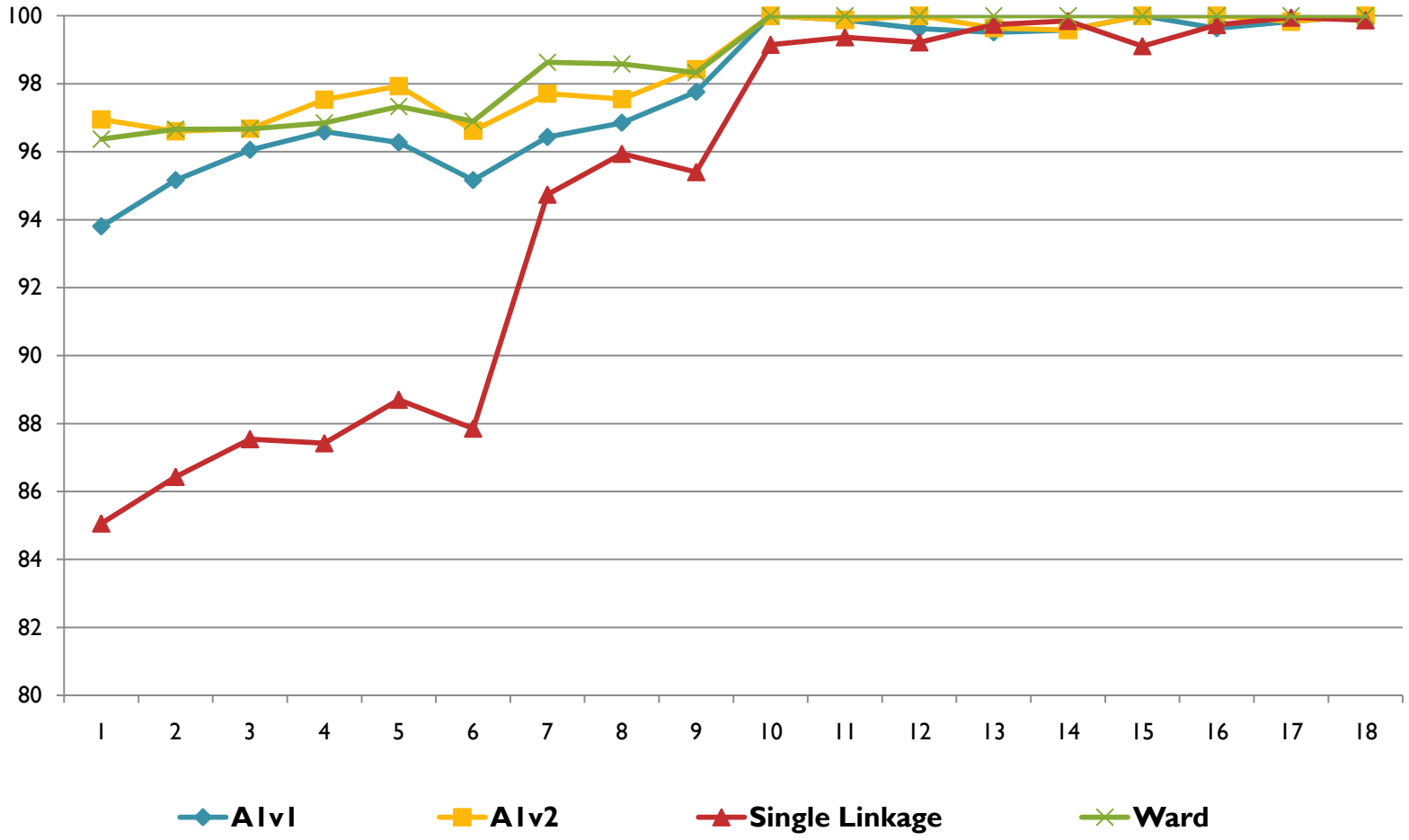
All neighbors: Take into account all adjacent.

2

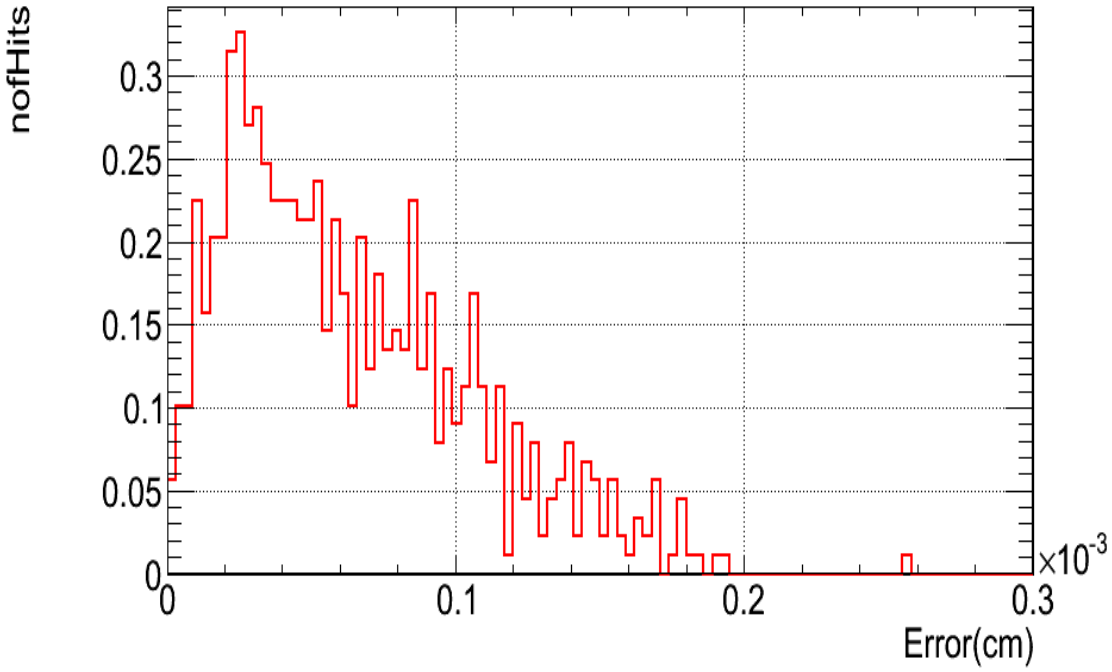


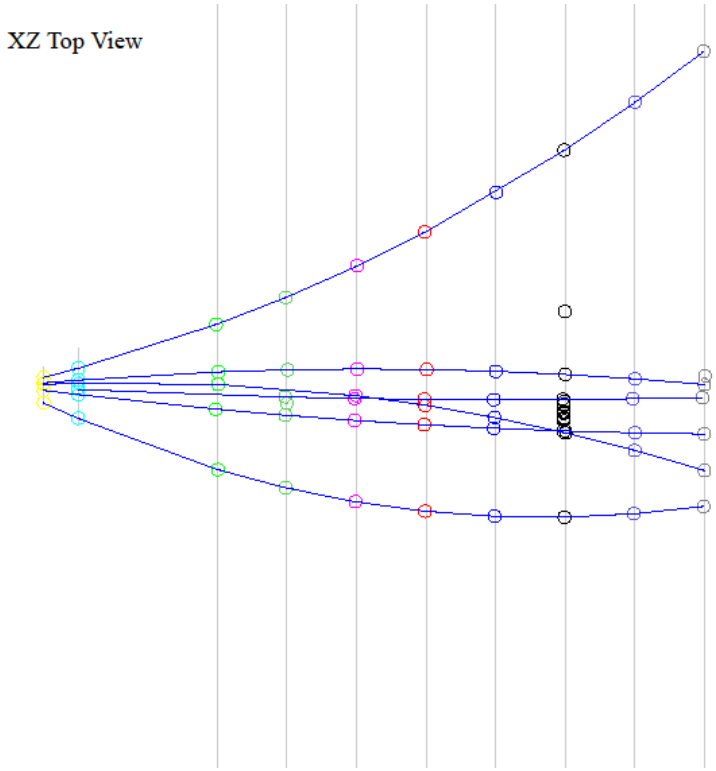
Nearest neighbors: Do not take into account angular pads.

Test results for MuCh: Efficiency



Test results for MVD: pC 30Gev centr

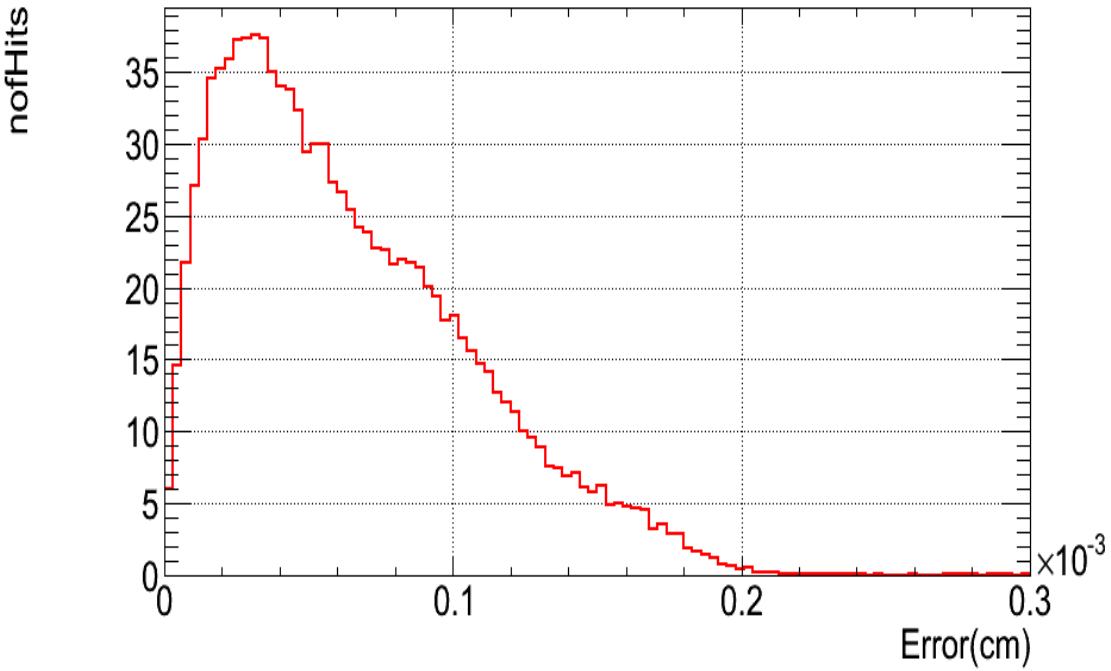


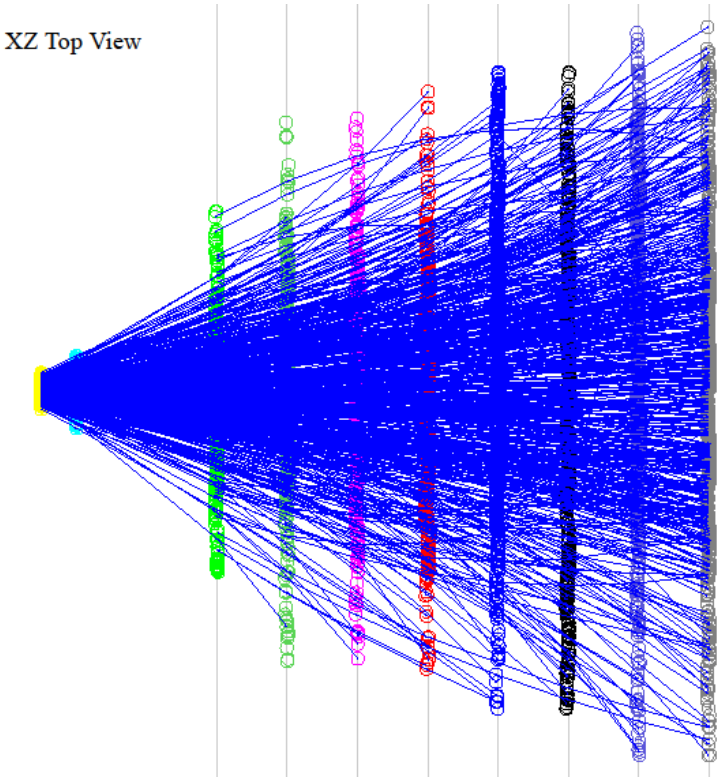


	Total	Station 1	Station 2
Lost Clusters	**0.16**	**0.07**	**0.09**
Fake Clusters	**0.06**	**0.03**	**0.03**
Efficiency	**98.29%**	Mean error	**0.0000707 cm**

7 MC Points per event

Test results for MVD: AuAu 25Gev centr





	Total	Station 1	Station 2
Lost Clusters	**29.96**	**10.22**	**19.74**
Fake Clusters	**4.26**	**1.84**	**2.42**
Efficiency	**97.81%**	Mean error	**0.0000905 cm**

1150 MC Points per event

Summary

- Clustering algorithms were discussed.
- Initial implementation of the clustering algorithms was developed:
  - tested for MuCh and MVD detector;
  - integrated with standard CBMROOT reconstruction;
- MVD clustering algorithm was investigated for different collision types:
  - pC, CC, pAu, AuAu centr and AuAu mbias;
  - efficiency is more than 97% for all cases:
    - algorithm is robust to track multiplicity;
  - errors are on the level of 2 μm;
  - low number of fake clusters.

Outlook:

- improvement of the Qa:
  - implementation based on littrack Qa;
- generalization of developed approaches.

Thank you for your attention