9th CERN Patatrack Hackathon @TUE

GROUP 3, 14 JULY 2021

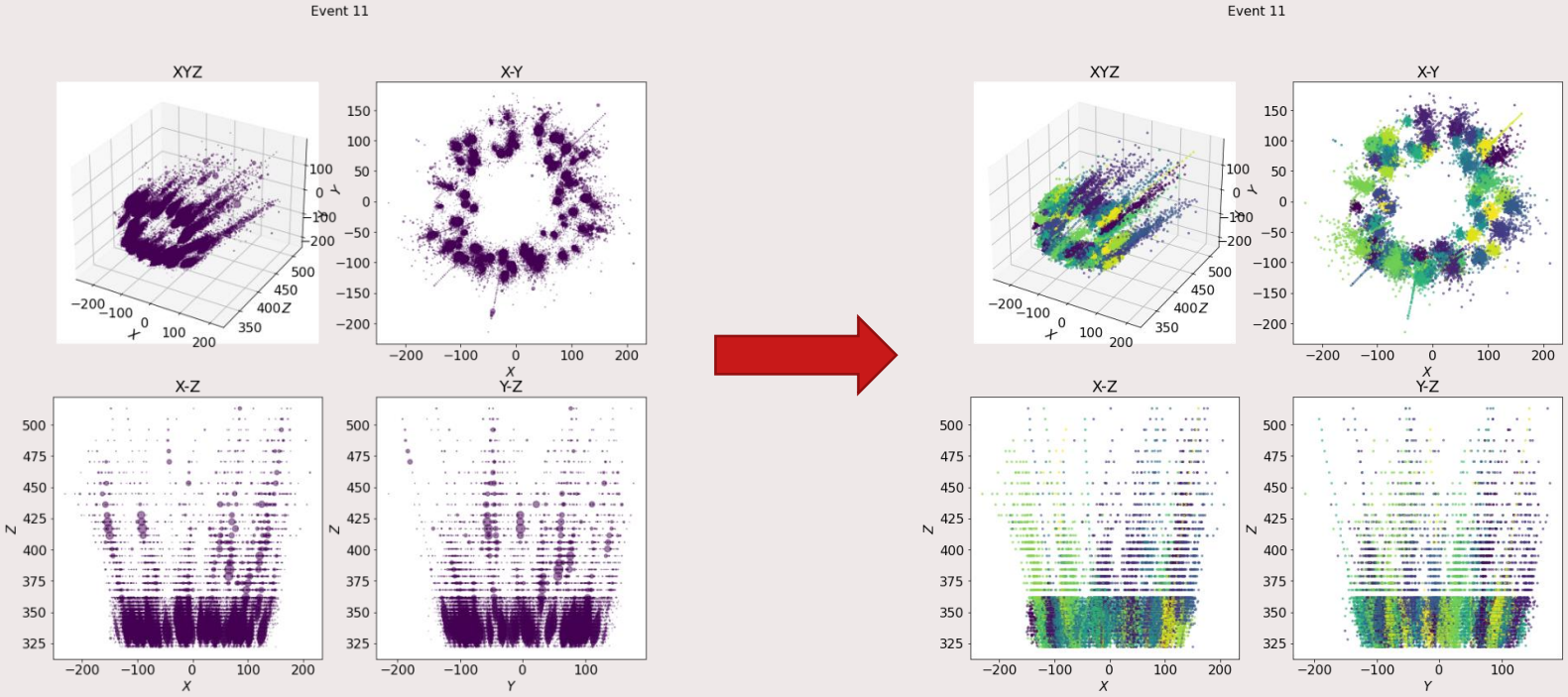
Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham

Department of Hacking

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Goal

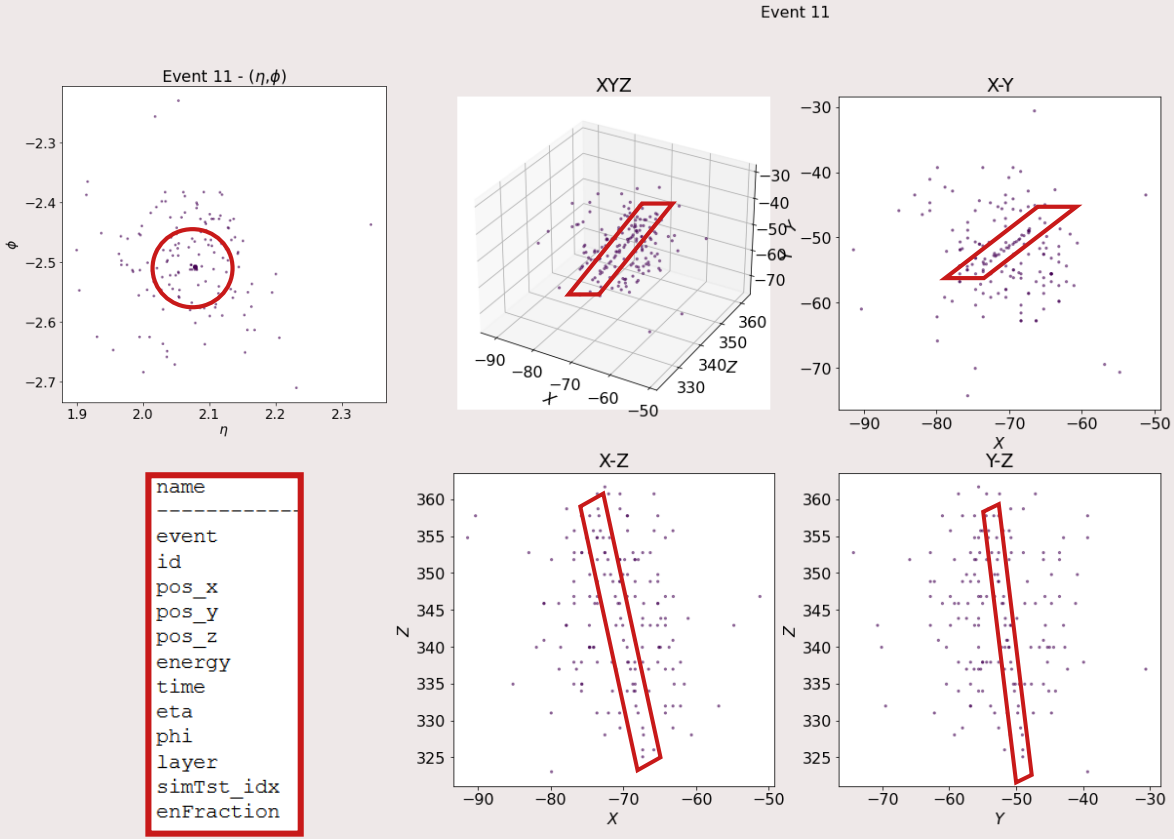
Identifying tracksters and the points belonging to them



Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

Datasets

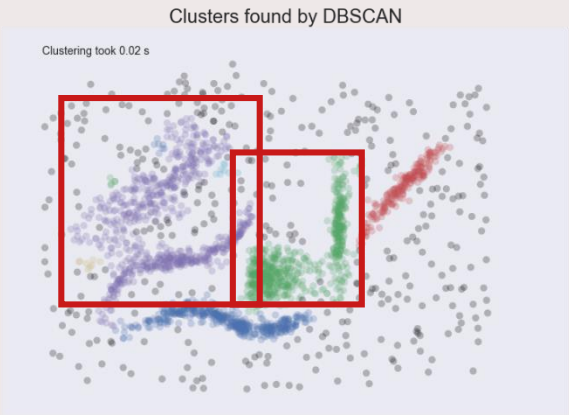
- Contains information about positions, energy, time, and layers
- Intuitively, there are areas that are more **densely** populated
- Some **pattern intuition** involving **Z**

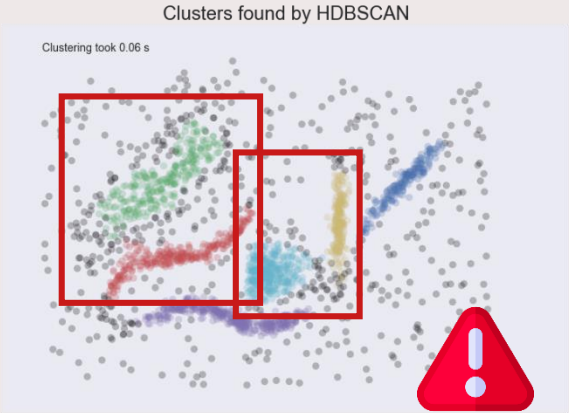


Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

Main idea

- Focusing on **density-**> DBSCAN algorithm
- DBSCAN uses **core samples** to build the clusters around them -> More **energetic points**
- HDBSCAN = DBSCAN + varying density clusters





Source

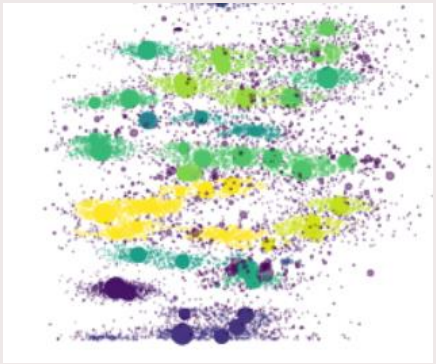
Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

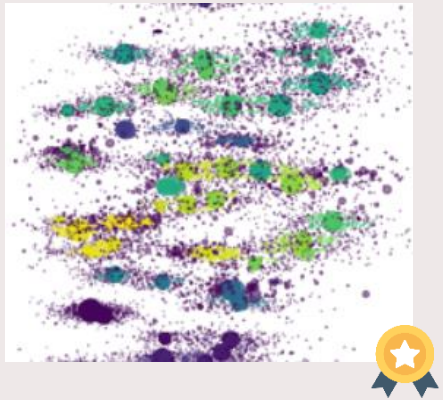
Findings

```
clusterer = hdbscan.HDBSCAN(min_cluster_size=15,
                            min_samples=7,
                            cluster_selection_epsilon=0.002)
```

```
labels = clusterer.fit_predict(df[['eta','phi']])
```

- Clustered on density based on η and ϕ
- You can fine-tune HDBSCAN with these parameters
  - min_cluster_size and min_samples -> # to consider a cluster
  - cluster_selection_epsilon -> minimum separation between clusters



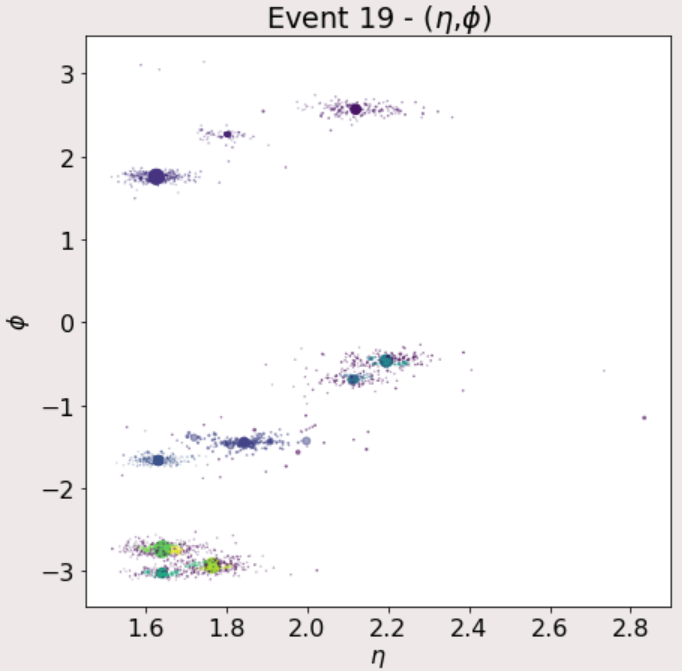


ε=0.02

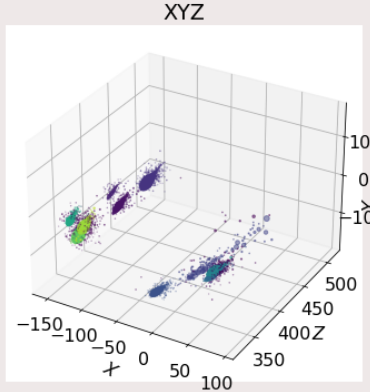
ε=0.002

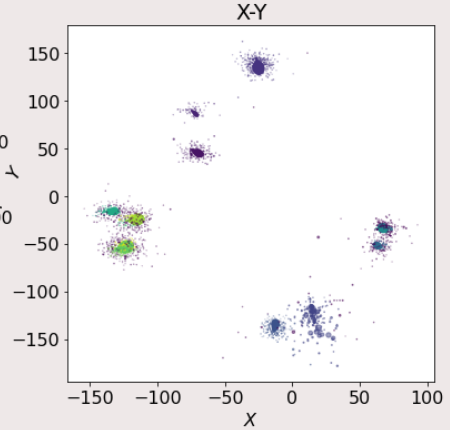
Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

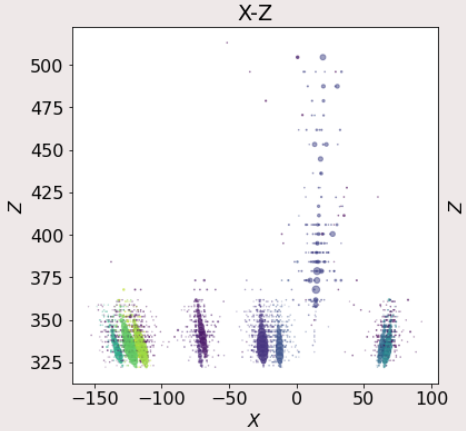
Results (I)

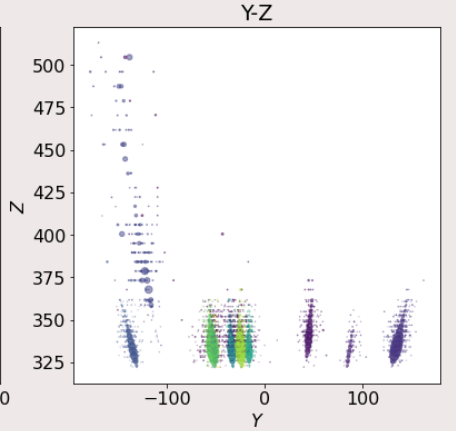



```
The total clusters are:  21
```









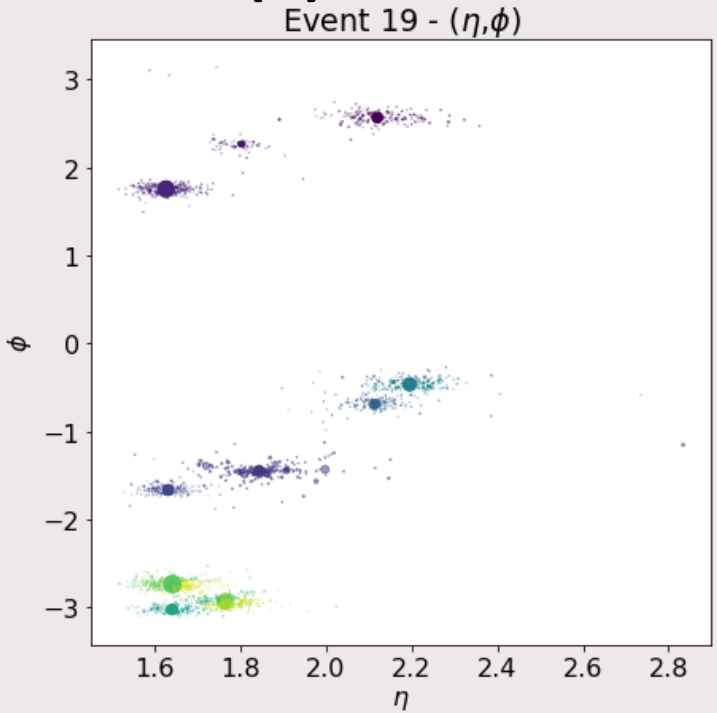
Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

Classifying outliers

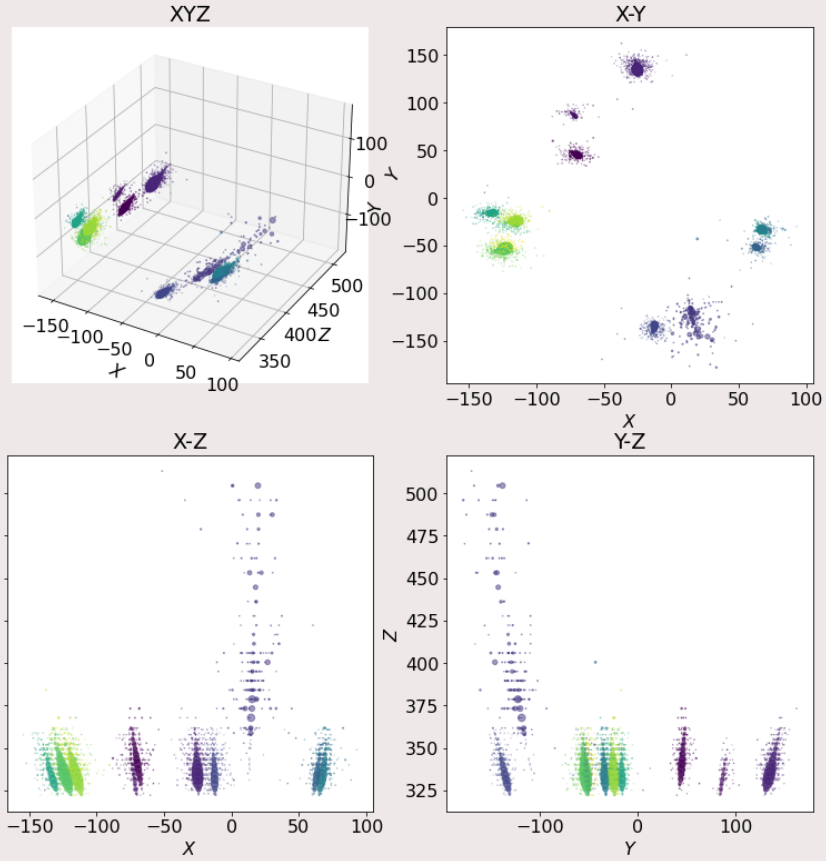
- KDTree on the classified points
- For the not classified:
  - Find the nearest point in the classified tree
  - Take the cluster from that point
  - Assign that cluster to the unclassified point

Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

Results (II)



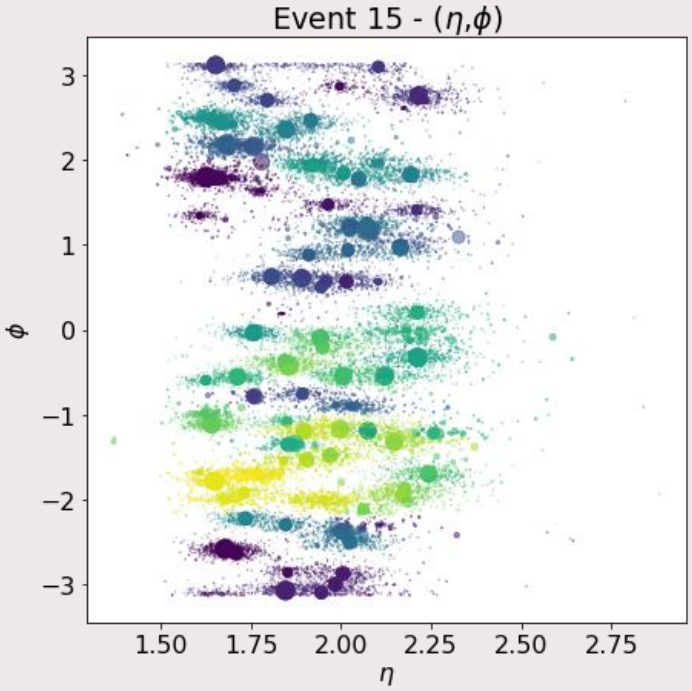

```
The total clusters are:  21
```



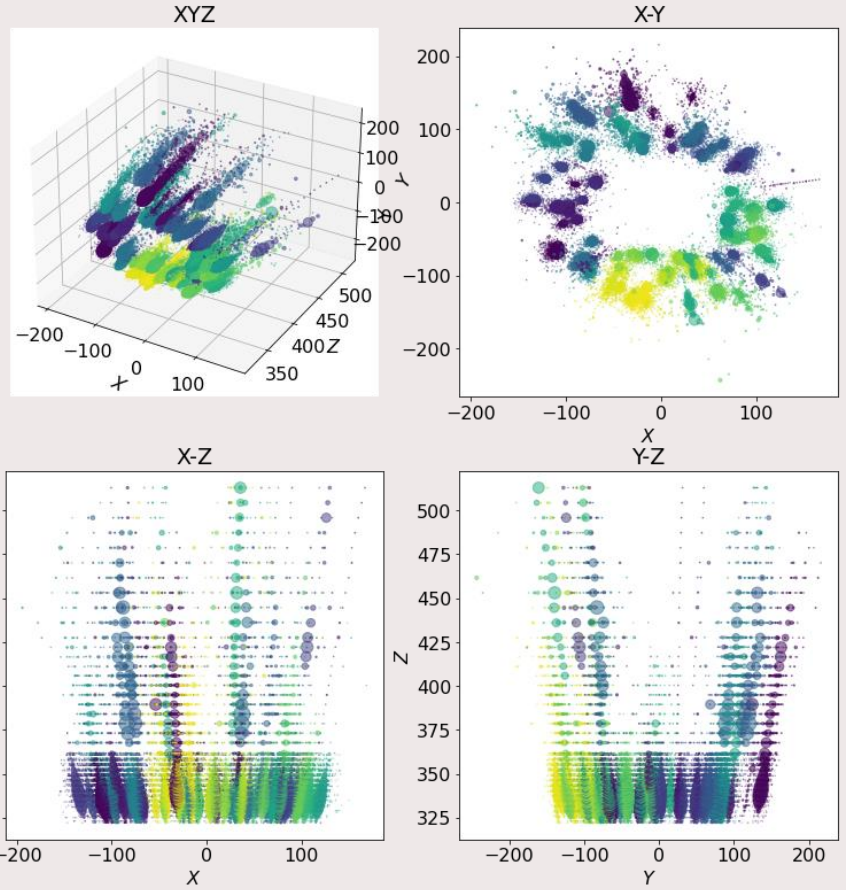
Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

TU/e

Results (II)



The total clusters are: 210



Luis Roma Barge, Mees Willem Bieling, Ngoc Quan Pham
Group 3 - Final Solution

Future possible improvements

Best approach is raw clustering to get the number of clusters following with precise (density-based) clustering method such as:

1. Flatten the image in each clusters' angle and use a stricter cluster algorithm based on the energy density
2. Use HBDSCAN to create a tree based on energy density and pick the appropriate level contour that matches to expected energy of the particles
3. Create circles around points and expand until it includes n points, the radius of each circles corresponds whether it is a core point, boundary point or noise point (the outliers)
4. **Extra:** Use temporal dependencies to distinguish different particle showers for files with overlapping showers