Track identification in SiW ECal

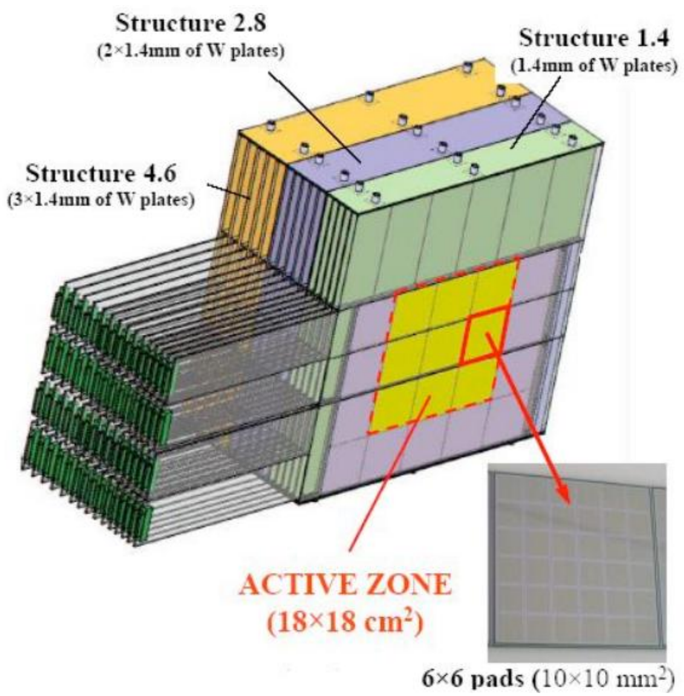
Sviatolav BILOKIN, Roman PÖSCHL
LAL, Orsay

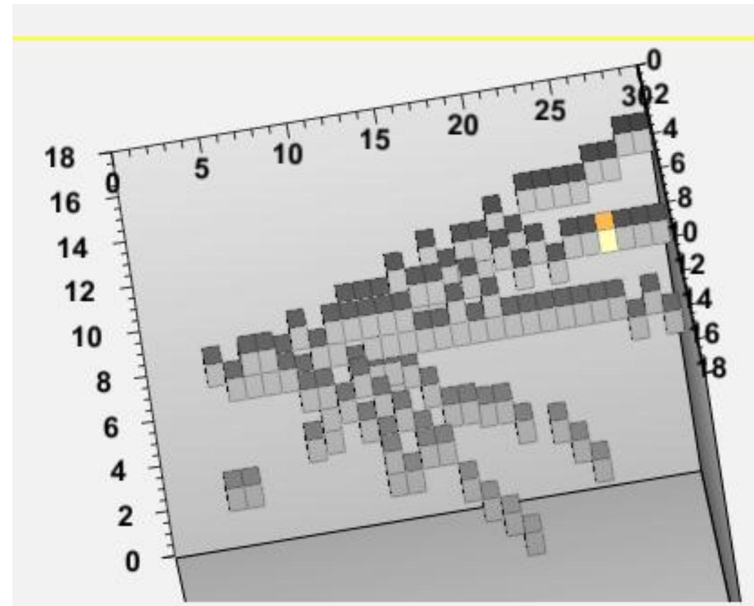
Third JCL @ LPSC Grenoble

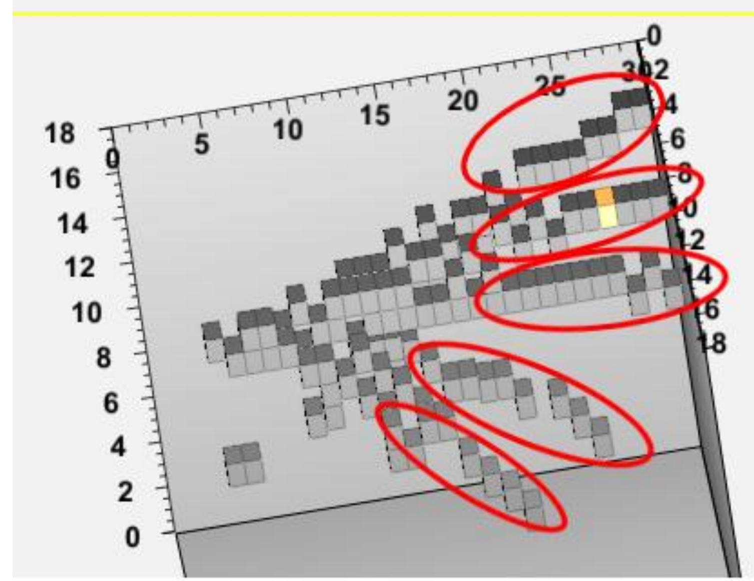
LABORATOIRE DE L'ACCÉLÉRATEUR LINÉAIRE

Objective

- Detect secondary particles from hadronic interactions in SiW ECal.
- The number of outgoing tracks can help to improve MC simulations and PFA
- Data studied: π^- between 2 and 10 GeV (FNAL, 2008)





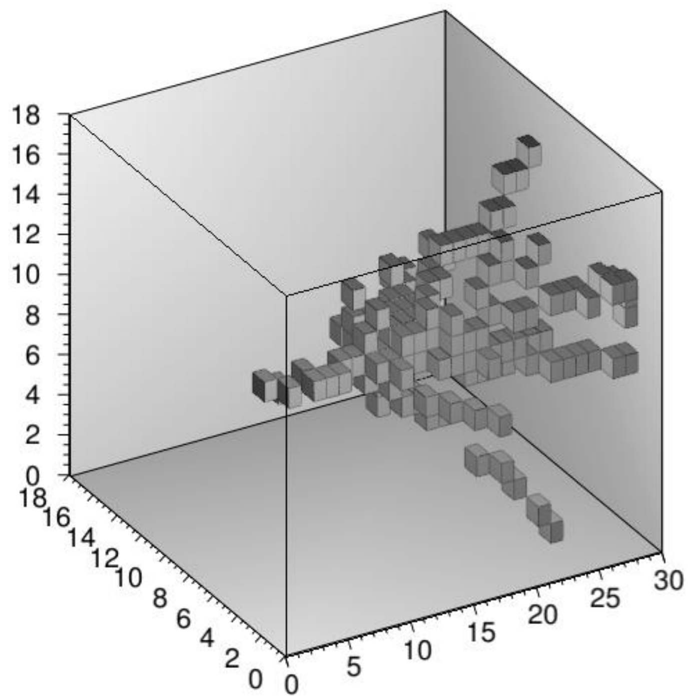


Analysis strategy

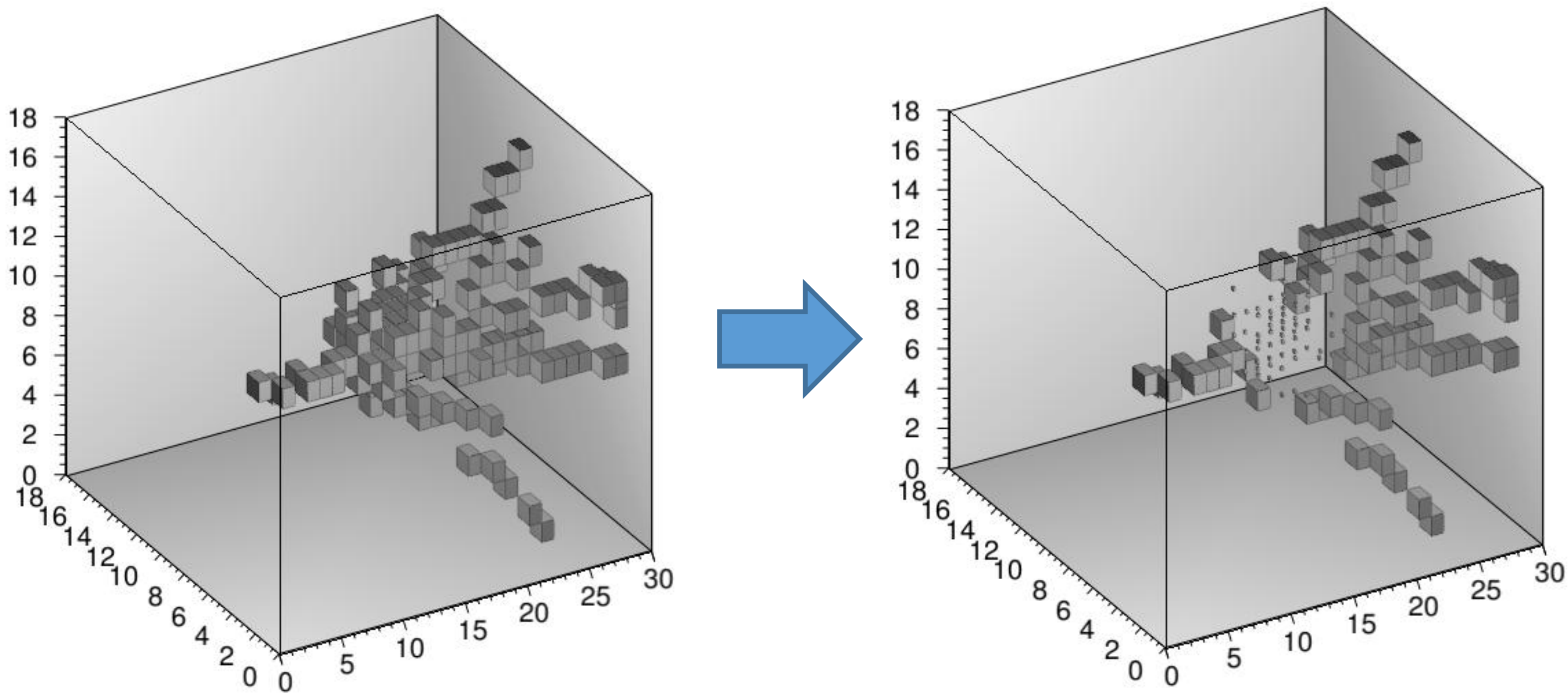
- Preselect the events, which have an energy deposition at the last layers of the ECal and interaction in the first half of the Ecal
- Remove an interaction zone from the events
- Collect all possible topologically connected energy depositions in ECal to clusters (clusterization stage)
  - At the last layers of the ECal the tracks are more distinguishable, so starting clusterization from the last layers will have an advantage.

- Classify each cluster as track like or not track like (analysis stage)
  - Build an observable quantity and condition for classification.
  - We are using 'calorimeter coordinates' for pad position:

$$\begin{cases} x = 1,2, \dots 18 \\ y = 1,2, \dots 18 \\ z = 1,2, \dots 30 \end{cases}$$

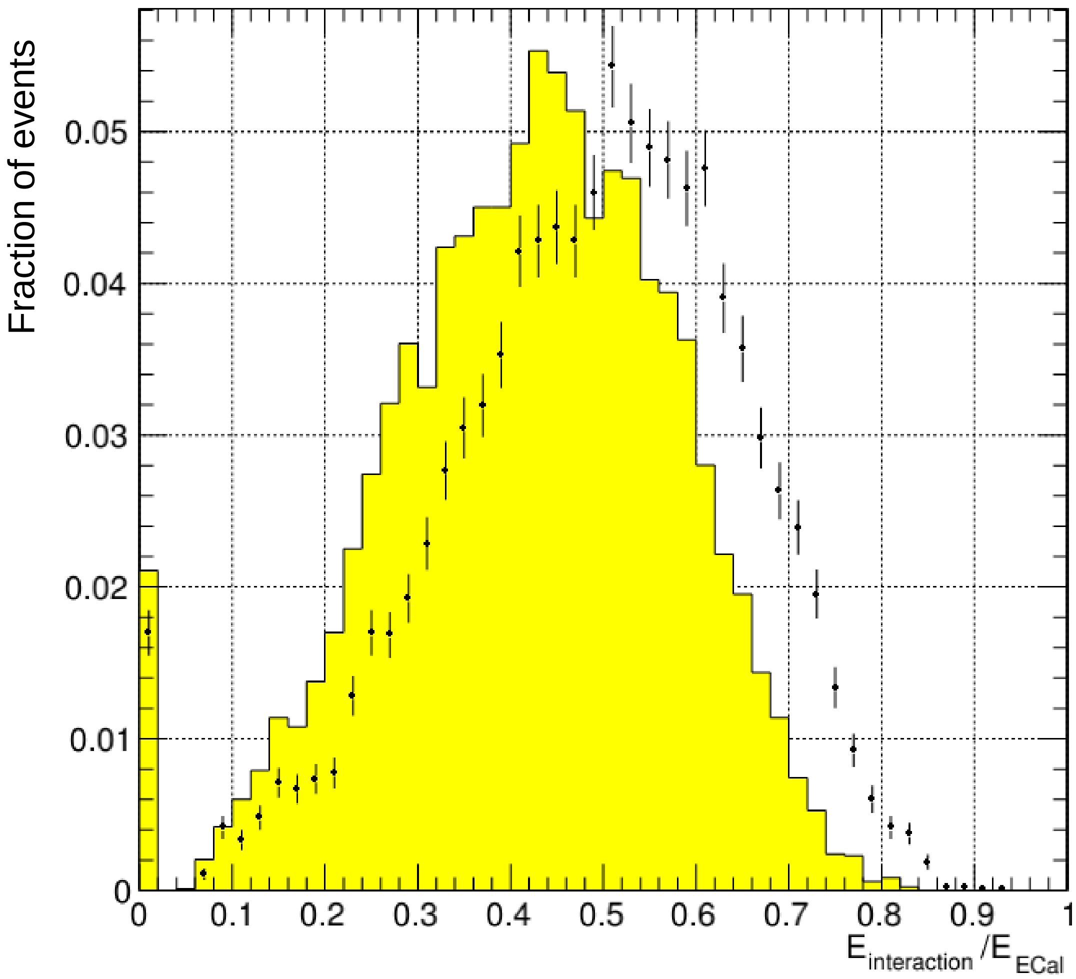


Cut out of interaction zone



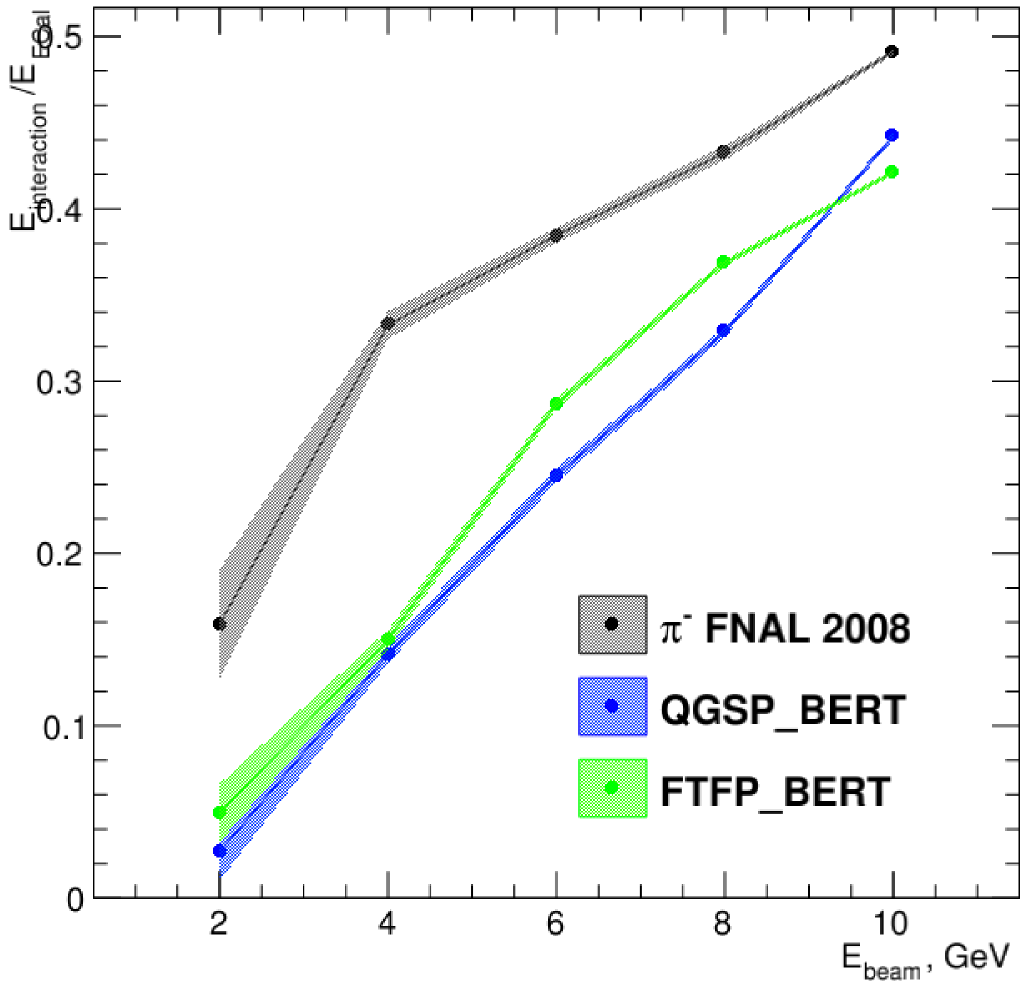
The interaction zone is removed by the topological criteria and ignored by further track finding algorithm.

Energy of interaction zone



Energy fraction of the interaction zone for $10\ \mathrm{GeV}\ \pi^{-}$ for the the data (dots) and FTFP Bertini MC (yellow). Data events have more energy deposition in interaction zone.

Energy of interaction zone

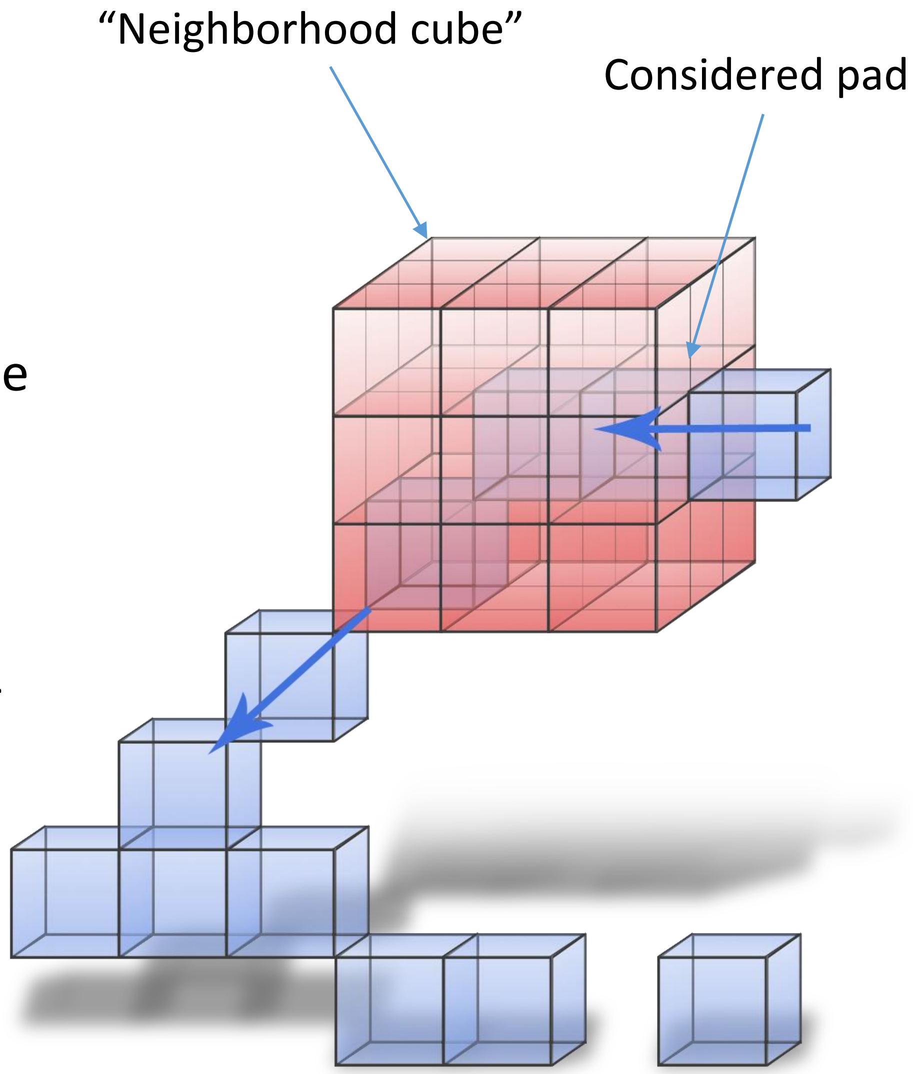


The $< E_{interaction}/ E_{ECal} >$ for π^- incoming particles of different energies. Error regions are statistical errors on the corresponding mean values. In the data there is more energy deposition in the interaction zone.

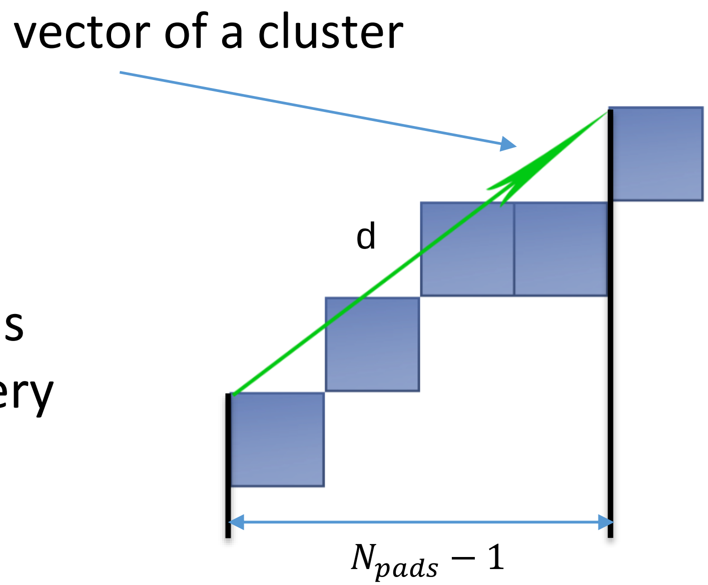
Clusterization

Clusterization is based on recursion algorithm:

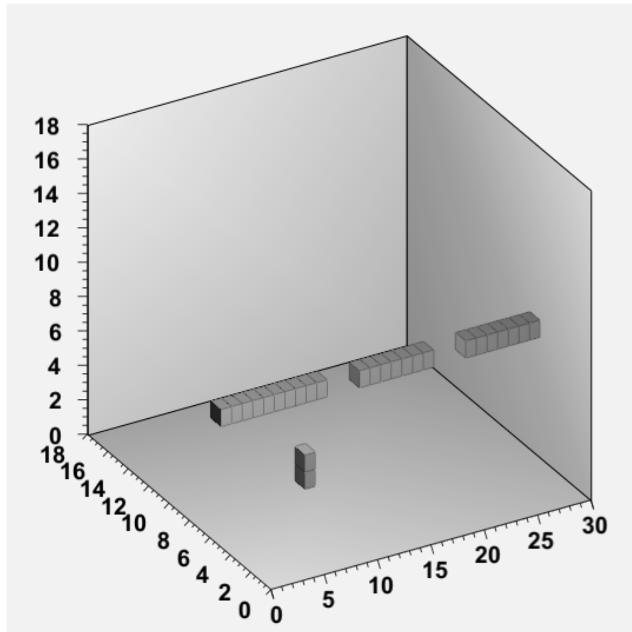
- On each step, for a given active pad algorithm finds the active neighbors, and put them into one cluster.
- The process ends if:
    - all neighbors found are already in another clusters
    - no more active neighbors of a given pad found
- As a result, all topologically connected energy depositions will form some amount of clusters



Classification



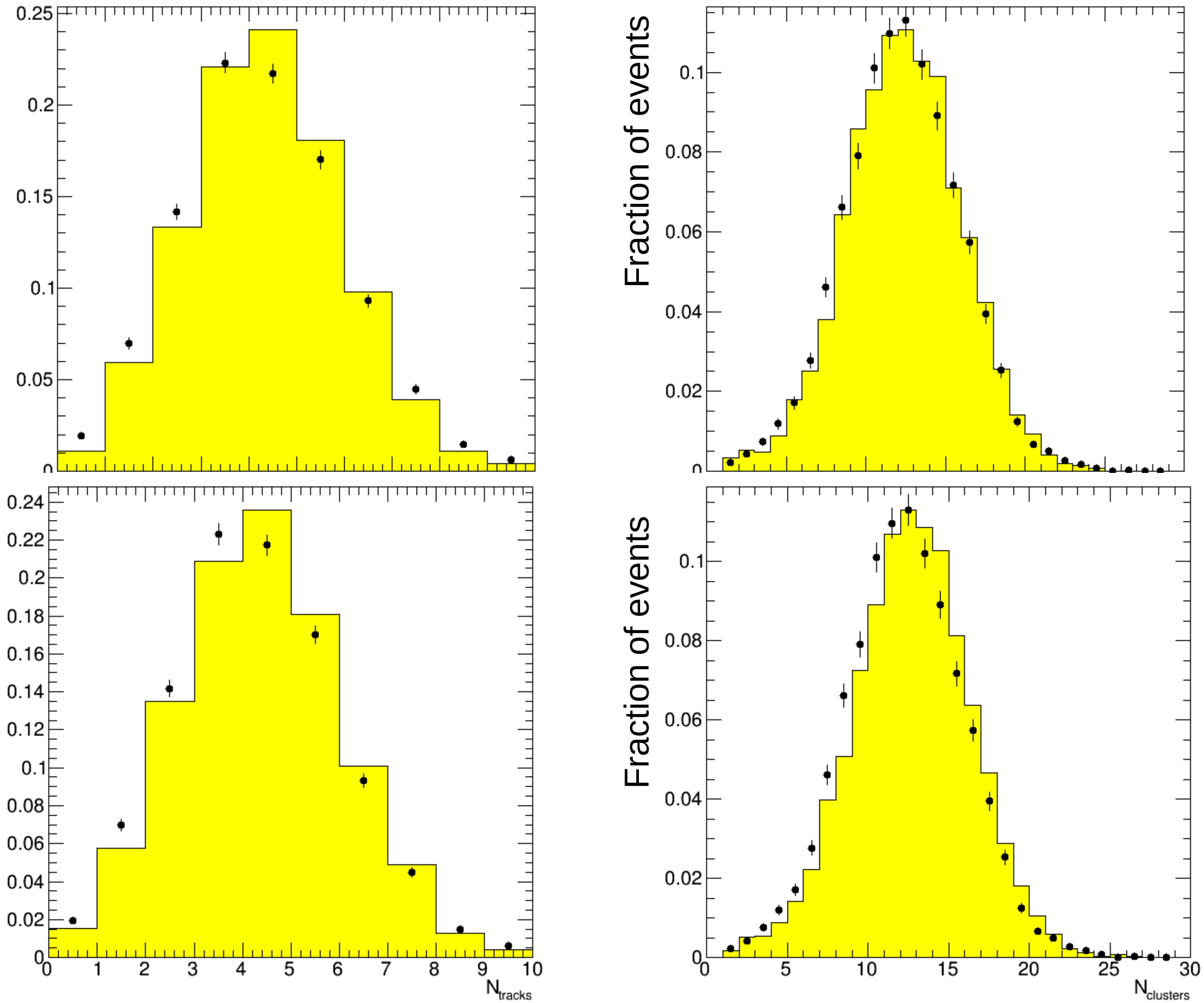
- Classification has following steps:
  - cluster with < 3 pads is noise
  - compute the module of cluster as maximal distance d between every pad in cluster;
  - compute an observable:
    $$\xi = \frac{d}{N_{pads} - 1} + \epsilon N_{pads},$$
    where ϵ is a correction for imperfect tracks, $\epsilon \ll 1$
  - if $\xi \geq 1$ then cluster is a track
  - other clusters could be classified as two close MIPs;
- After classification it is necessary to reconstruct segmented tracks using cone algorithm



Comparison of N_{tracks} and $N_{clusters}$

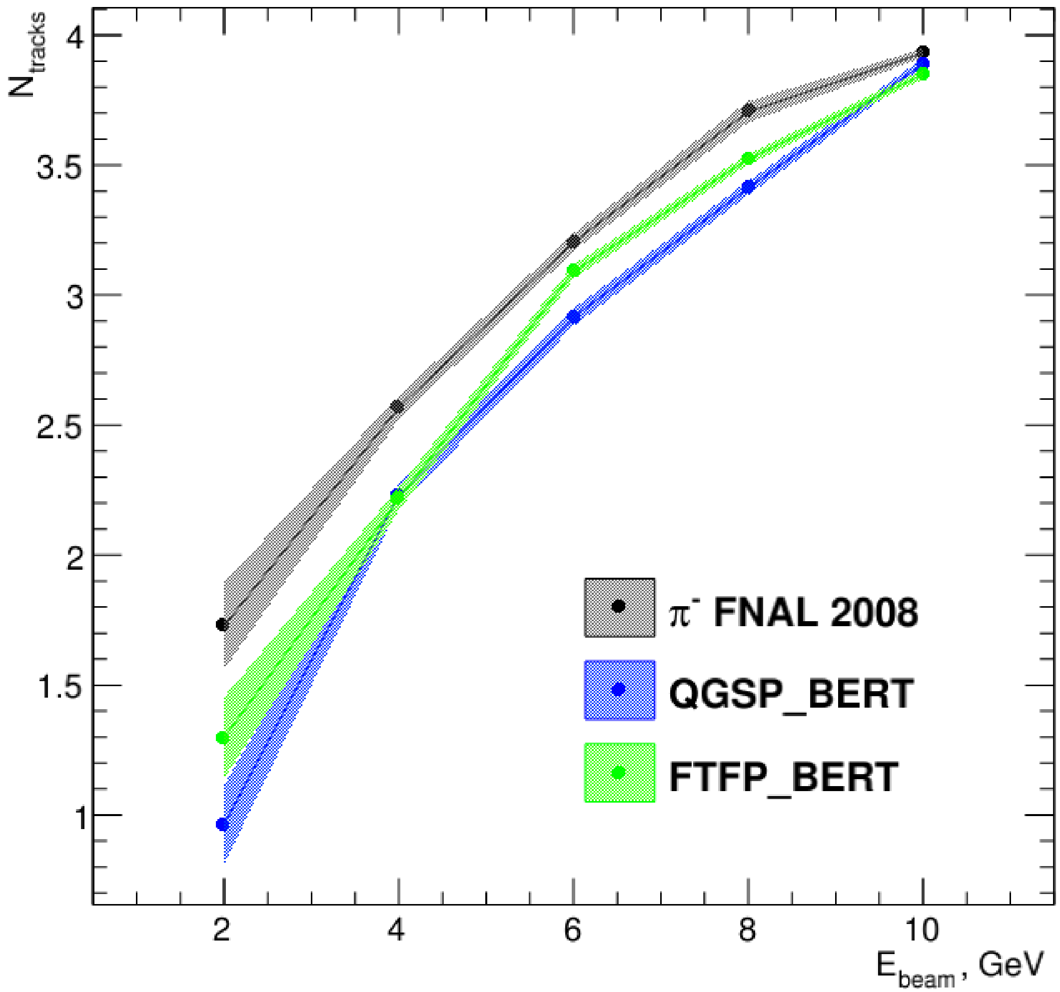
FTFP_BERT

QGSP_BERT



Comparison plots of 10 GeV pion data (dots) and MC (Geant 9.6 p01) (yellow) for N_{tracks} (left) and $N_{clusters}$ (right). The data/MC agreement is good.

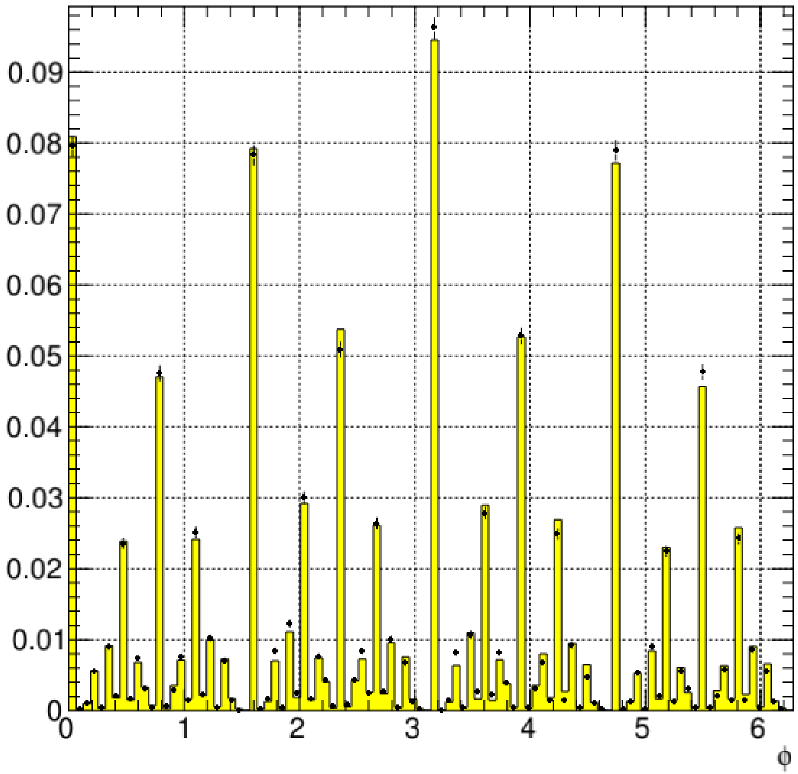
Number of tracks



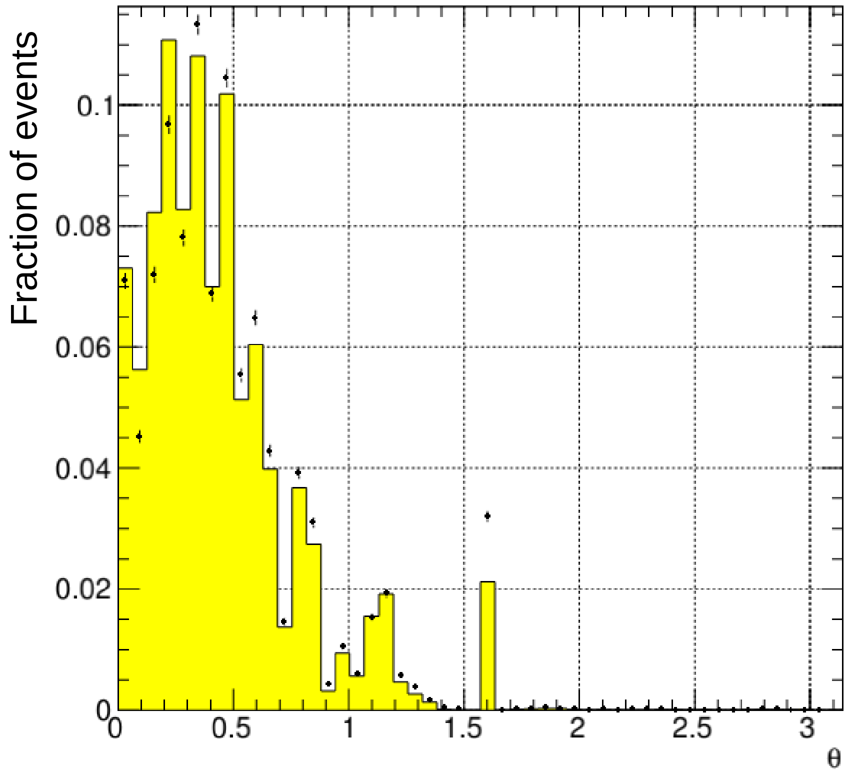
Average of the number of the tracks for different energy beams and MC physics lists. On average there is more tracks in the data.

Angular distribution of secondaries

ϕ distribution

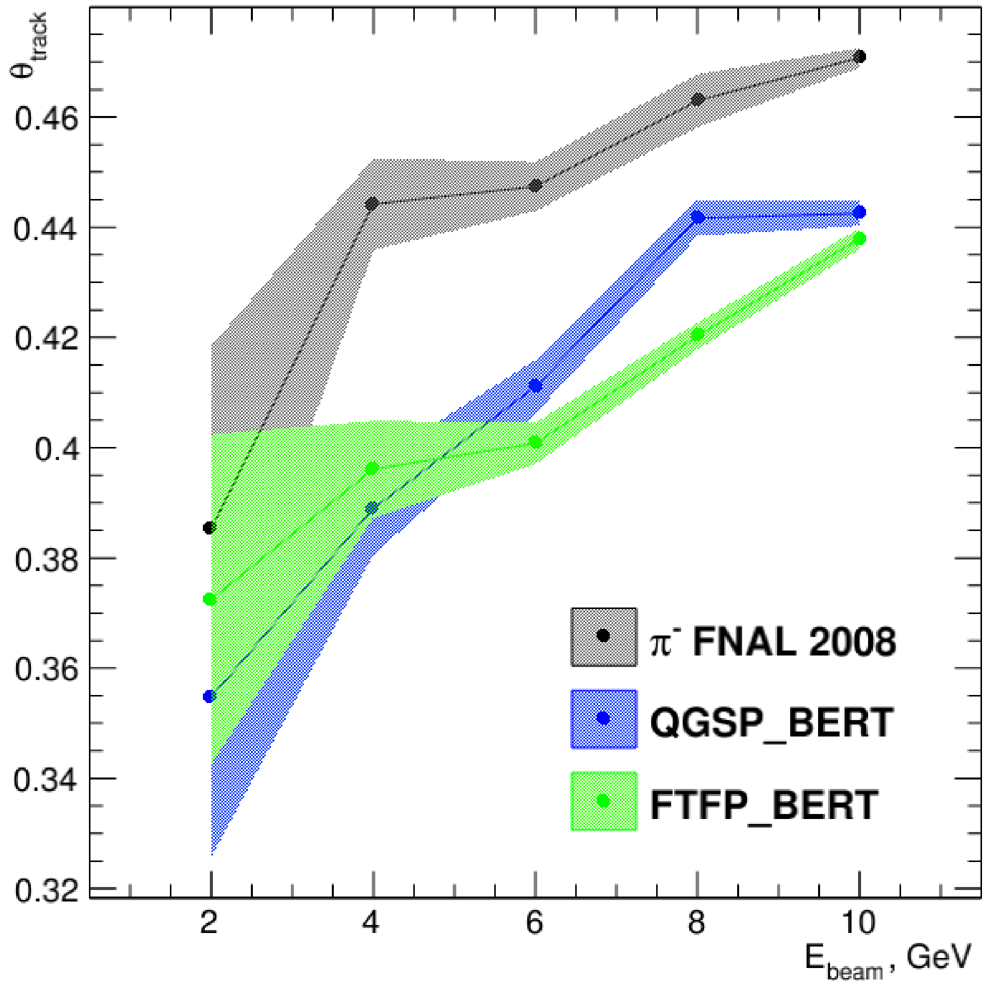


θ distribution



10 GeV $\pi^-$for the the data (dots) and QGSP Bertini MC (yellow).
Both distributions well coincide.

Mean θ angle

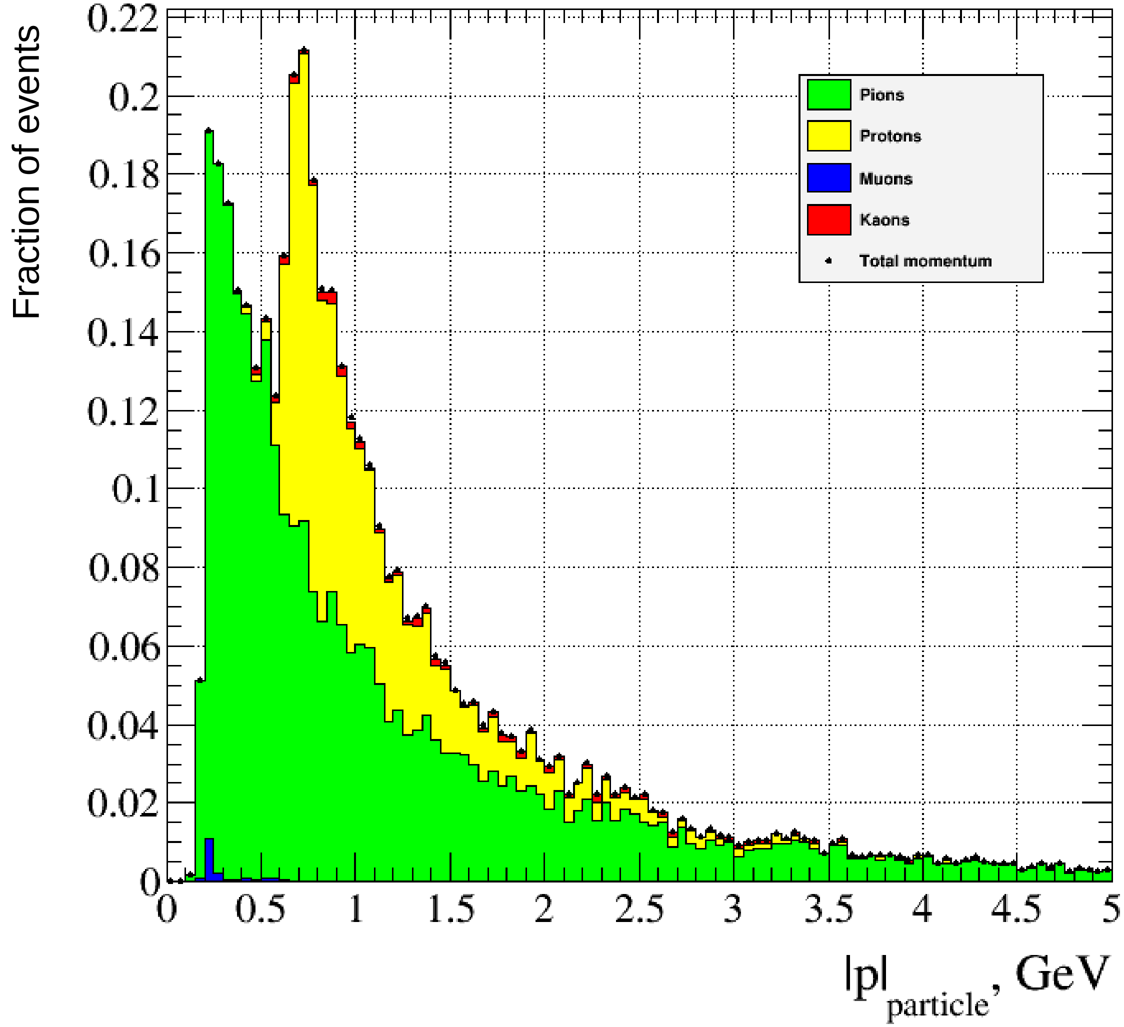


Average values of θ angles for the data. Cross section of π^- interaction increases with the E_{beam}, and cross section is larger for the data events.

Generator/Tracks correspondance

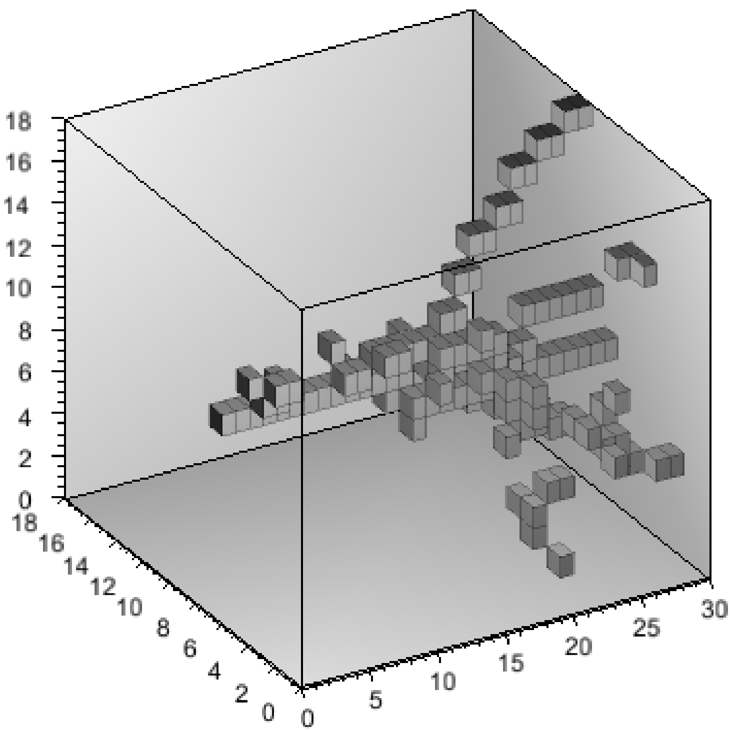
- One MC event can have between 2000 and 14000 MC particles in the electromagnetic calorimeter region.
- One needs to provide selection cuts on their kinematics to get the secondaries that can be reconstructed in the ECal:
    - Kinetic energy cut E_{kin}
    - Distance of flight cut d_f
    - Start point cut - particle should be produced within or slightly before the ECal
    - Neutral particles cut - only charged particles are considered
    - Merging of collinear particles by $sin\theta$ criteria
- The number of MC particles can be used as training data for machine learning algorithms.

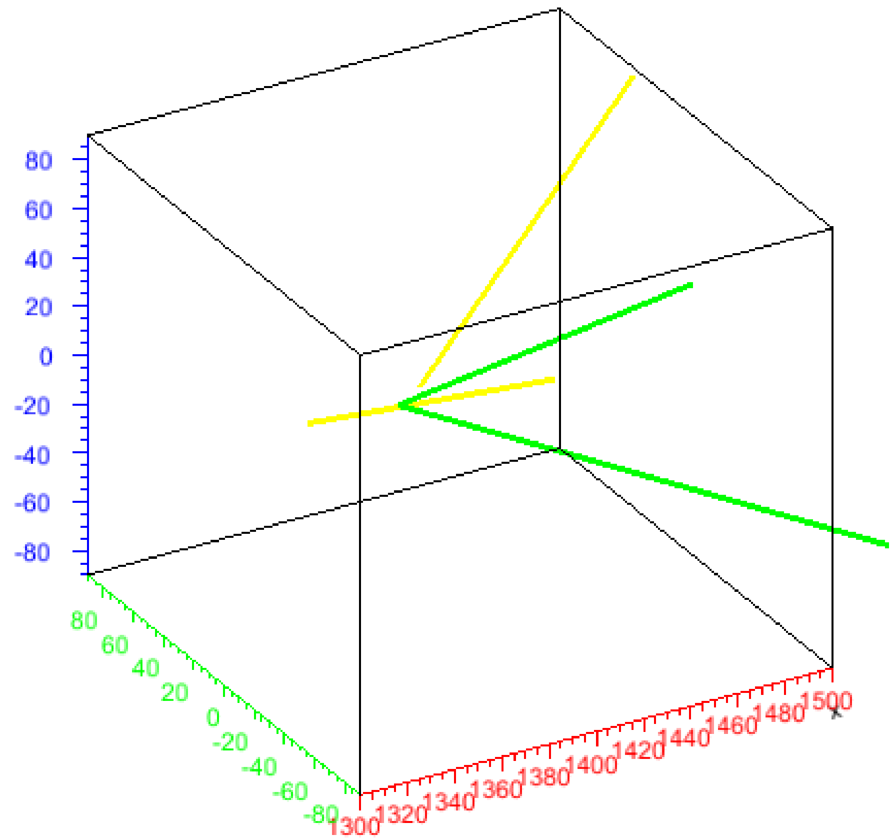
Momentum of MC particles



Momentum of MC particles for all events for 10 GeV pion interaction. The selection cuts are: $E_{kin} > 0.3\ GeV, d_f > 55\ mm$.

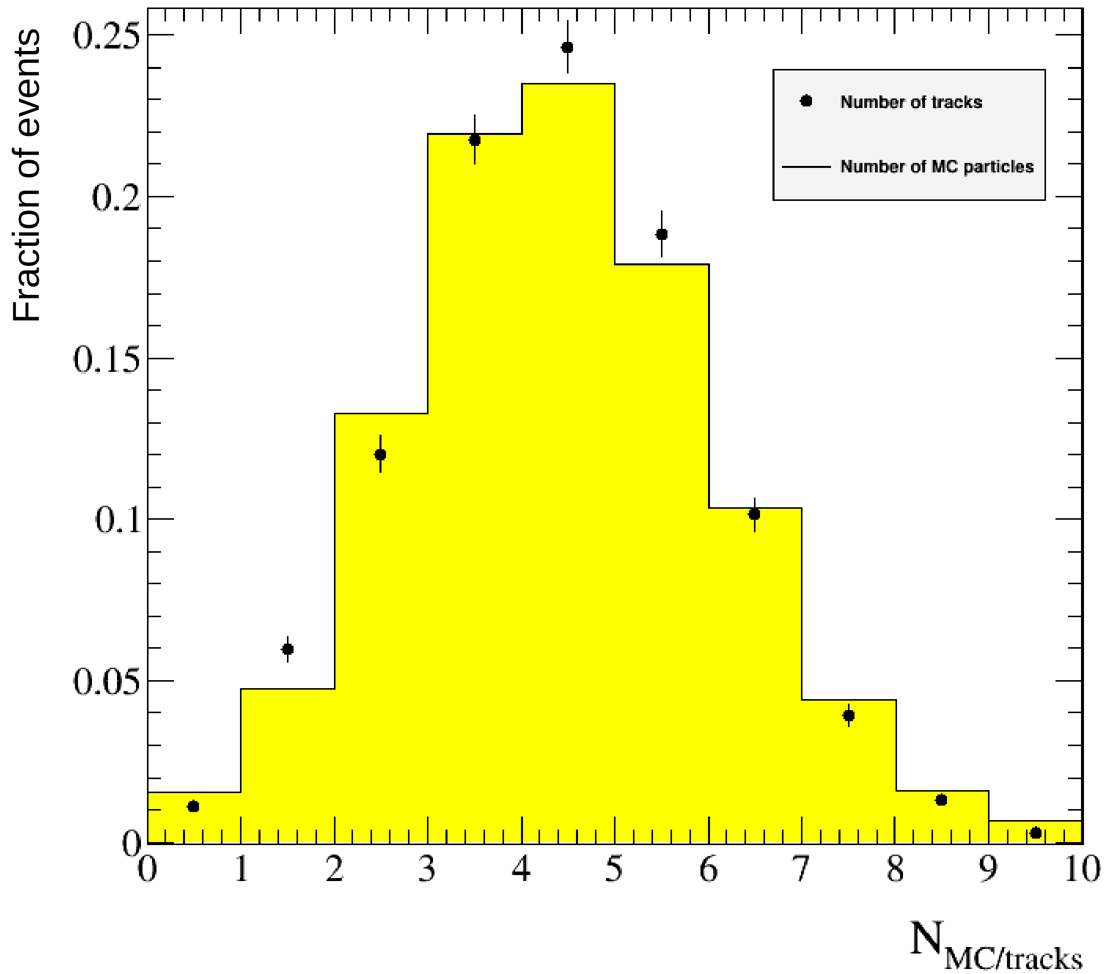
Comparison with MC truth





Study the correspondence between the number of found outgoing tracks and the number of secondary particles that are produced in the simulation. Yellow lines are protons, Green lines – pions.

Comparison with MC truth

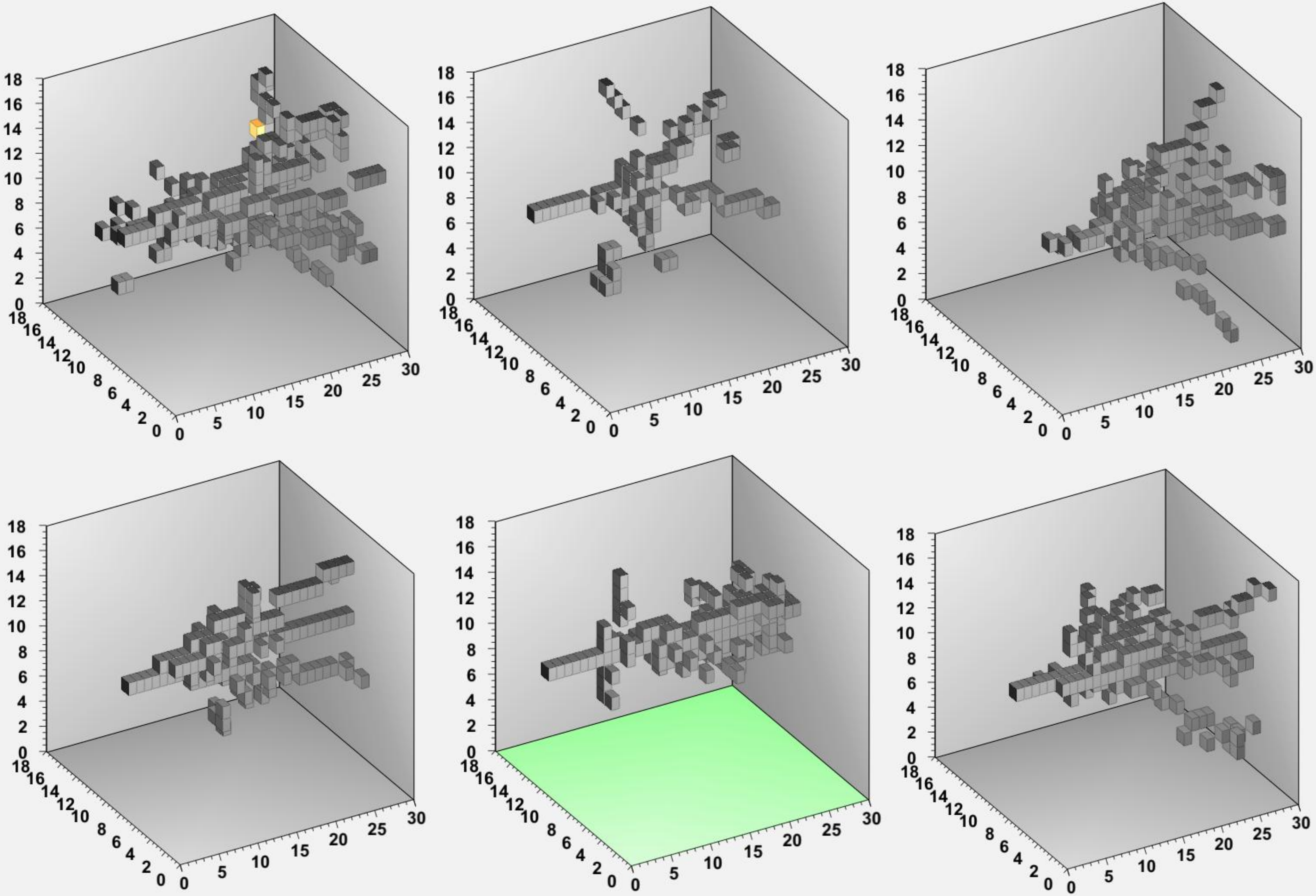


Study the correspondence between the number of found outgoing tracks and the number of secondary particles that are produced in the simulation. The selection cuts are: $E_{kin} > 0.3\ GeV, d_f > 55\ mm, \sin\theta > 0.2$.

Conclusion

- We have developed and tested an algorithm that finds the outgoing tracks in hadronic showers
- A good agreement between MC and data was found
- The number of found tracks was compared to the number of secondary particles produced in the simulation
- We can turn the MC secondary particles into an input data for machine learning algorithms
- Further work:
    - Study the systematics
    - Add information from HCal
    - Test the algorithm by MC with magnetic field
    - Compare to other algorithms

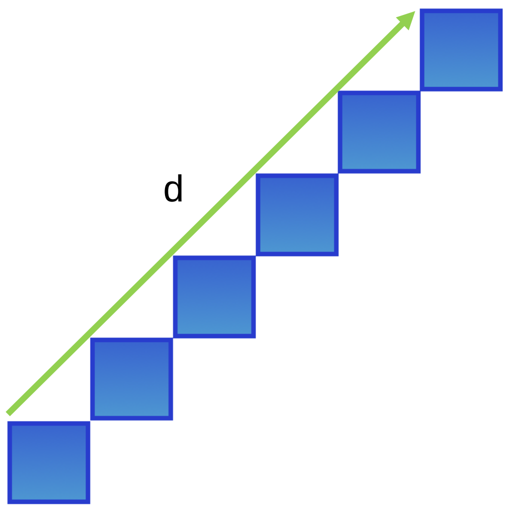
Thank you!


0
2
4
6
8
10
12
14
16
18
0 2 4 6 8 10 12 14 16 18
0 5 10 15 20 25 30

Choice of the observable

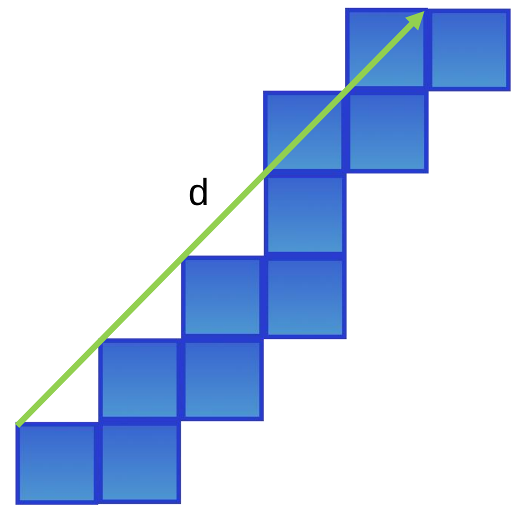
$$\xi = \frac{d}{N_{pads} - 1} + \epsilon N_{pads}$$

Ideal track



$$\frac{d}{N_{pads} - 1} > 1$$

Real track

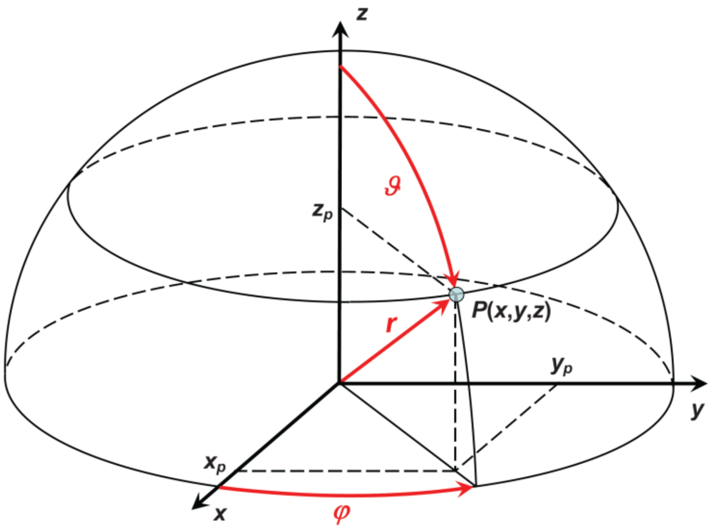


$$\frac{d}{N'_{pads} - 1} < 1$$

- One needs to apply a correction ϵN_{pads} to define second cluster as a track
- Free parameter ϵ is defined empirically and $\epsilon \sim 10^{-2}$

Angular distribution study

- We use common spherical coordinates to define the (ϕ, θ) angles for vector of the track
- The start and end points of a track have integer values in calorimeter units – the (ϕ, θ) space is discrete.



$$\phi = \mathrm{acos}\frac{y}{x}$$

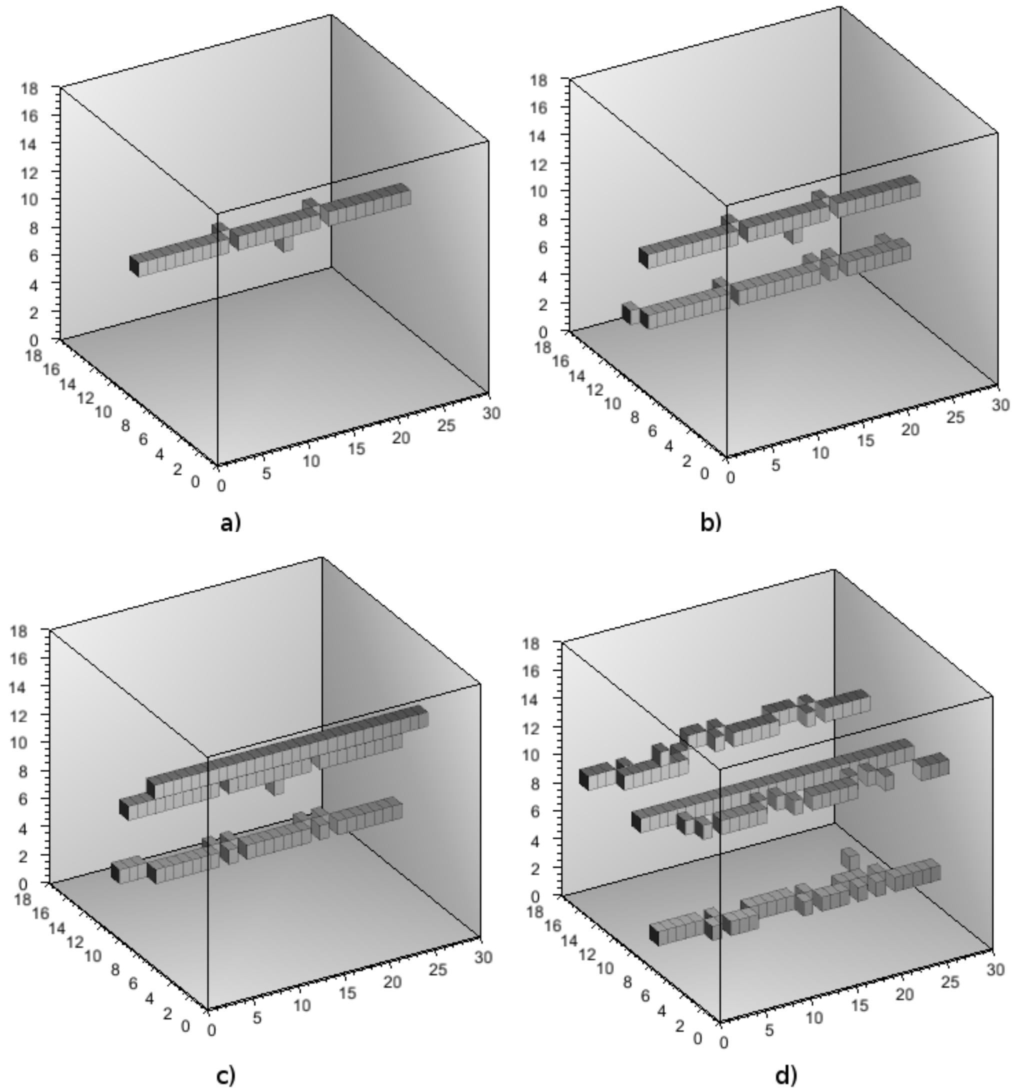
integer (y), integer (x)

$$\theta = atan\frac{z}{r}$$

integer (z), real (r)

- Due to this definition ϕ angle is quantized and θ is discrete for small tracks and more smooth for longer tracks

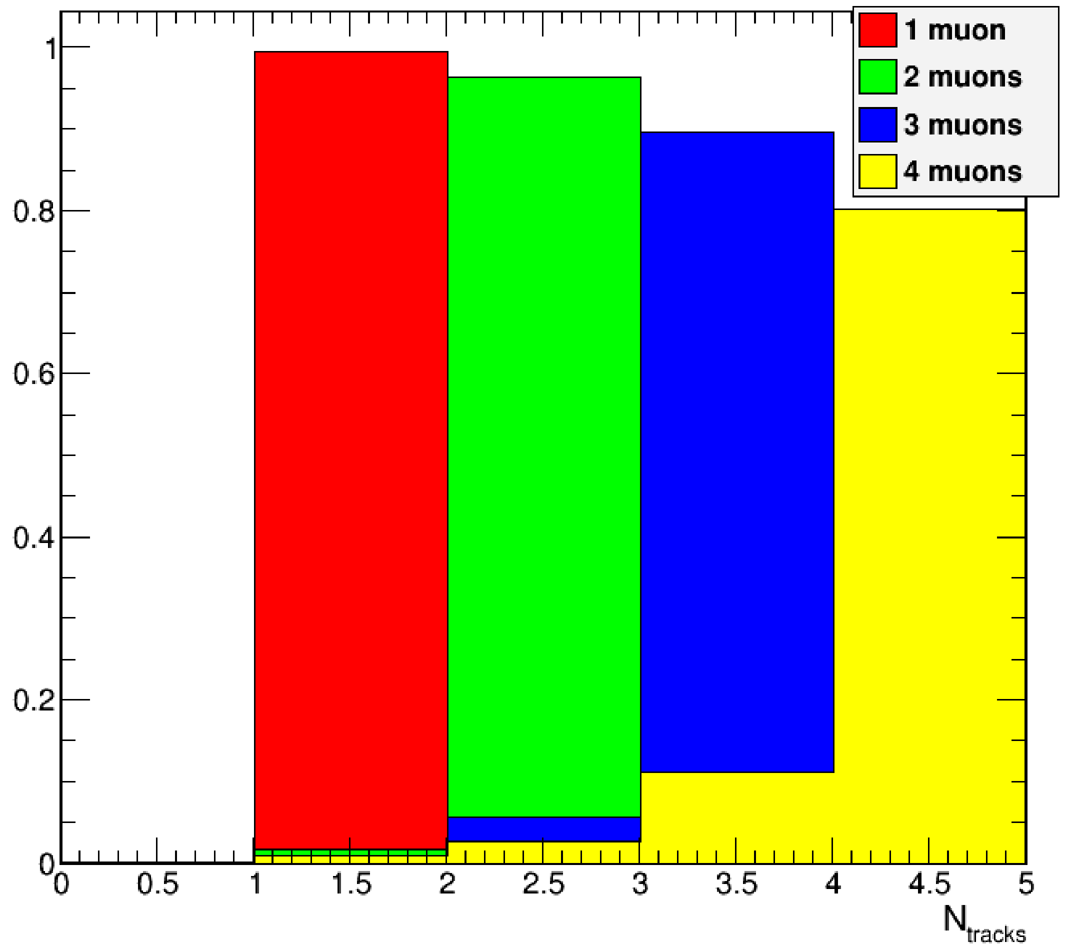
Consistency test



a) b) c) d)

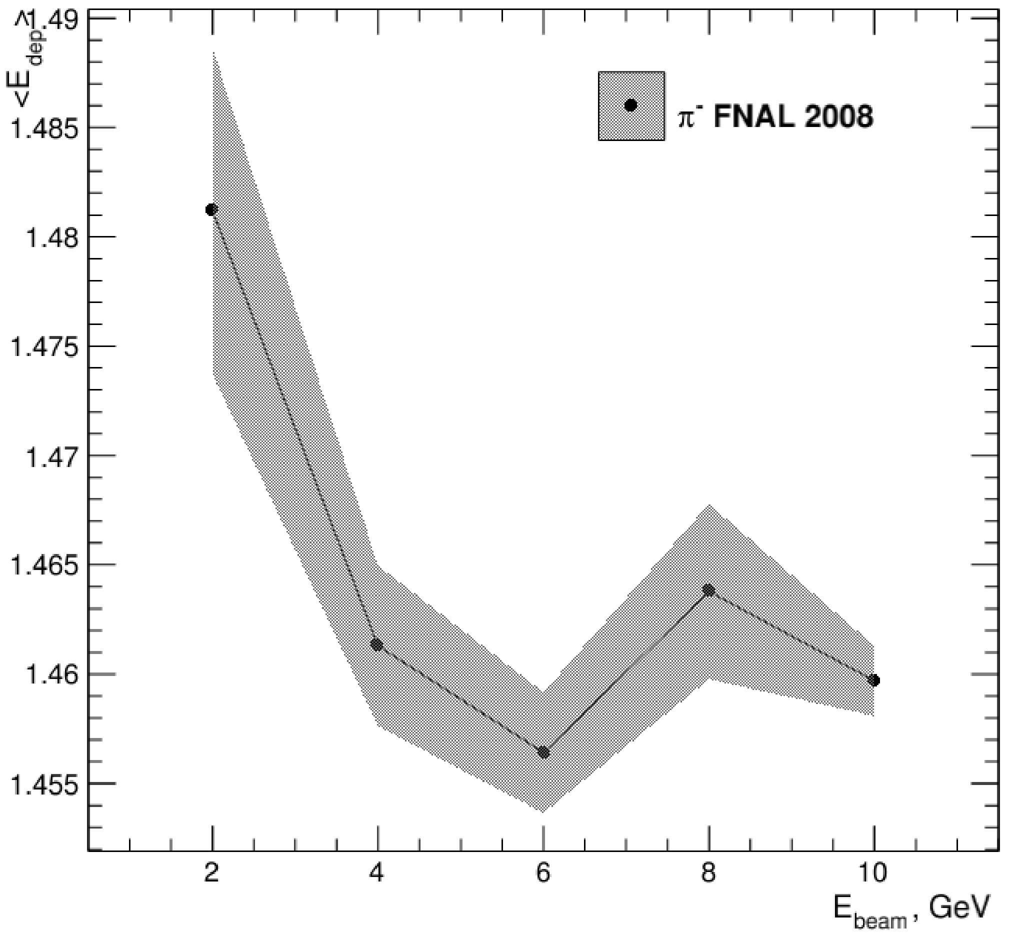
MC simulation of muons in the ECal with overlaid events.
Starting point of muon is smeared by a Gaussian with σ around 25 mm

Consistency test



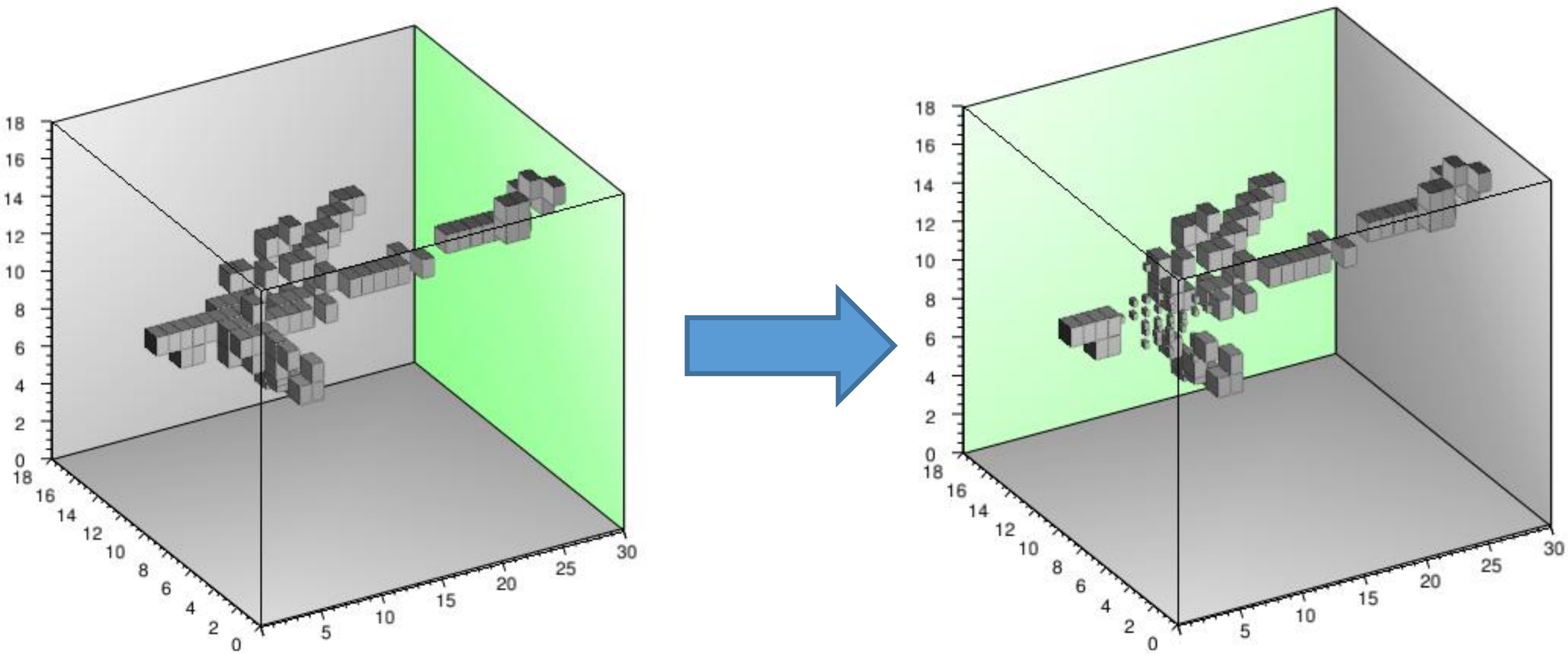
Track finding efficiency with 6 GeV muon MC events, which are multiple overlaid one over each other. Not stacked plot. Efficiency decreases with the number of tracks, but this is the lower bound of efficiency, because no minimal distance is required.

Mean energy deposition



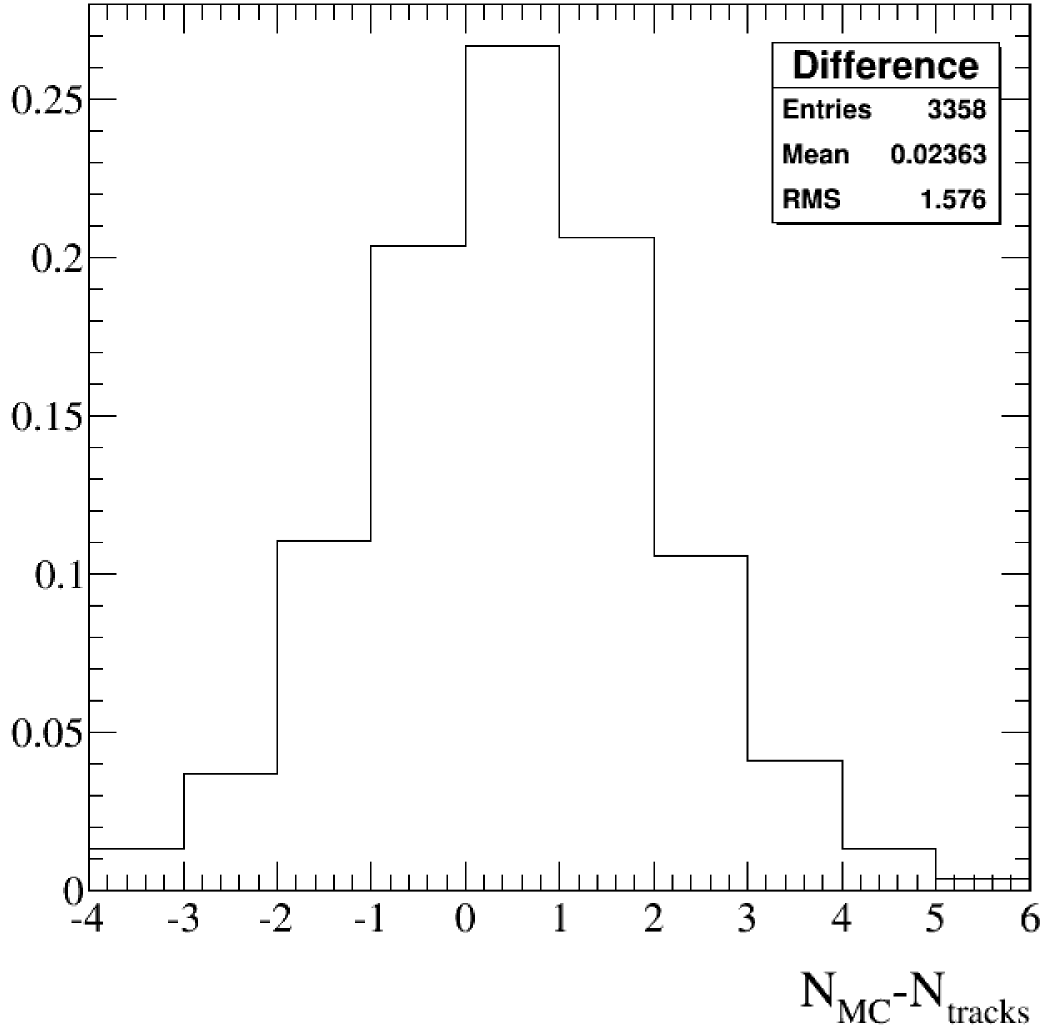
Mean energy deposition per pad in MIP units for π^- incoming particles of different energies. For this graph we selected events with incoming π^- as MIPs. This information can be used for ‘in situ’ calibration of the ECal

Cut out of interaction zone



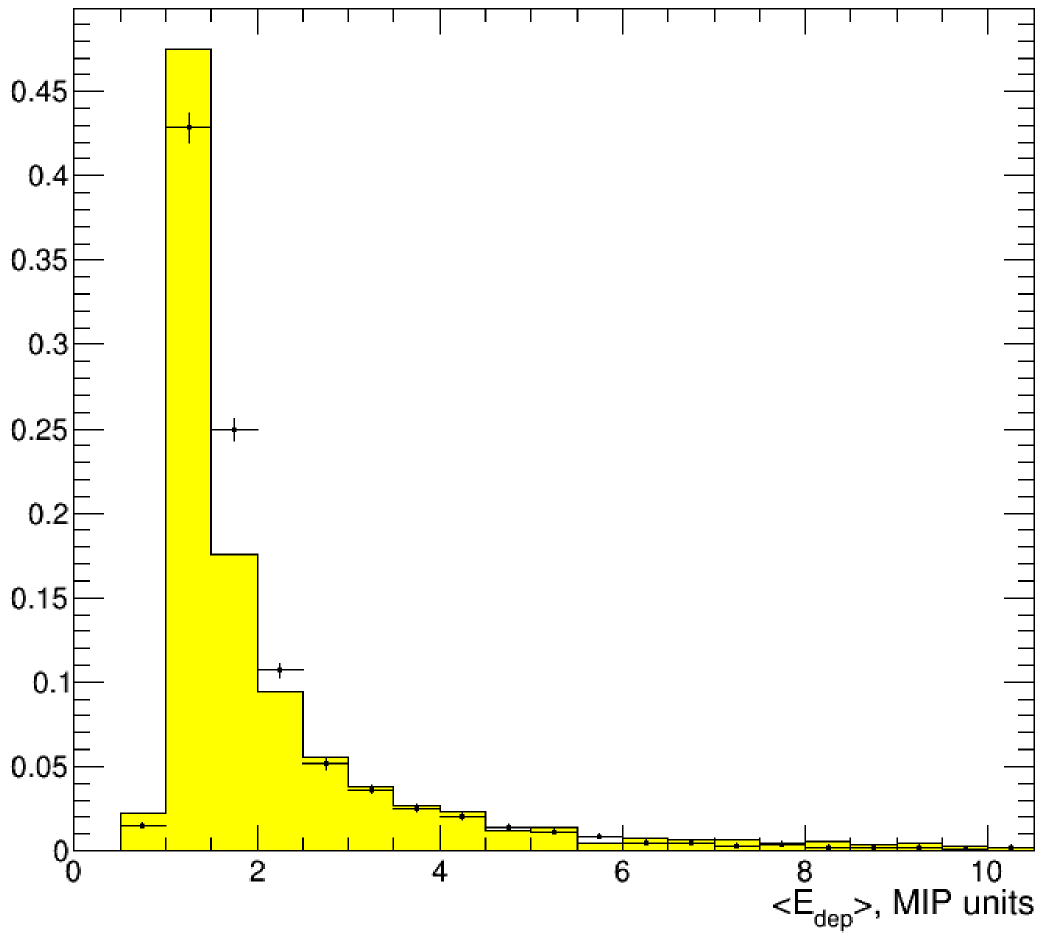
The interaction zone is removed by the following topological criteria: if some pad has more than 6 active neighbor pads, the considered pad and its neighbors will be marked as ‘shower pads’ and ignored by further analysis

Comparison with MC truth



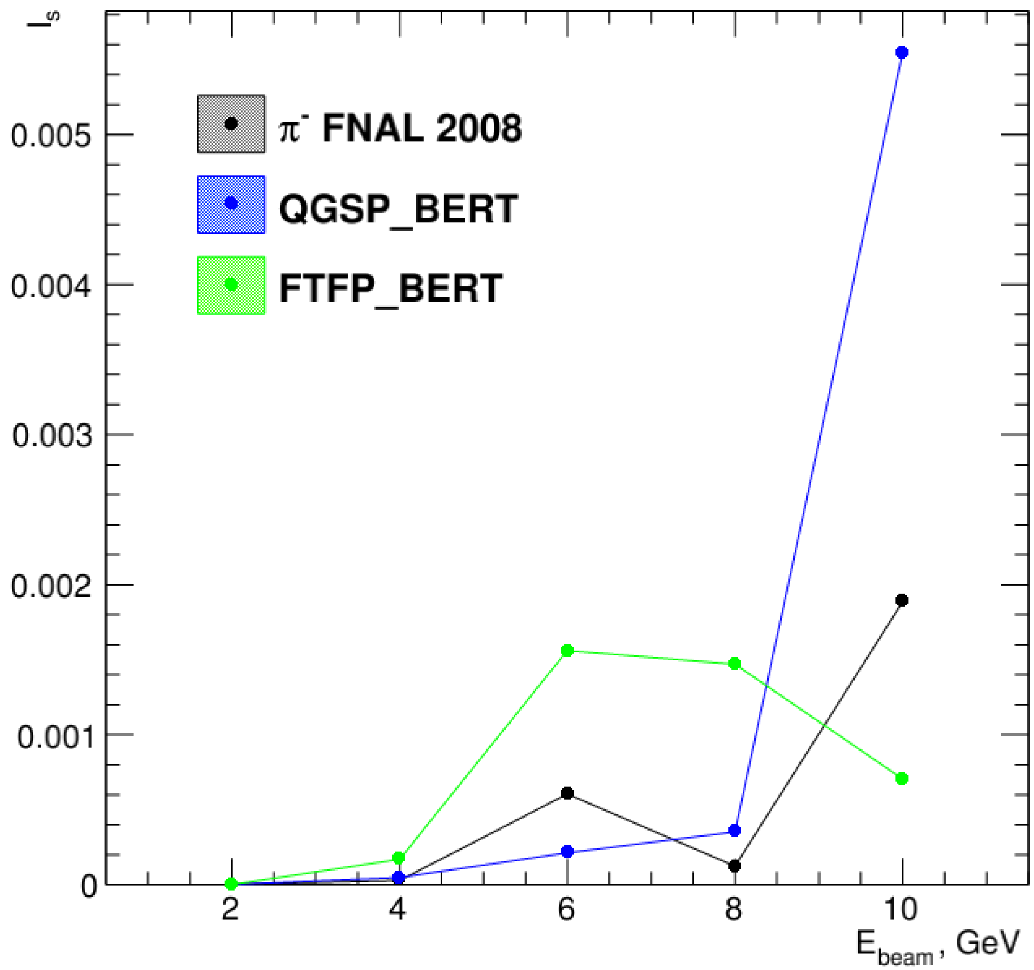
The histogram of event-per-event difference between number of MC particles and the number of tracks without a shower identified

π^- and secondaries comparison



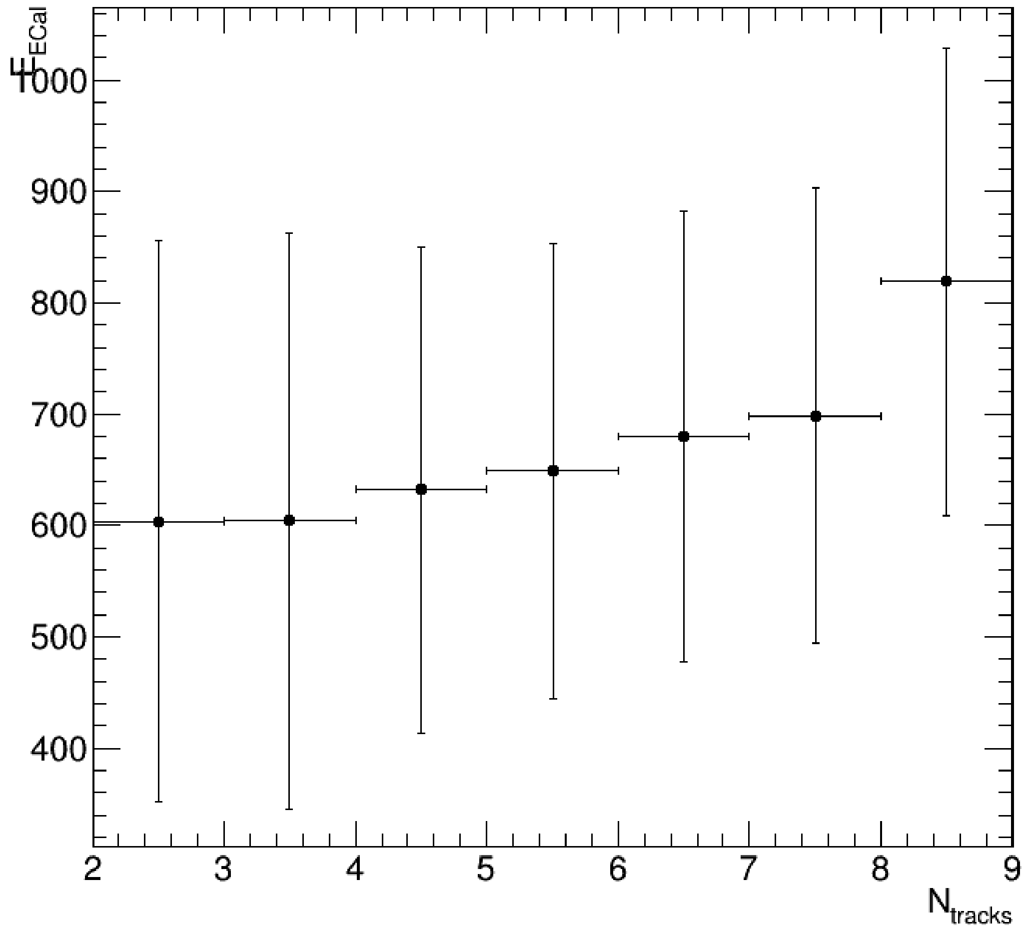
The histogram of the mean energy deposition per pad in MIP units for π^- incoming particles (yellow) and outgoing particles (dots) for 10 GeV energy. These spectra almost coincide -> most of outgoing particles are $\pi^{\pm}$.

Graph of ϕ isotropy



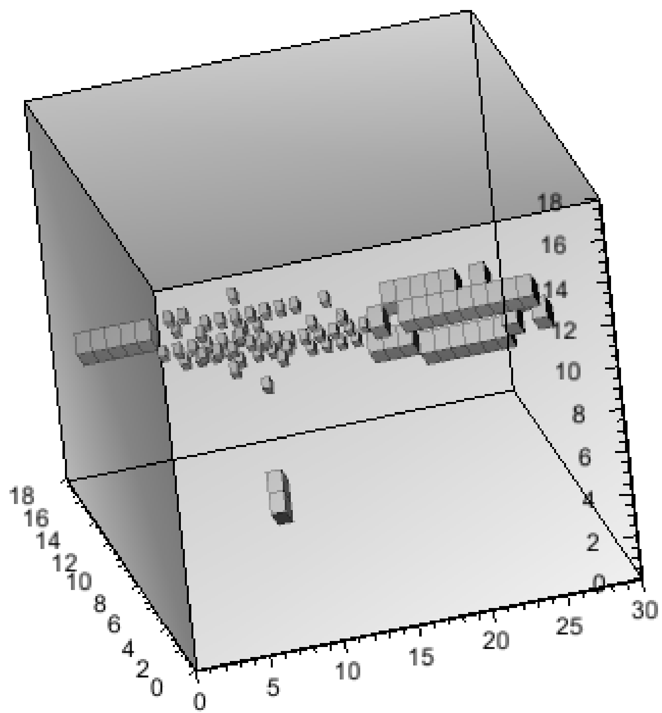
Graph of 'ϕ isotropy' $I_s = |\sum_i \frac{dv_i}{d\phi}|$, where v_i is a value of i bin of ϕ angle histogram.

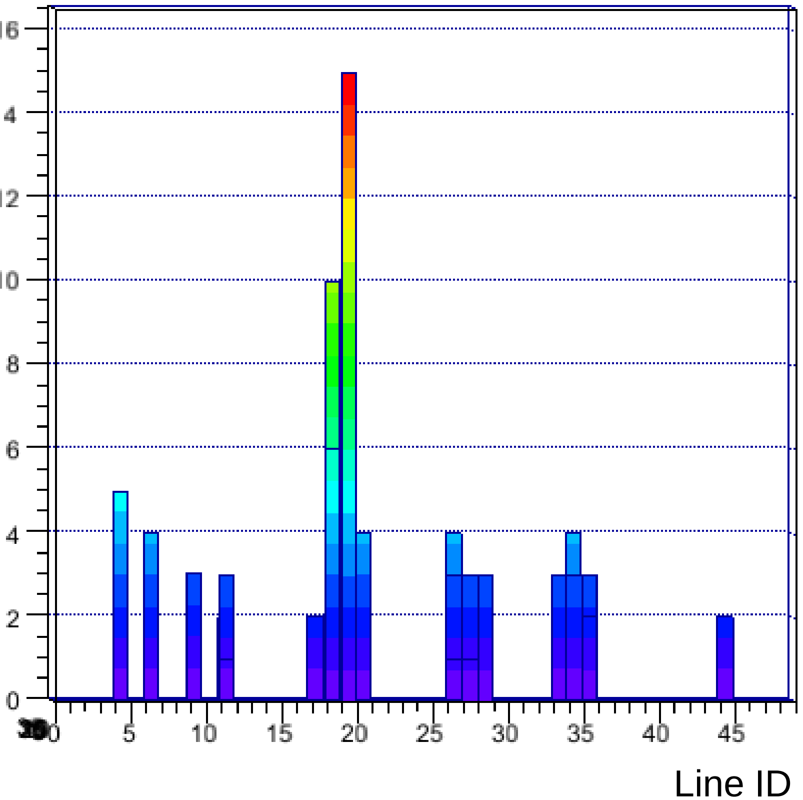
Energy/Tracks dependence



Total energy deposition (MIP) as function of number of tracks (10 GeV pion data at Fermilab, 2008). Error bars are a standard deviation of a Gaussian fit for each bin. Mean values are increasing with track number, but error bars are too large

3D Hough transform





This algorithm uses 49 precalculated lines per pad. It can single out the interaction zone to get rid of most of the noise.