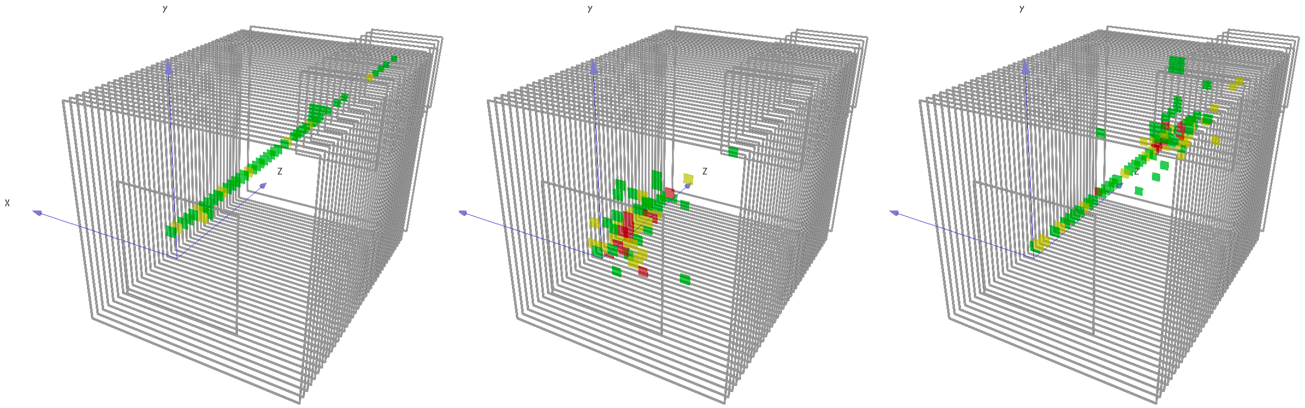
Particle Identification in the 2018 AHCAL test beam data

CALICE Collaboration Meeting



Vladimir Bocharnikov (DESY),
Utrecht, 10-12 April 2019

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

CALICE

P.N.Lebedev Physical Institute of the Russian Academy of Science

NATIONAL RESEARCH NUCLEAR UNIVERSITY NRNU MEPhI MEPhI

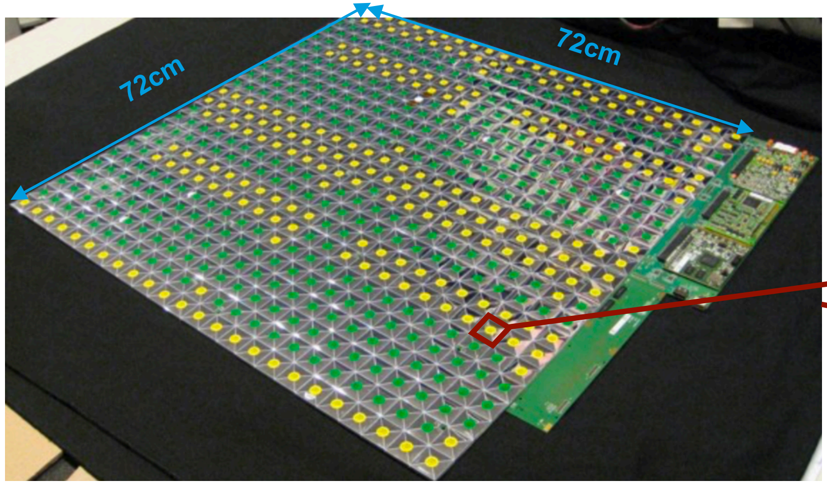
DESY.

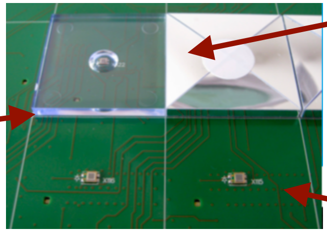
Outline

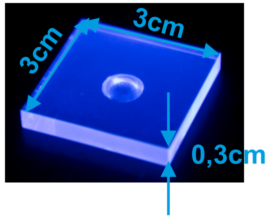
- Test beam **particle ID**
  - Introduction
  - Cut-based method
  - BDT method
- **Event filtering** studies

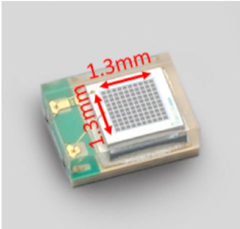
The Analog Hadron Calorimeter (AHCAL)

The AHCAL Technological Prototype 2018



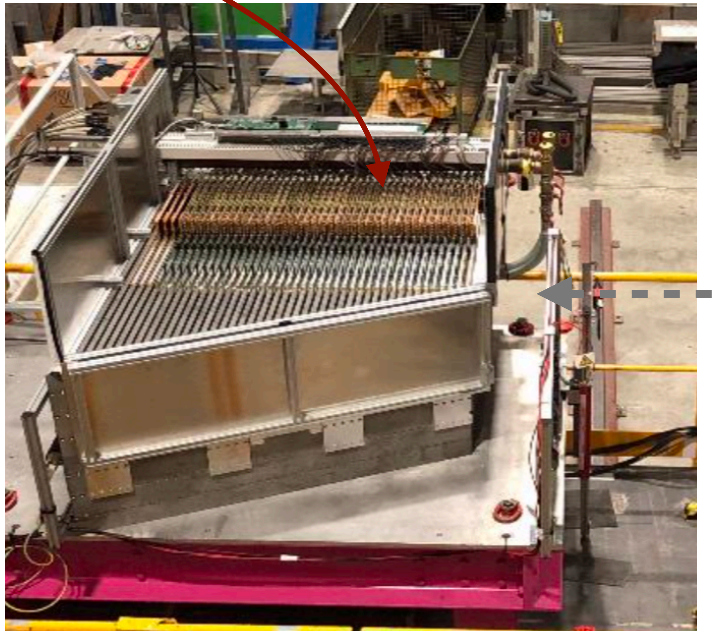






AHCAL Technological Prototype 2018:
38 layers (of 24x24 tiles each) alternating with 2cm steel absorber plates.
In total: ~22000 channels, ~4λ.

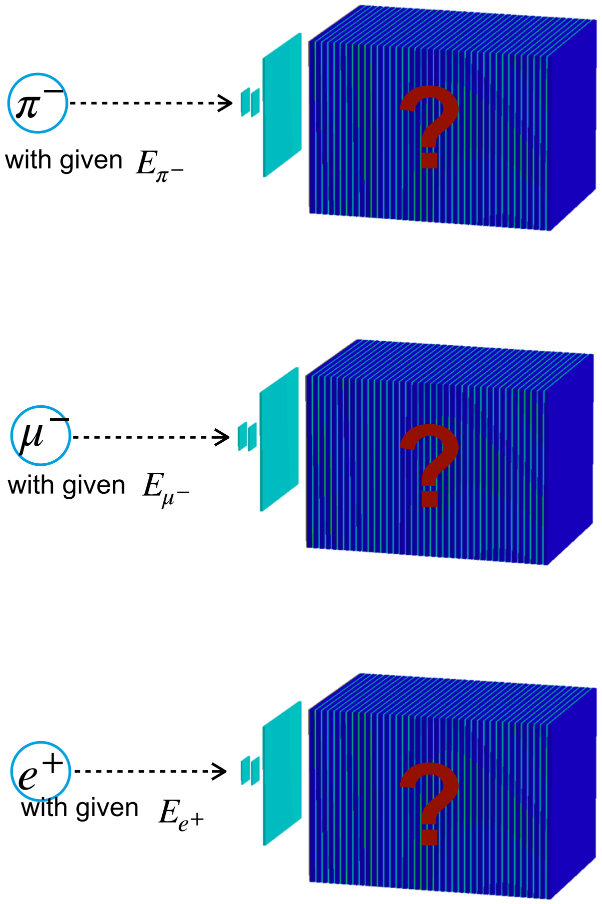
CERN SPS test beams: in May, June and October 2018 prototype was tested with the e/μ/π beams



beam line

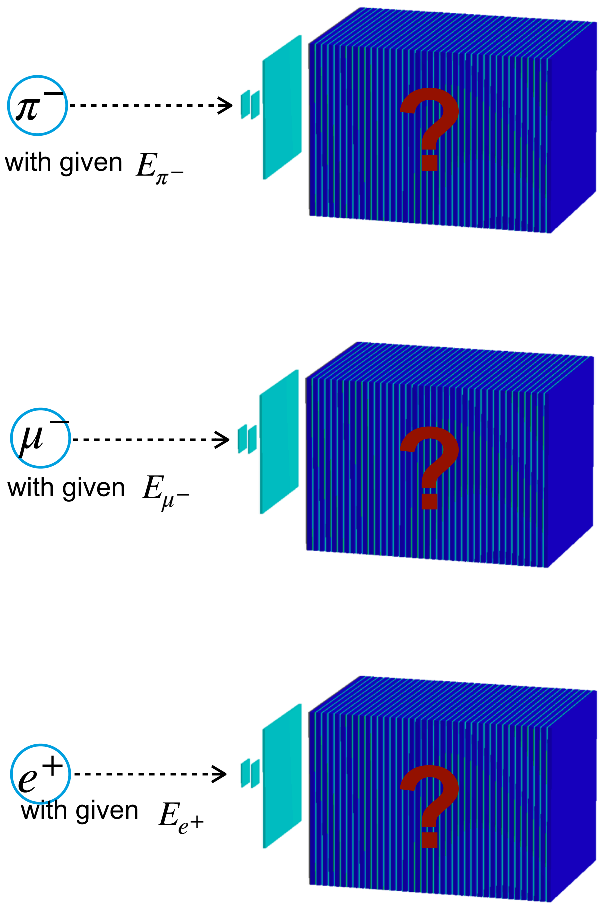
Test beam studies

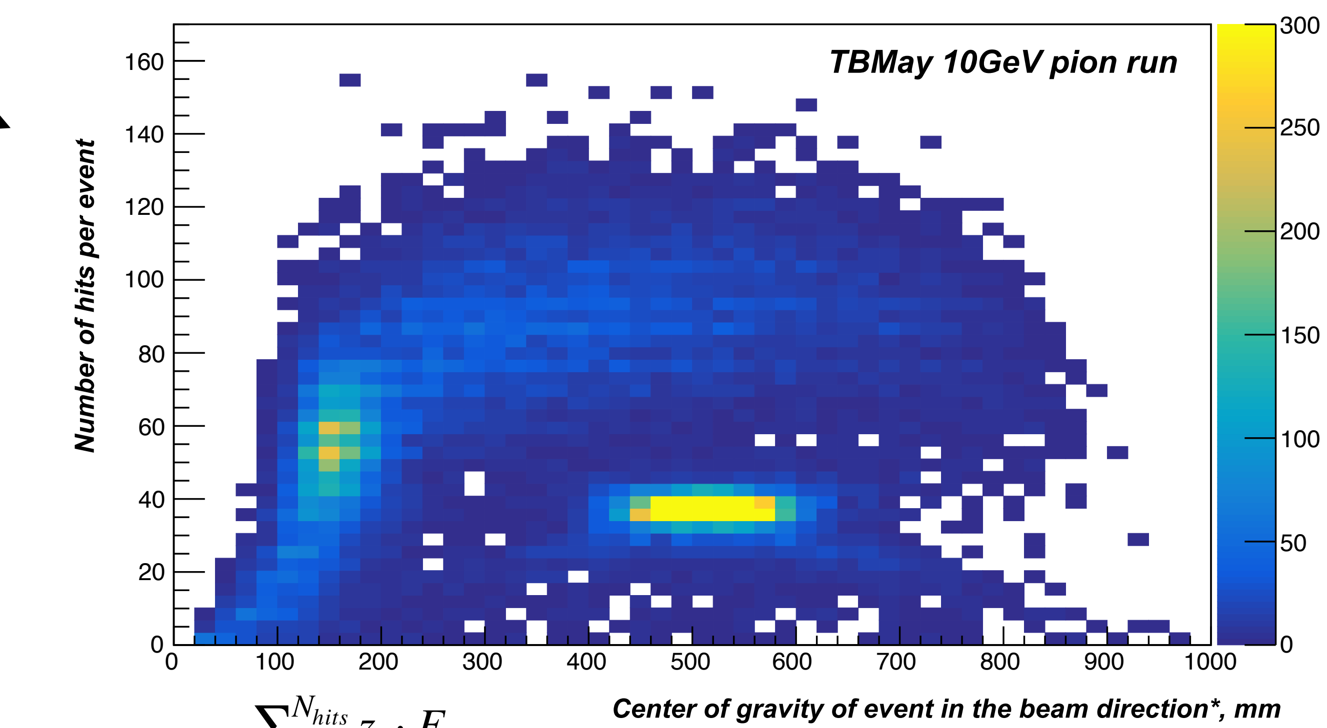
Motivation for particle ID



Test beam studies

Motivation for particle ID

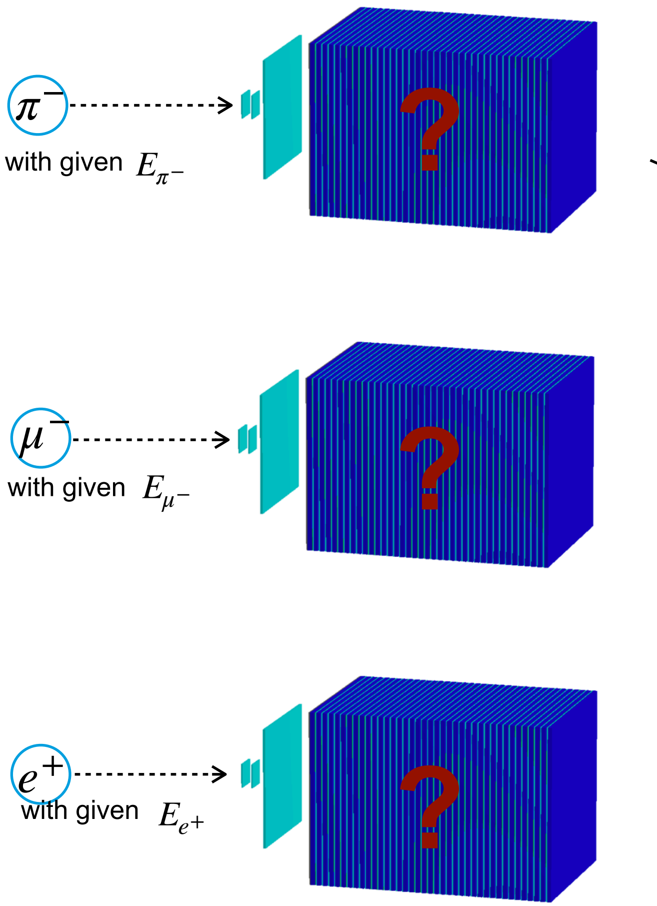


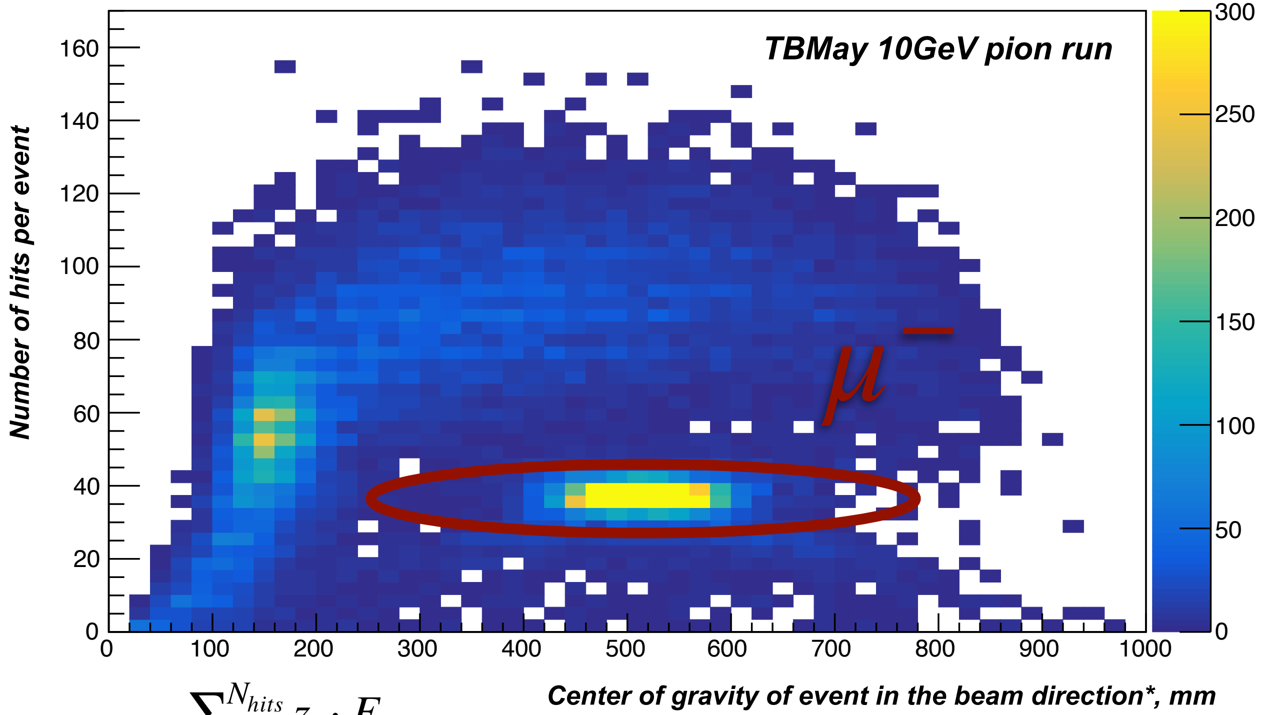


$$* \, z_{CoG} = \frac{\sum_{i=1}^{N_{hits}} z_i \cdot E_i}{E_{sum}}$$

Test beam studies

Motivation for particle ID

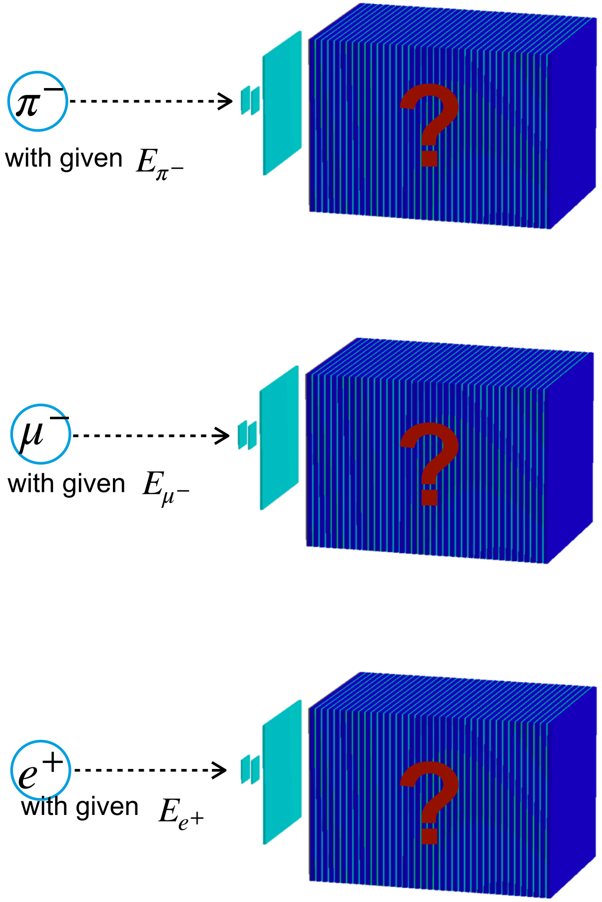


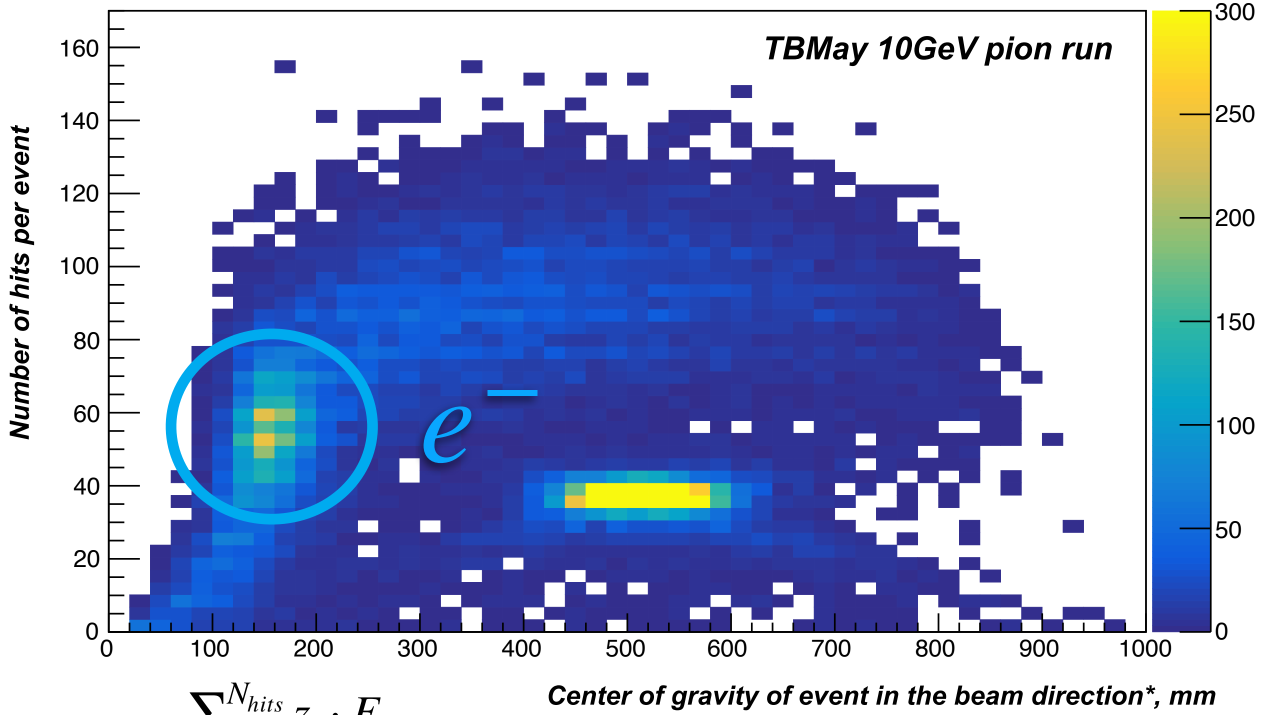


$$* \, z_{CoG} = \frac{\sum_{i=1}^{N_{hits}} z_i \cdot E_i}{E_{sum}}$$

Test beam studies

Motivation for particle ID

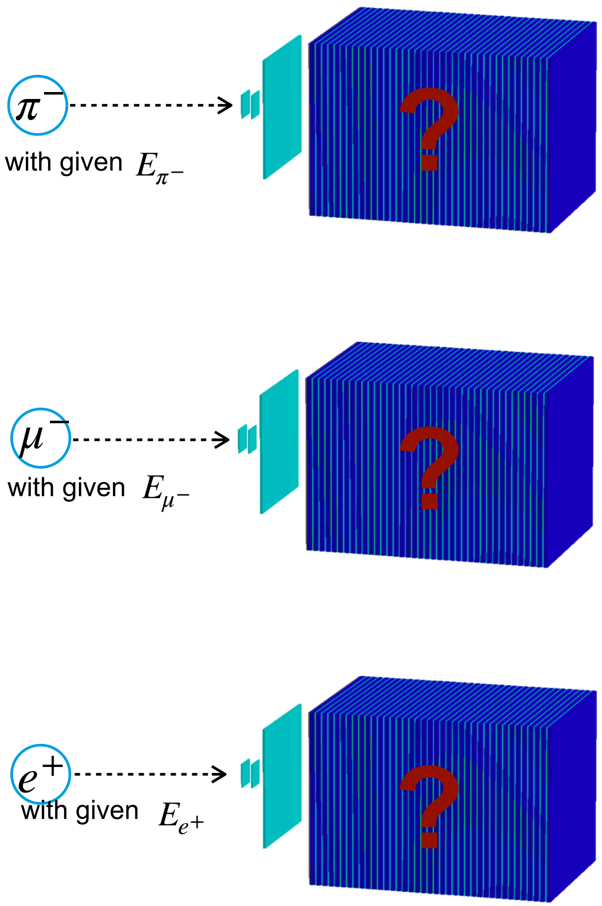


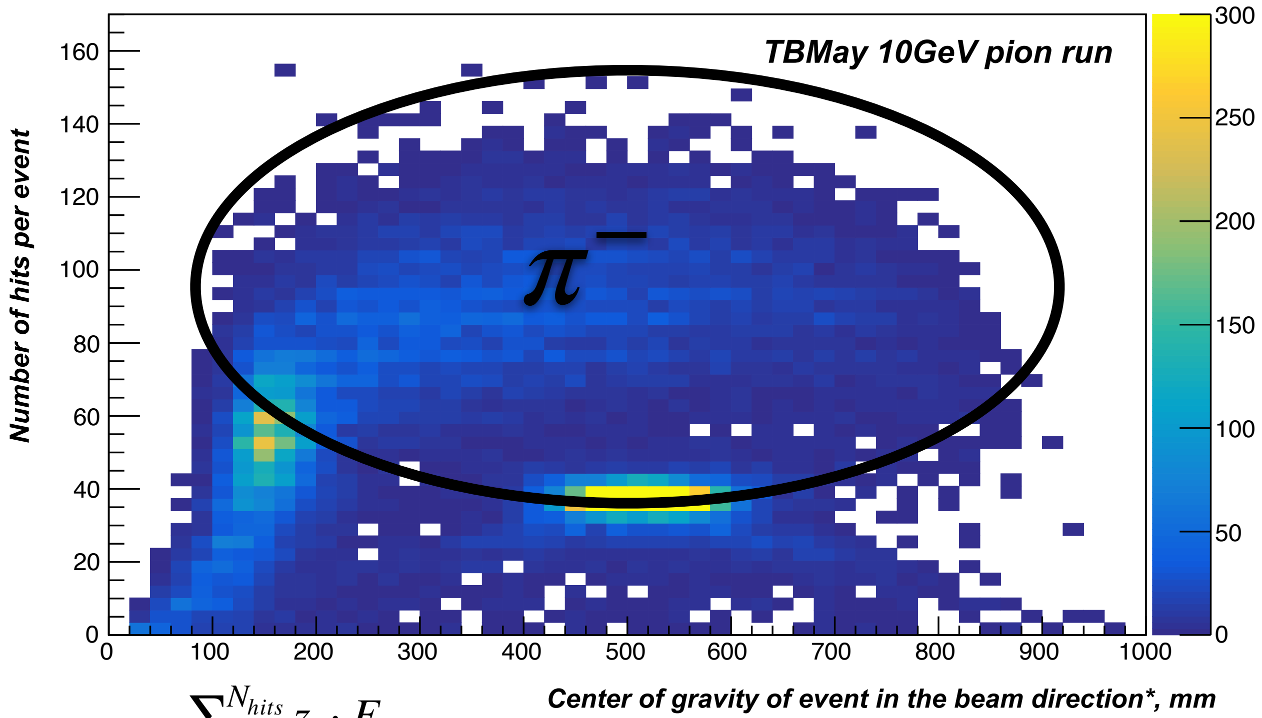


$$* \, z_{CoG} = \frac{\sum_{i=1}^{N_{hits}} z_i \cdot E_i}{E_{sum}}$$

Test beam studies

Motivation for particle ID

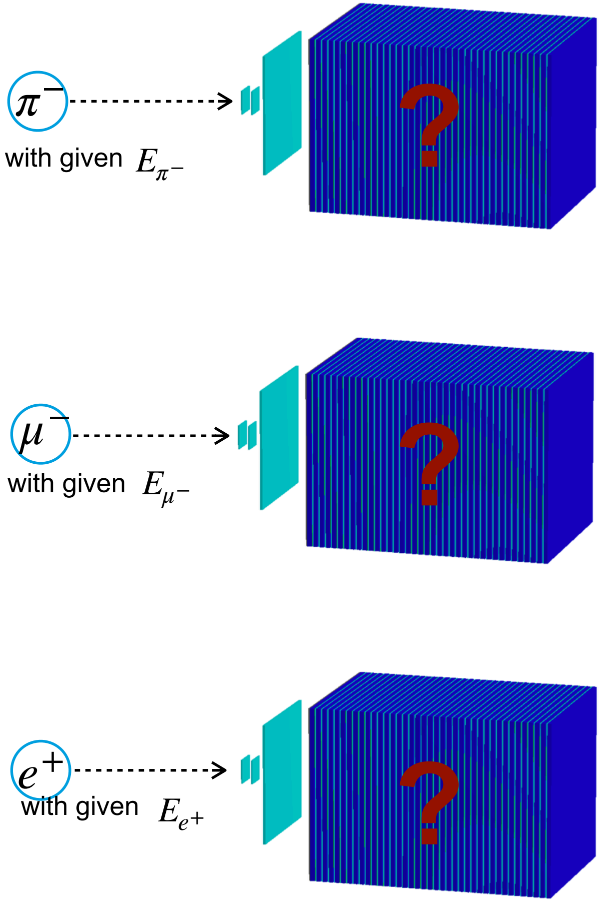




$$* \, z_{CoG} = \frac{\sum_{i=1}^{N_{hits}} z_i \cdot E_i}{E_{sum}}$$

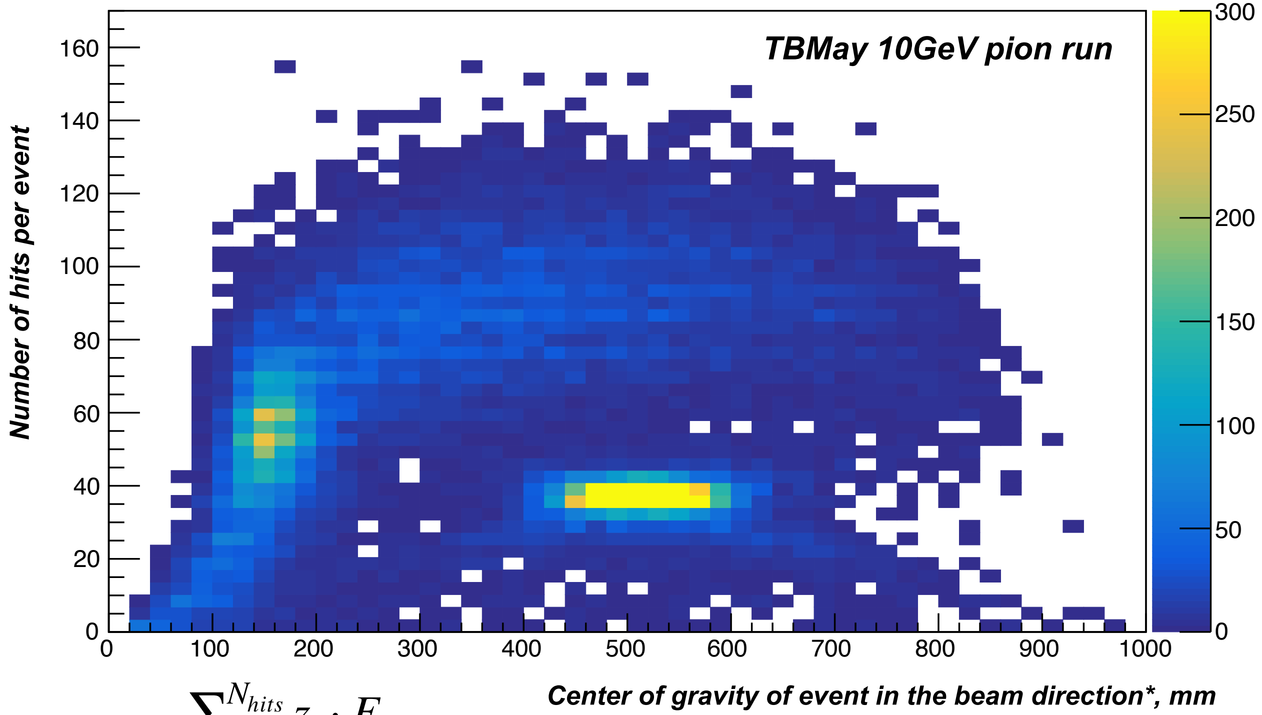
Test beam studies

Motivation for particle ID



We always have a deal with contamination by other particles.

⇒To investigate detector response to particles of given type we need to perform particle identification

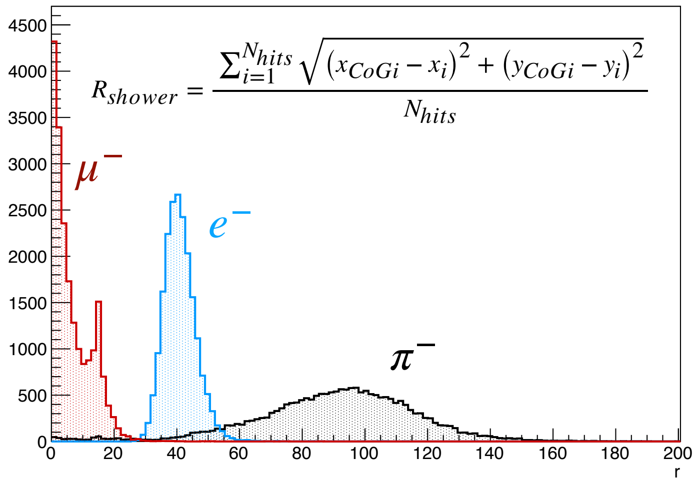


$$* \, z_{CoG} = \frac{\sum_{i=1}^{N_{hits}} z_i \cdot E_i}{E_{sum}}$$

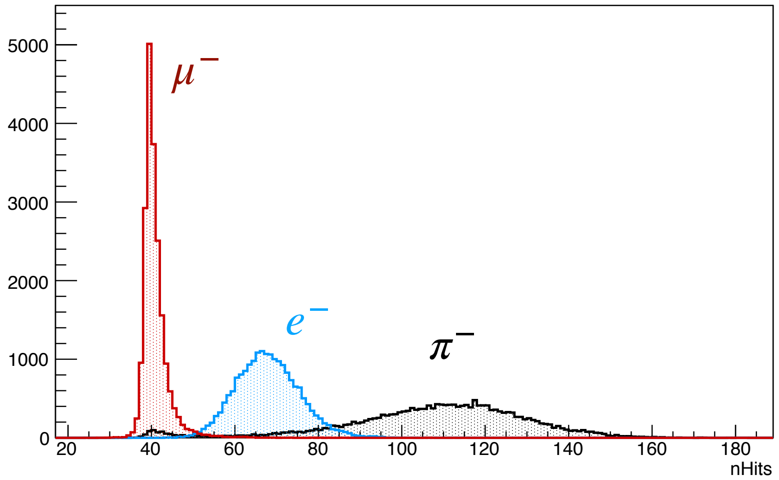
Observables

For particle ID

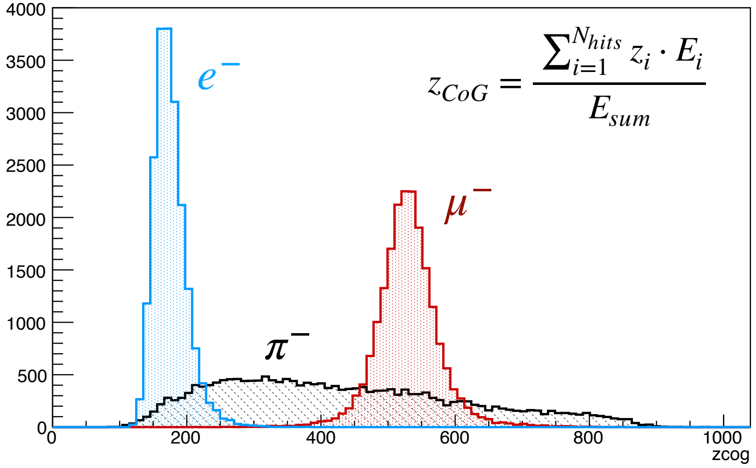
Shower radius, [mm]



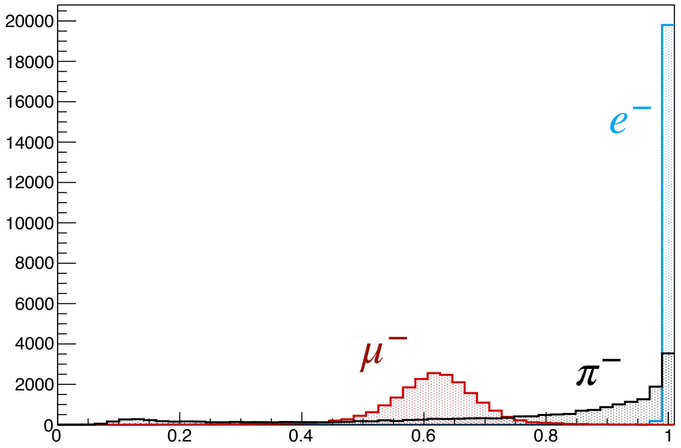
Number of hits per event



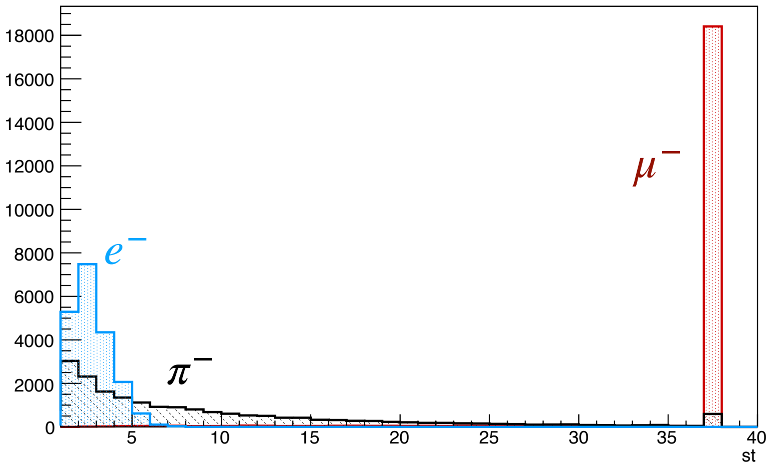
Center of gravity in z, [mm]



Fraction of energy in 1st 25 layers



Shower start layer



10GeV MC pions
20000 events

10GeV MC muons
20000 events

10GeV MC electrons
20000 events

Cut-based method

For particle ID (10 GeV particles)

Bad quality events:

- **Hits**: nHits < nHits_min
- **multi-particle** and **early shower** event filtering will be discussed in the second part of the talk (**Not applied** in following results)

Electron events:

- **Hits**: 45 < nHits < 95
- **Shower start layer:** st < 10
- **Shower radius:** 0 < r < 65 [mm]
- **COGz:** zcog < 400 [mm]
- **Fraction in first 25 layers:** frac25 > 0.9

Muon (muon-like) events:

- **Not an electron event**
- **Hits**: 0 < nHits < 70
- **Shower radius:** 0 < R < 30 [mm]
- **COGz:** 260 < zcog < 800 [mm]
- **Fraction in first 25 layers:** frac25 < 0.95

Remaining events are classified as **hadron events.**

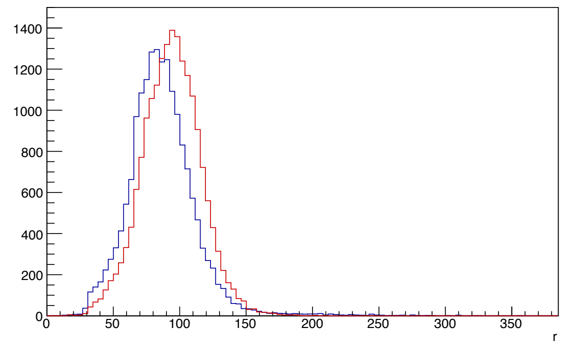
Simulations vs. data

Hadron events. Work in progress…

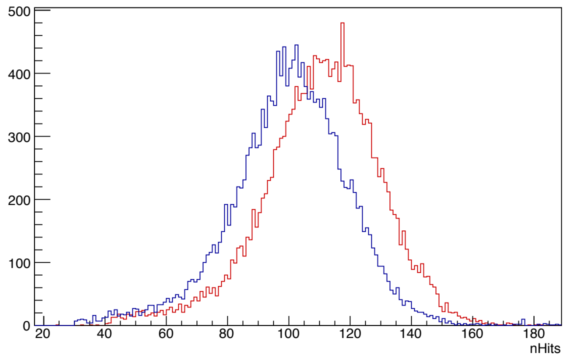
MC data sets simulated and fully reconstructed according to test beam setups using GEANT4 routine (QGSP_BERT_HP physics list)

Used for **cut adjustment** and **training/testing of BDT**

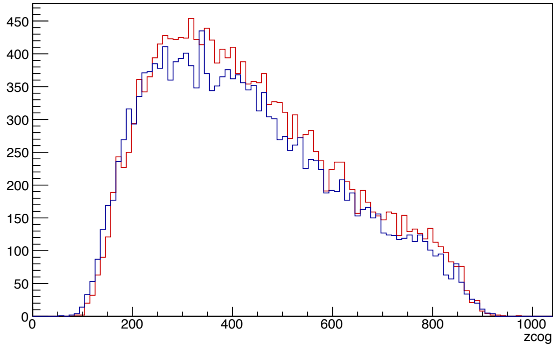
Shower radius, [mm]



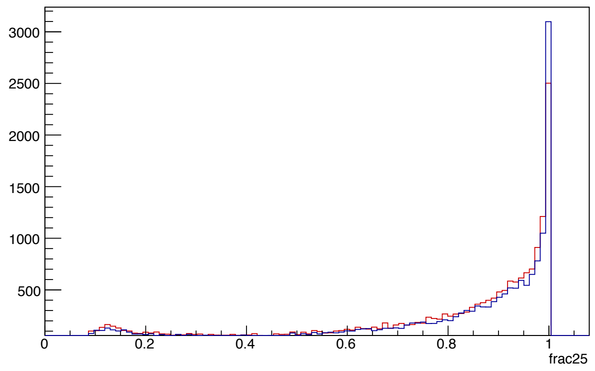
Number of hits per event



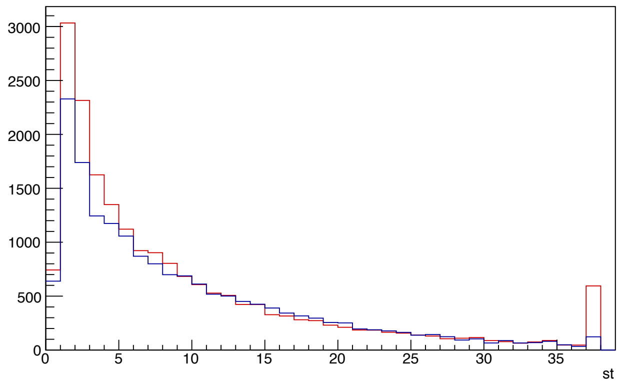
Center of gravity in z, [mm]



Fraction of energy in 1st 25 layers



Shower start layer



TBJune 10GeV **pion data** 17505 events

10GeV **MC pions** 18811 events

Same cuts to select hadron events

Performance

of cut-based ParticleID for 10GeV particles.

Input. ↓ Output →	Hadron events	Muon-like events	Electron events	Empty (nHits < 30)
MC 10GeV pions 20000 events	18811 events **94%**	536 events **2,7%**	653 events **3,3%**	Cut is not applied (0 events)
MC 10GeV muons 20000 events	56 events **0,3%**	19941 events **99,7%**	3 events **~0%**	Cut is not applied (0 events)
MC 10GeV electrons 20000 events	57 events **0,3%**	0 events **0%**	19943 events **99,7%**	Cut is not applied (1 event)

Performance

of cut-based ParticleID for 10GeV particles.

Input. ↓ Output →	Hadron events	Muon-like events	Electron events	Empty (nHits < 30)
MC 10GeV pions 20000 events	18811 events **94%**	536 events **2,7%**	653 events **3,3%**	**Cut is not applied (0 events)**
MC 10GeV muons 20000 events	56 events **0,3%**	19941 events **99,7%**	3 events **~0%**	**Cut is not applied (0 events)**
MC 10GeV electrons 20000 events	57 events **0,3%**	0 events **0%**	19943 events **99,7%**	**Cut is not applied (1 event)**
Data:				
10GeV pion run TBMay18 23000 events	7630 events **33,2%**	9824 events **42,7%**	3664 events **15,9%**	1887 events **8,2%**
10GeV pion run TBJune18 23000 events	17505 events **76,1%**	3685 events **16%**	807 events **3,5%**	1003 events **4,3%**

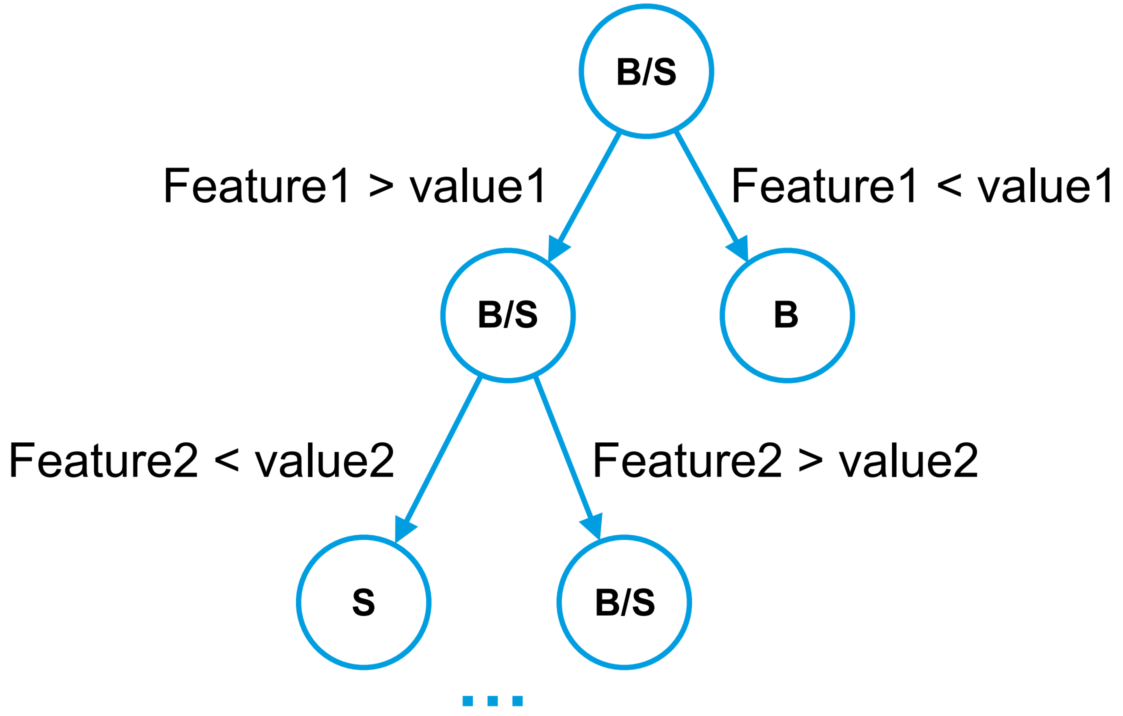
>12 cut values for signal/background tuning

Boosted decision tree (BDT)

Machine learning technique.

A **decision tree** takes a set of input features and splits input data recursively based on those features.

Training: Build and optimise the tree using set of known training events.



Boosted decision tree (BDT)

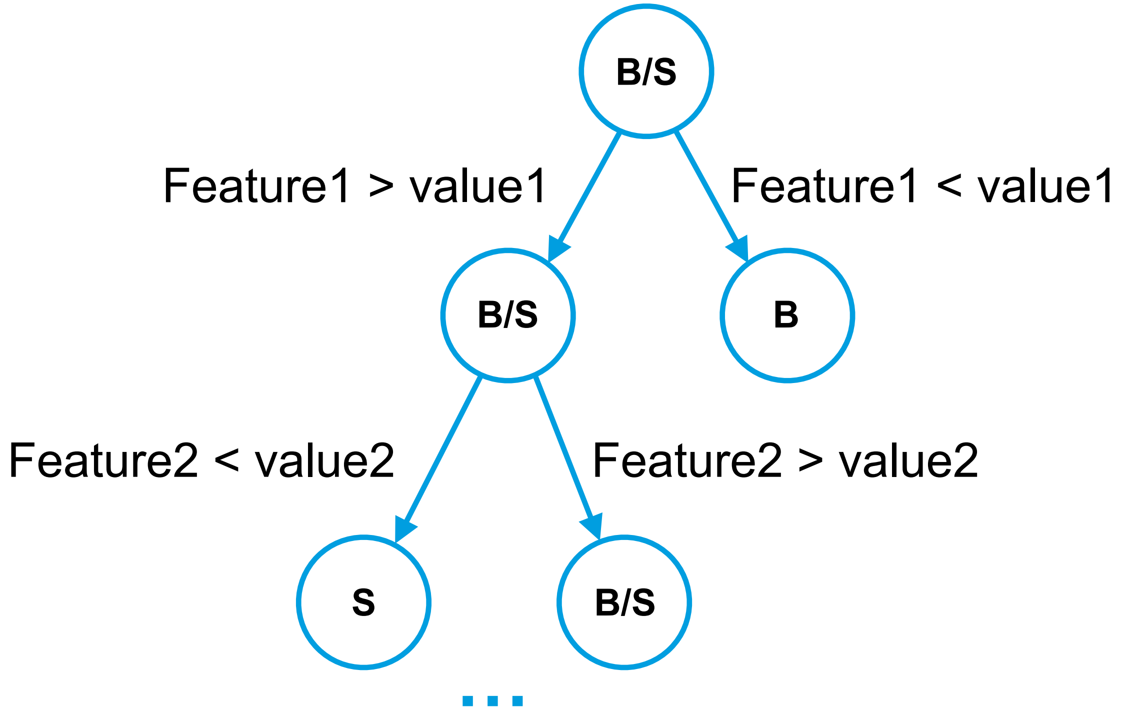
Machine learning technique.

A **decision tree** takes a set of input features and splits input data recursively based on those features.

Training: Build and optimise the tree using set of known training events.

Decision trees are powerful, but unstable. A small change in the training data can produce a large change in the tree. This is remedied by the use of **boosting**.

Boosting: Built many trees, so that the weighted average over all trees is insensitive to fluctuations.



BDT input

10GeV pion selection.

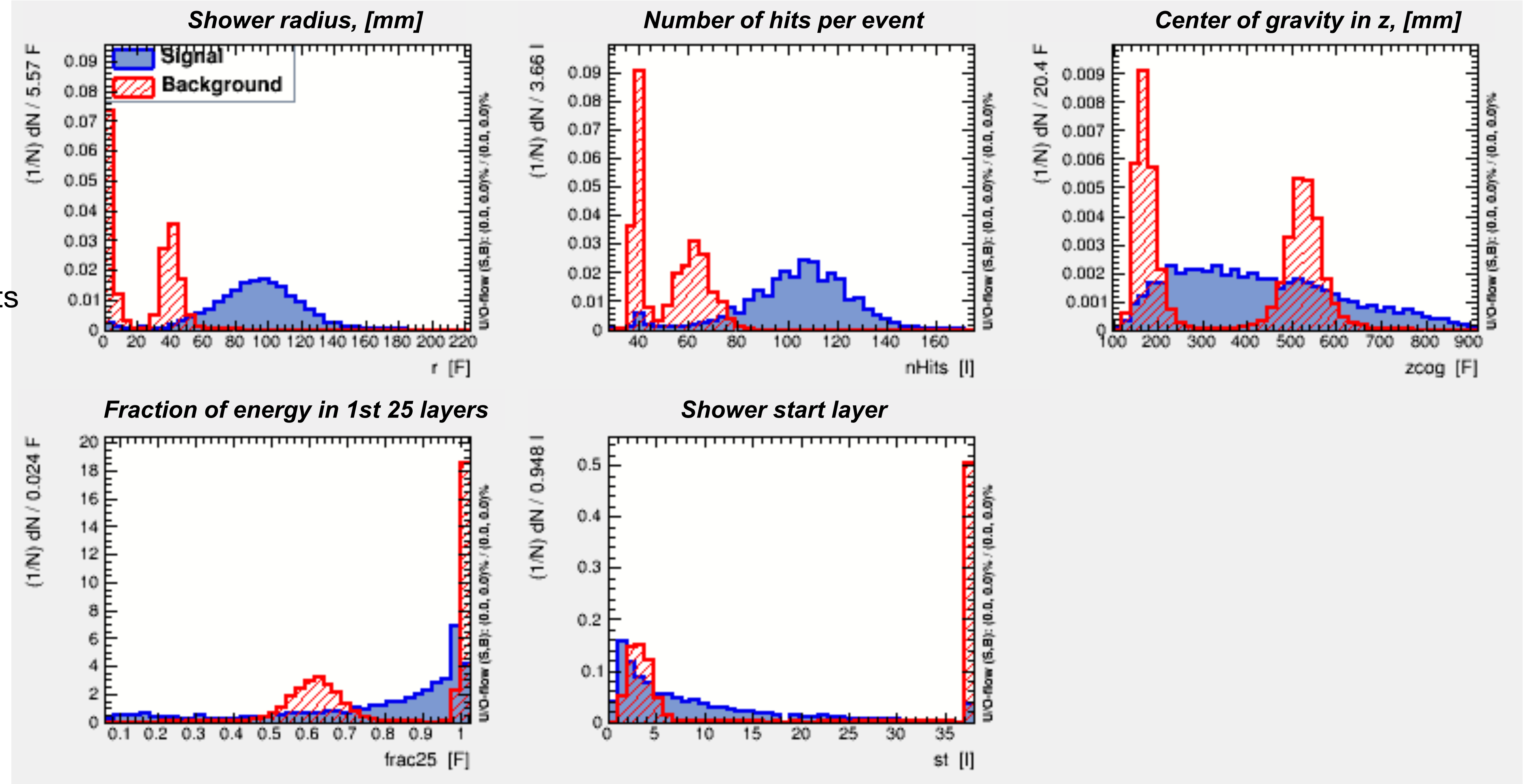
Training data:

- 10GeV MC **pions** - 20k events (**signal**)
- 10GeV MC **muons** - 20k events (**bckgr**)
- 10GeV MC **electrons** - 20k events (**bckgr**)

Variables:

- Shower radius
- Number of hits
- Center of gravity in z
- Fraction in first 25 layers
- Shower start layer

No cuts are applied



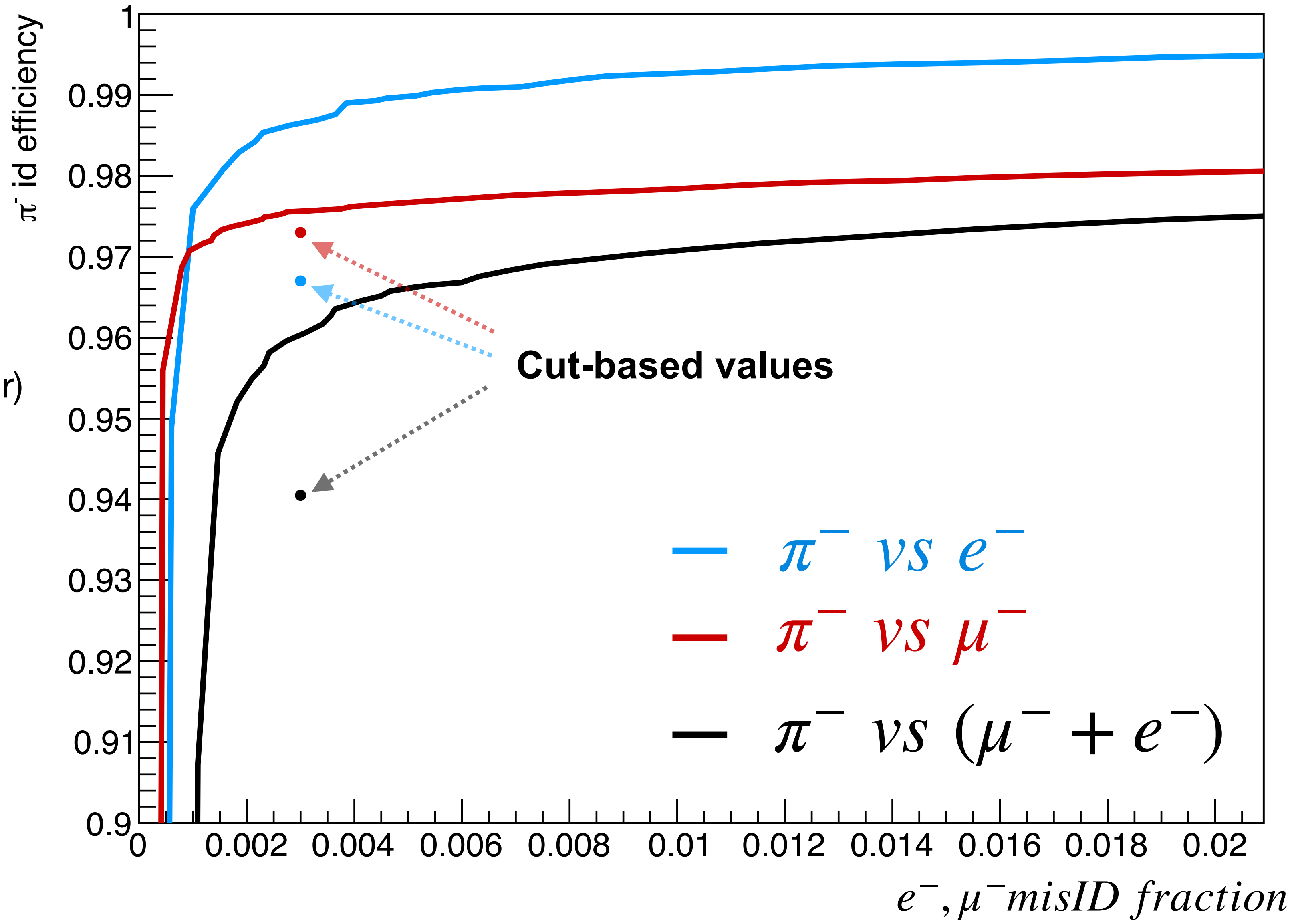
BDT output

10GeV pion selection.

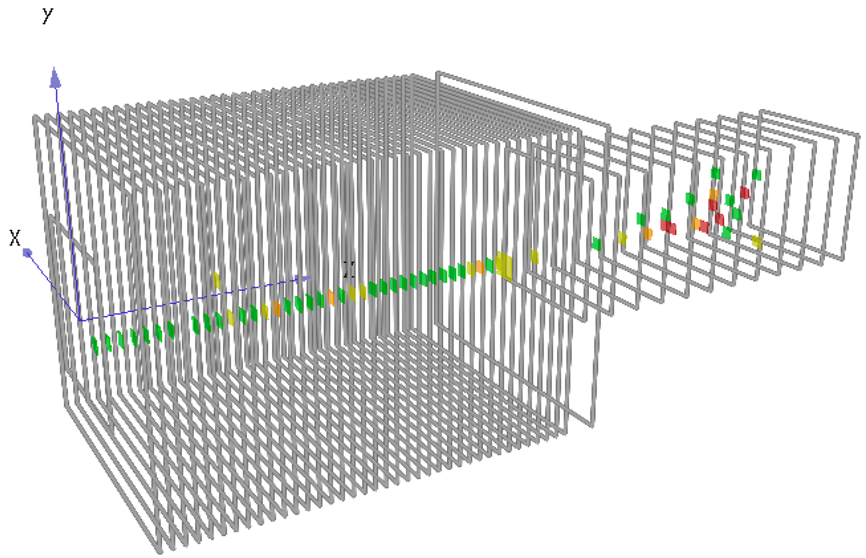
Pions vs. muons&electrons

Test data:

- 10GeV MC **pions** - 20k events (signal)
- 10GeV MC **muons** - 20k events (bckgr)
- 10GeV MC **electrons** - 20k events (bckgr)
- **No cuts**

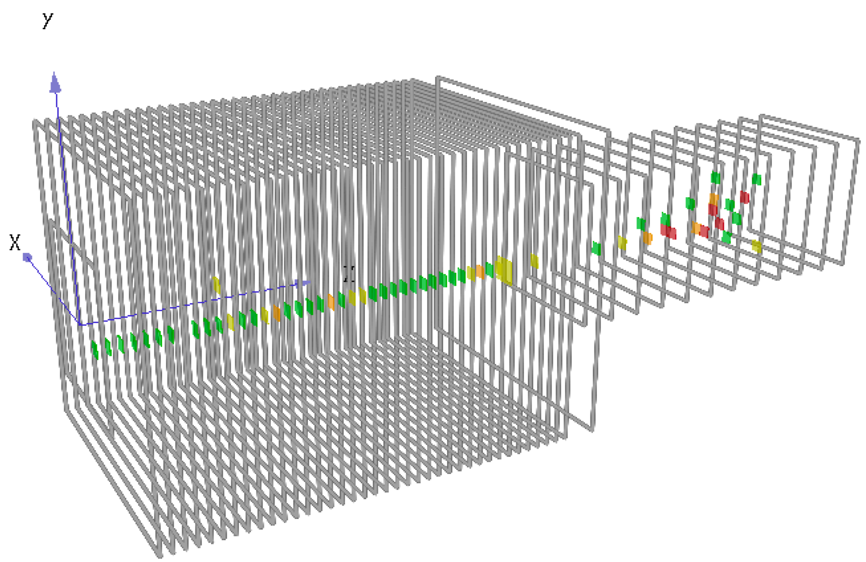


Punch-through rejection



Input. ↓ Output →	Hadron events	Muon-like events
MC 10GeV pions 20000 events	18811 events **94%**	! 536 events **2,7%**

Punch-through rejection

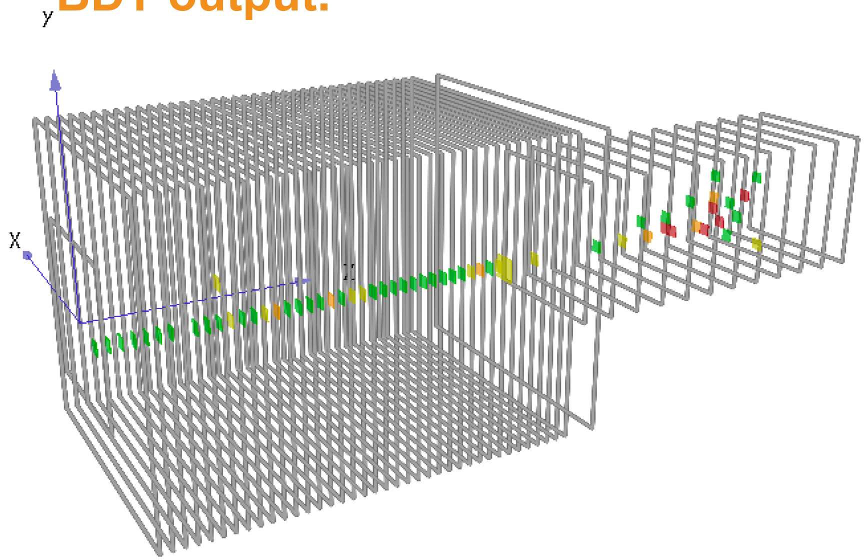


Showering pions vs. muon-like
Training/test data:

- 10GeV MC **pions** - 19,5k events (signal)
- **Cut:** Shower start < 38
- 10GeV MC **muons** - 20k events (bckgr)

Input. ↓ Output →	Hadron events	Muon-like events
MC 10GeV pions 20000 events	18811 events **94%**	! 536 events **2,7%**

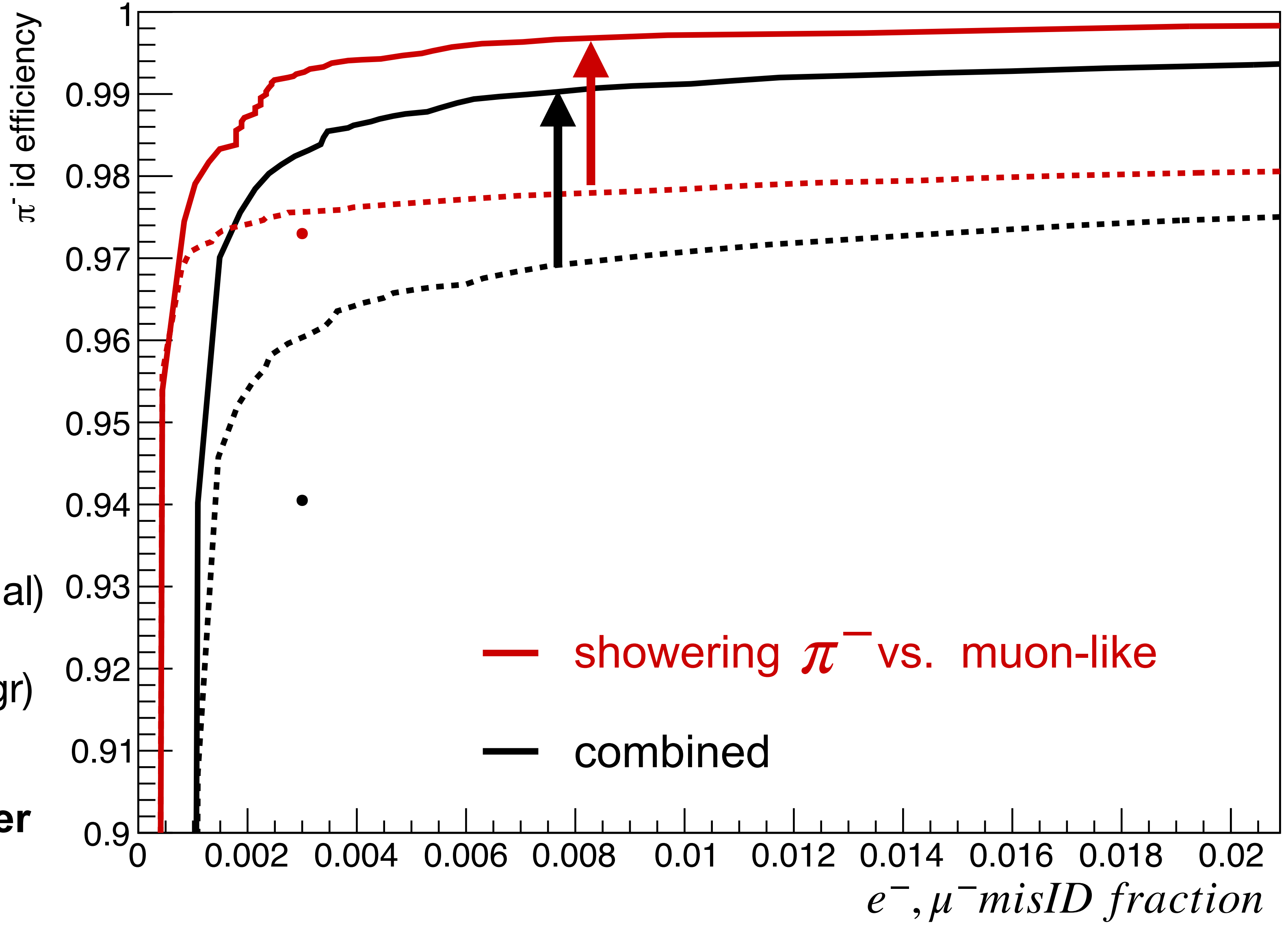
Punch-through rejection



Showering pions vs. muon-like
Training/test data:

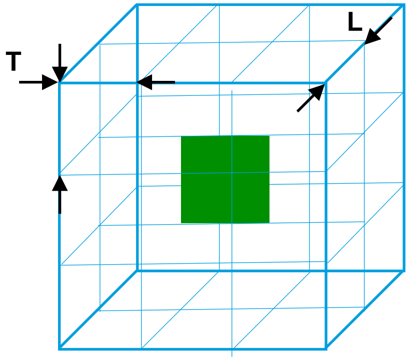
- 10GeV MC **pions** - 19,5k events (signal)
- **Cut:** Shower start < 38
- 10GeV MC **muons** - 20k events (bckgr)

Better performance than cut-based method can be reached. **One parameter** defines signal/background ratio



Event filtering for multi-particle and early showers rejection

Clustering



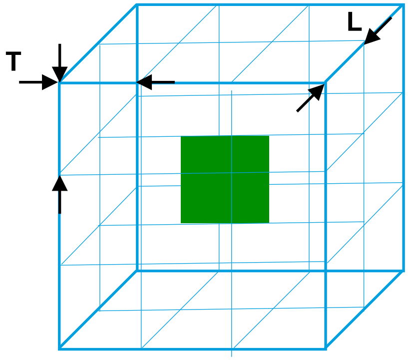
Algorithm: Hits are grouped in clusters if if they are neighbours in volume, {I,J,K}-space is used.

Steering parameters:

- Size of volume (T and L), *$T = 1$, $L = 2$*
- minimum nHits in cluster *nHits_min = 5*
- z_max of first hit in cluster (for filtering) *z_max = 100mm*

If nClusters > 1 => multi-particle event (or early shower)

Clustering



Algorithm: Hits are grouped in clusters if if they are neighbours in volume, {I,J,K}-space is used.

Steering parameters:
- Size of volume (T and L), **T = 1, L = 2*
- minimum nHits in cluster ** nHits_min = 5*
- z_max of first hit in cluster (for filtering) ** z_max = 100mm*

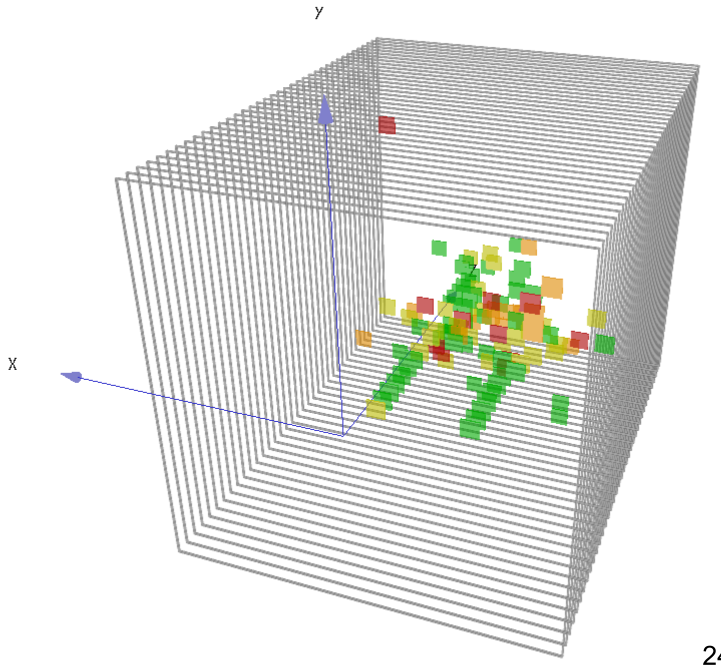
If nClusters > 1 => multi-particle event (or early shower)

Incoming MIP tracks

Construct towers with same I and J in first layers.

Steering parameters:
- Number of first layers **nPrimaryTrackLayers = 4*
- Minimum number of hits in track **MinHitsInPrimaryTrack = 3*

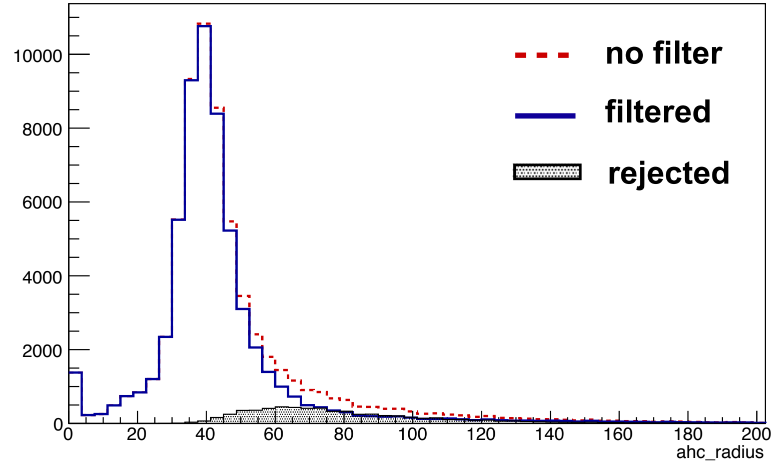
If nMIPTracks > 1 => multi-particle event

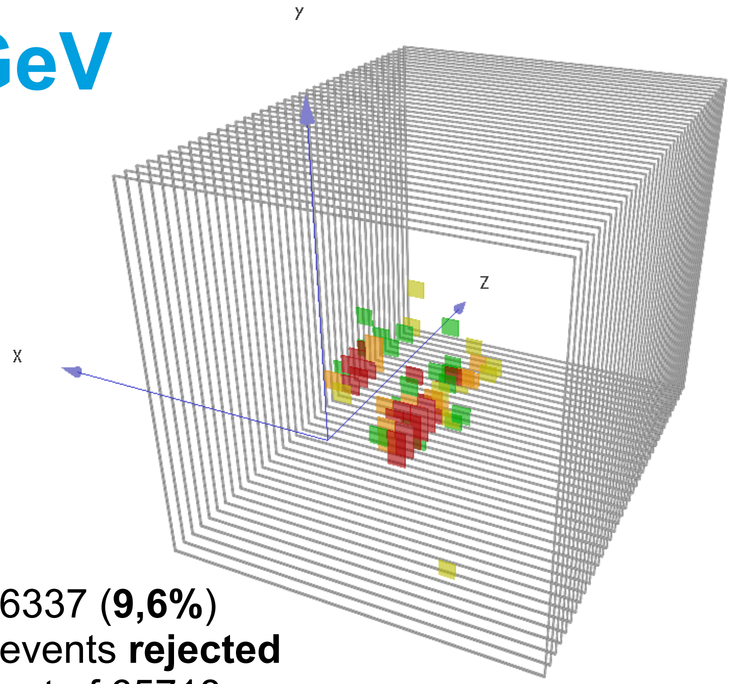


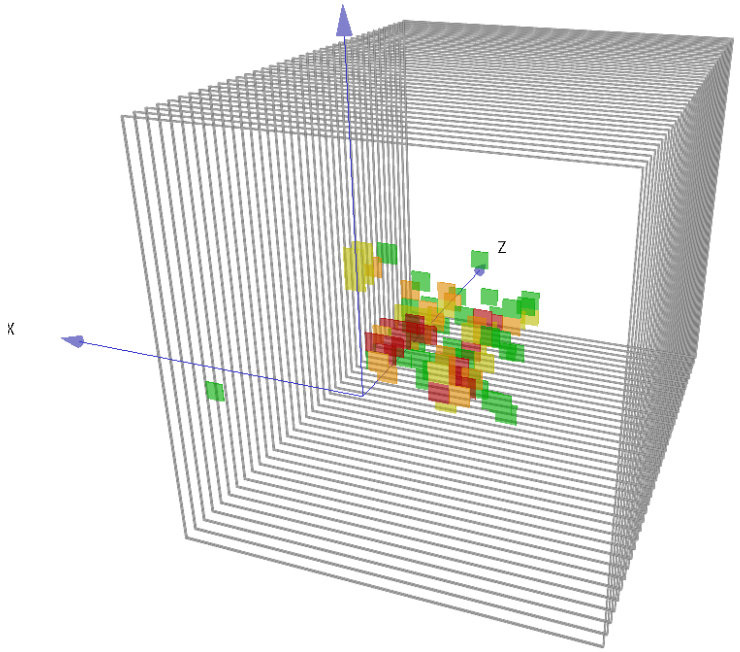
TBMay18 electron 10GeV

Apply filtering to data.

Shower radius, [mm]

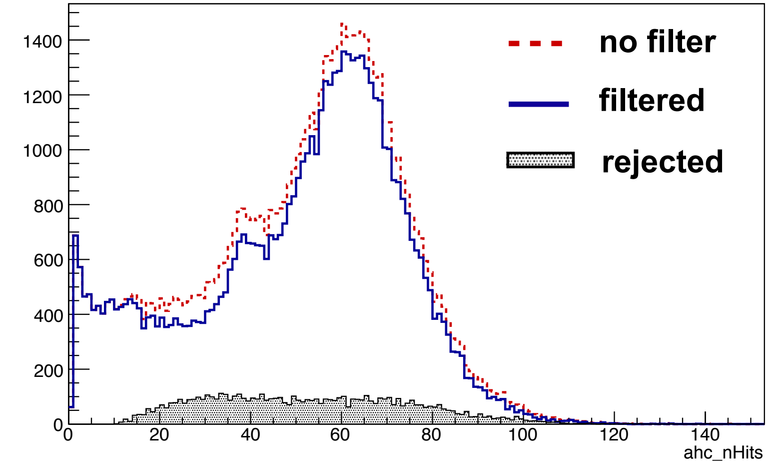




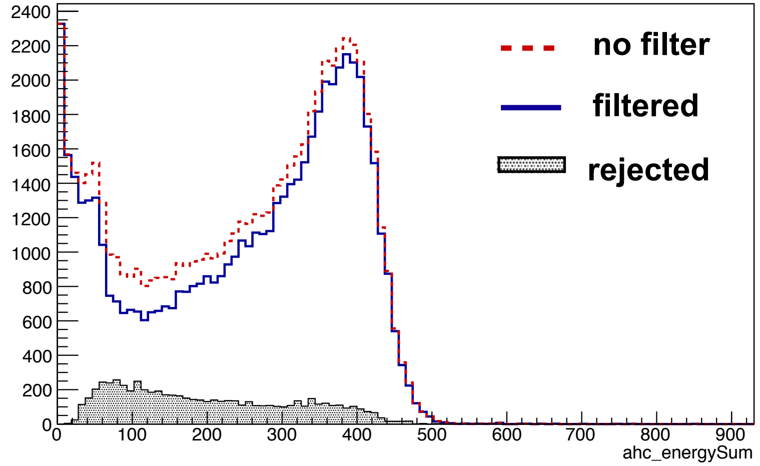


6337 (**9,6%**)
events **rejected**
out of 65718

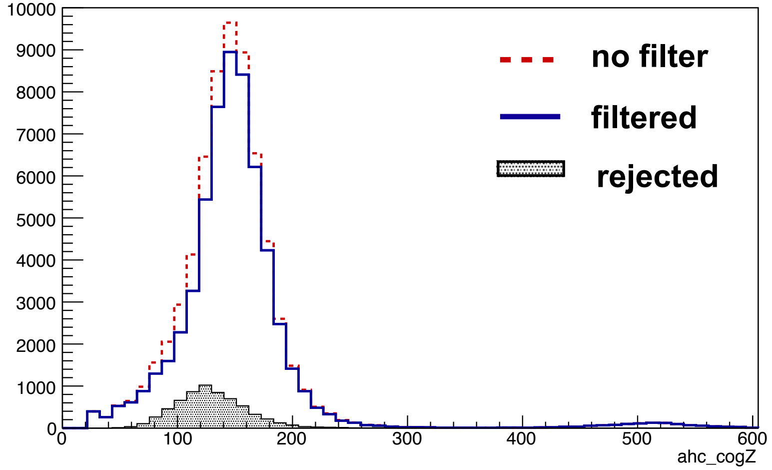
Number of hits



Energy sum, [MIP]



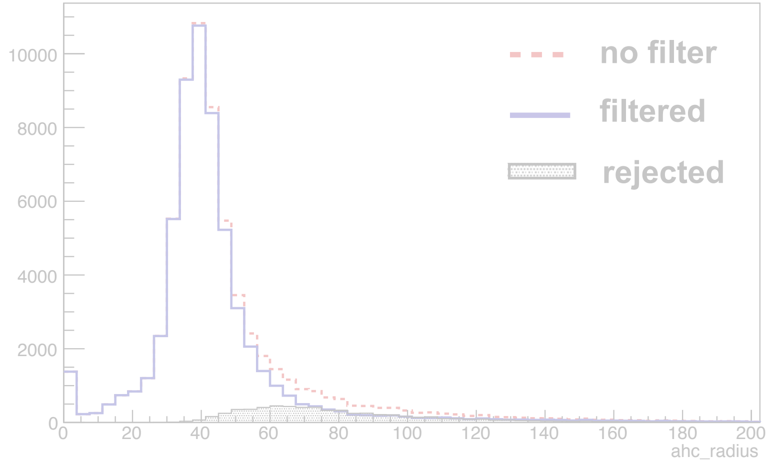
Center of gravity in z, [mm]

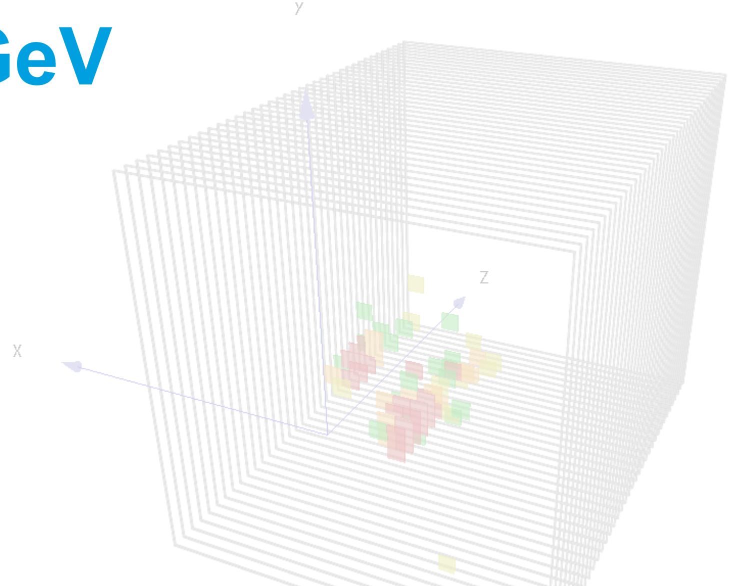


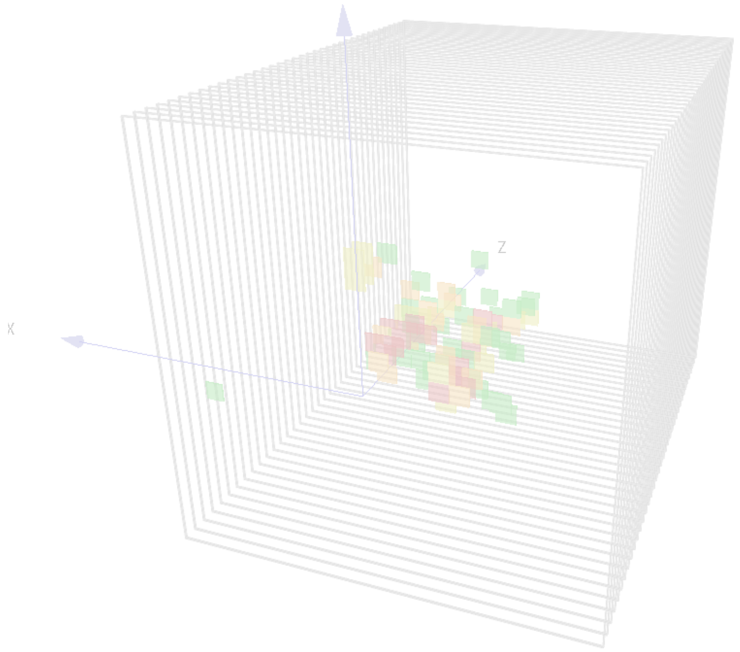
TBMay18 electron 10GeV

MC vs data

Shower radius, [mm]



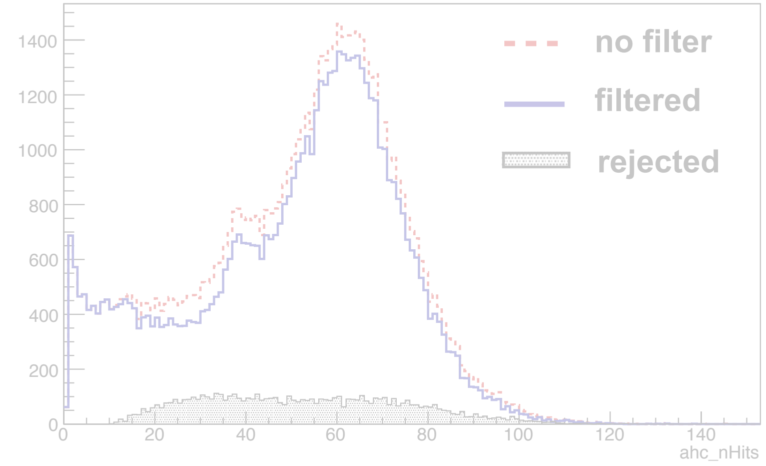




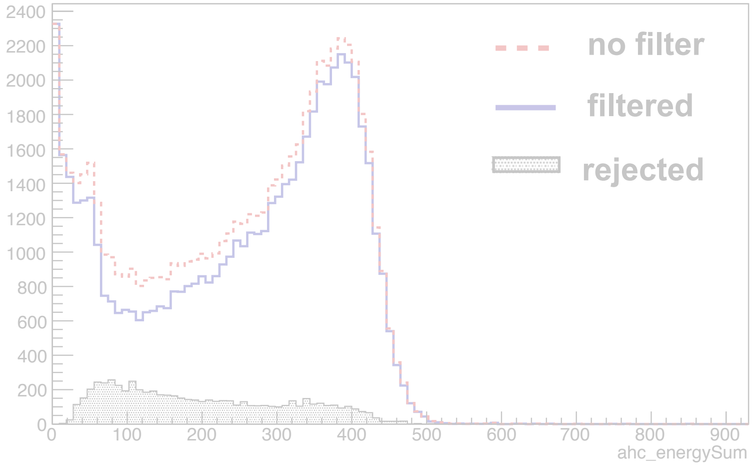
TBMay18 10GeV **MC electron:**

6 **(<<1%)** out of 10000 events

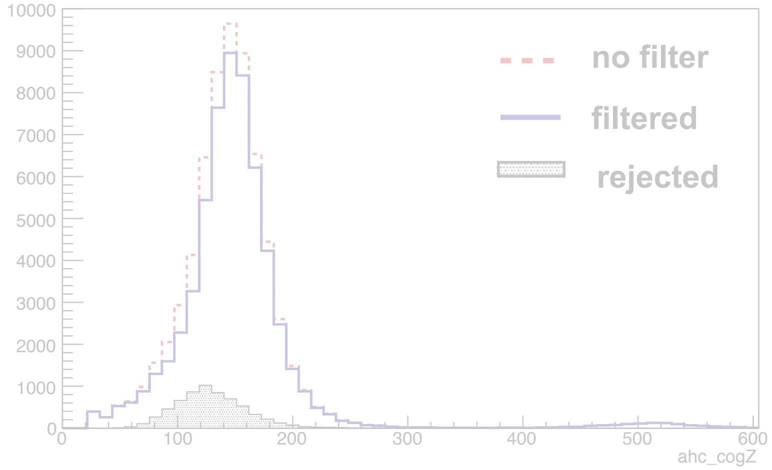
Number of hits



Energy sum, [MIP]



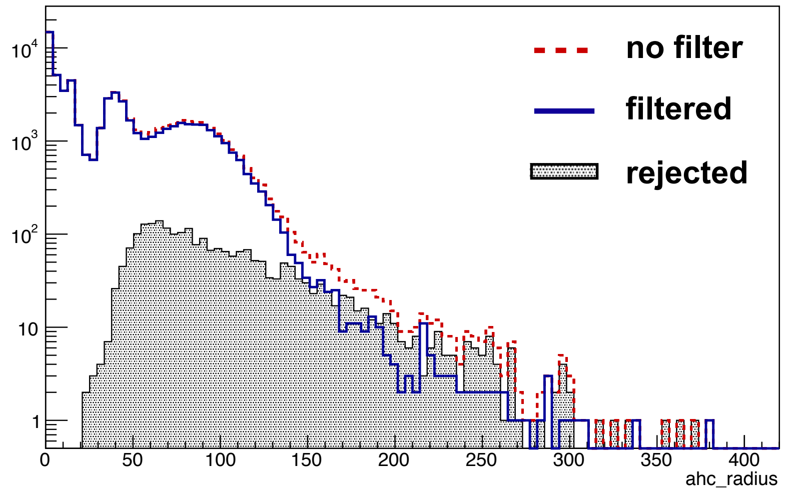
Center of gravity in z, [mm]

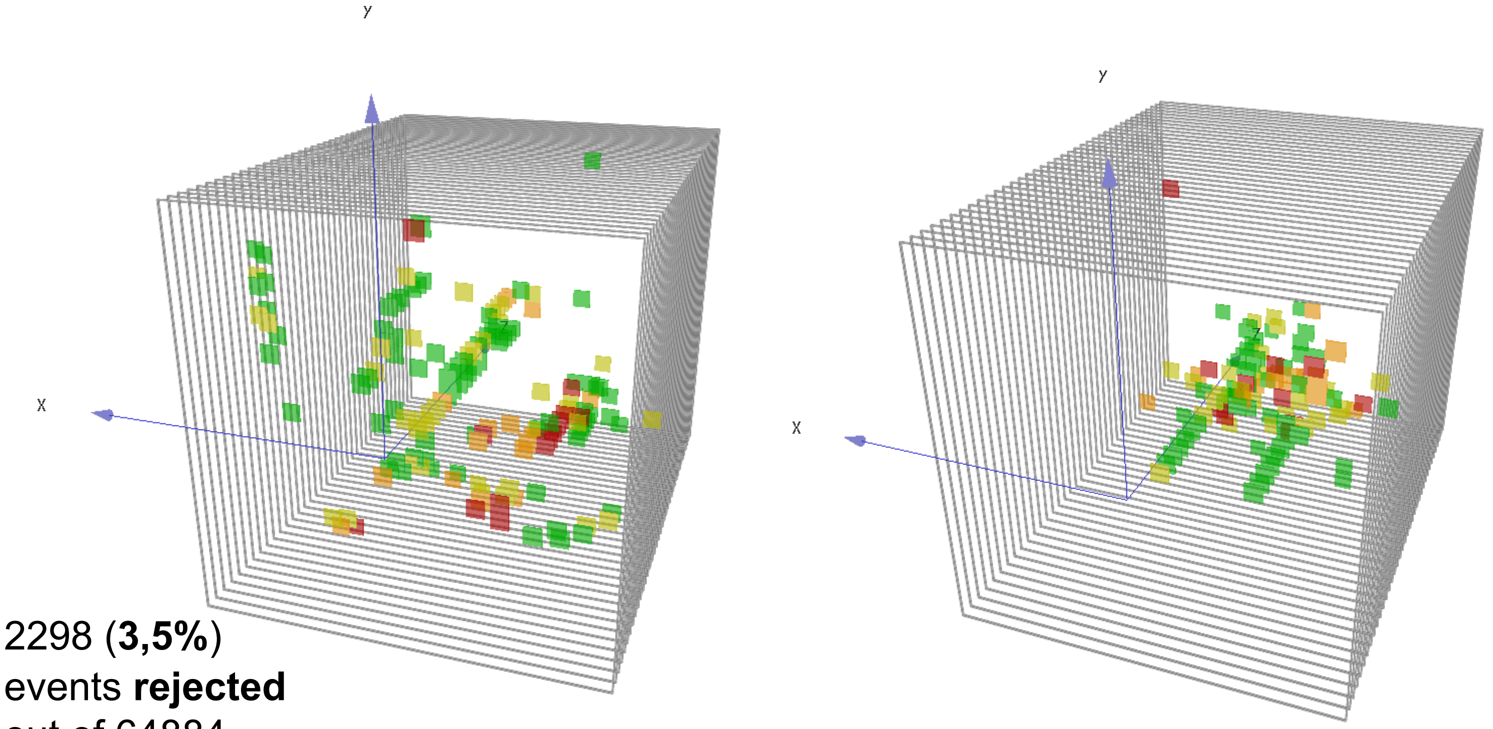


TBMay18 pion 10GeV

Apply filtering to data.

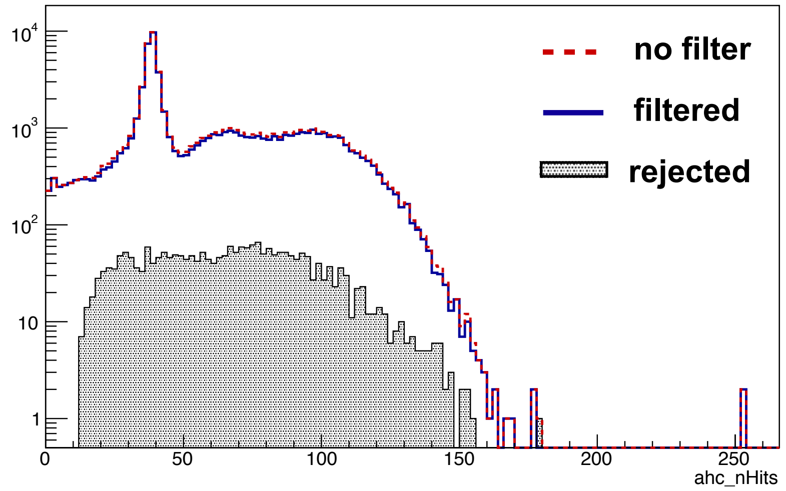
Shower radius, [mm]



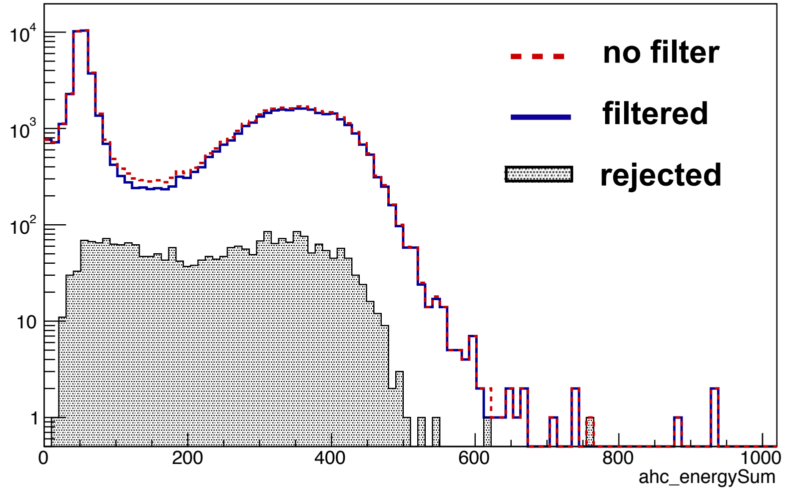


2298 (**3,5%**)
events **rejected**
out of 64884

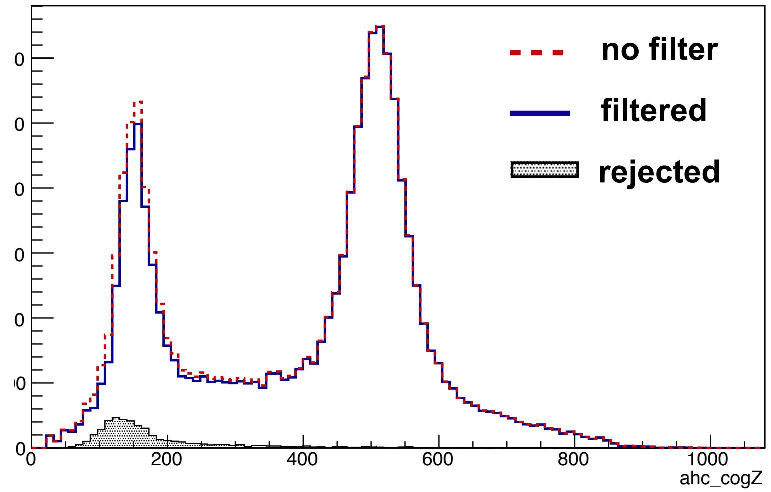
Number of hits



Energy sum, [MIP]



Center of gravity in z, [mm]



Rejected events. Simulations vs. data

10GeV pion.

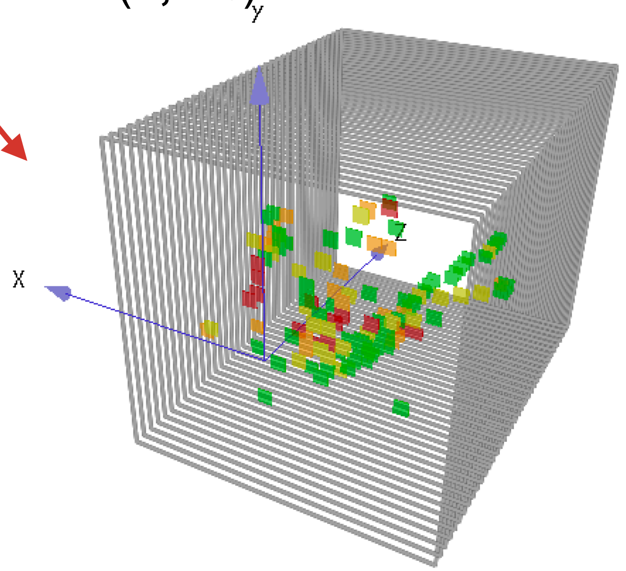
TBMay18 10GeV pion run:

2298 (**3,5%**) out of 64884 events

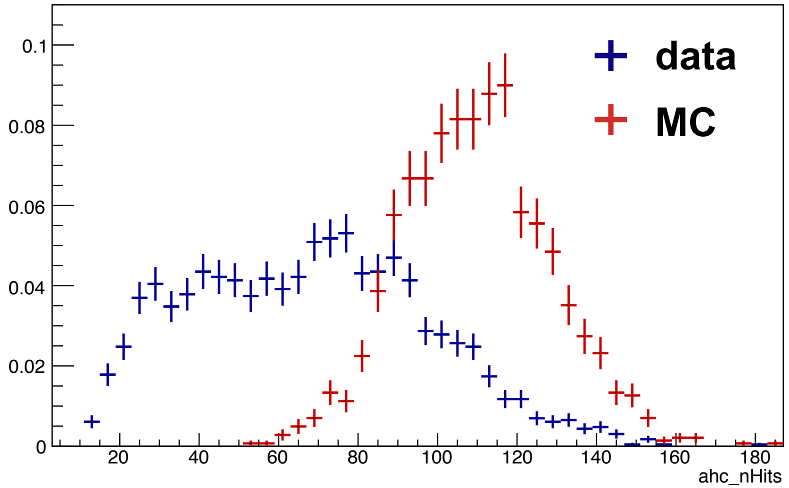
TBMay18 10GeV MC pion:

QGSP_BERT_HP physics list

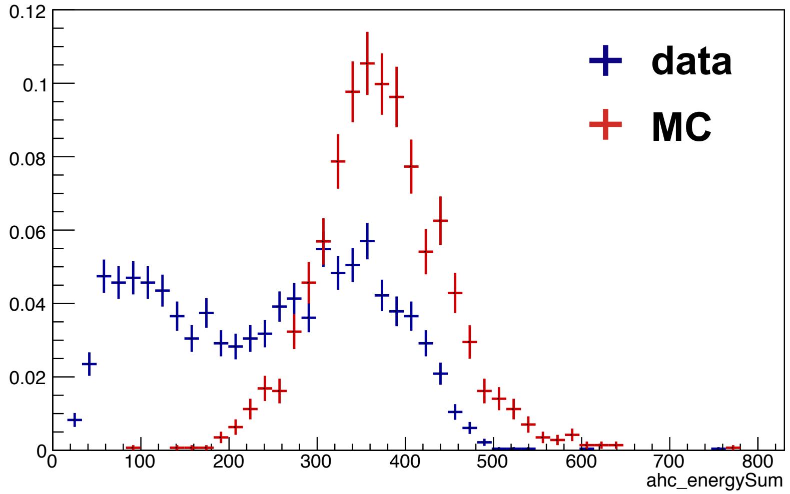
1423 (**3,3%**) out of 41000 events



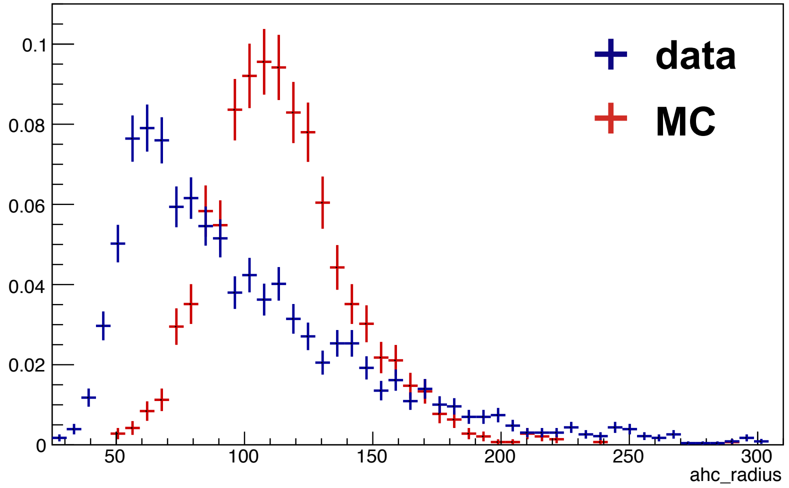
Number of hits



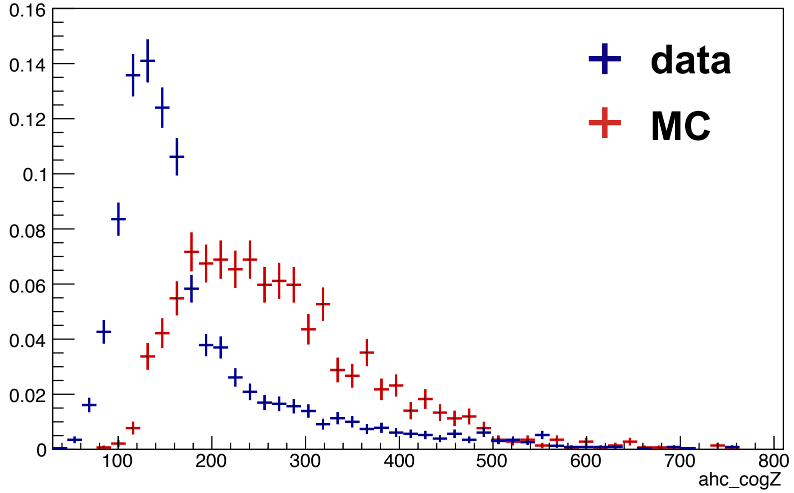
Energy sum, [MIP]



Shower radius, [mm]



Center of gravity in z, [mm]



Rejected events. Simulations vs. data

10GeV pion.

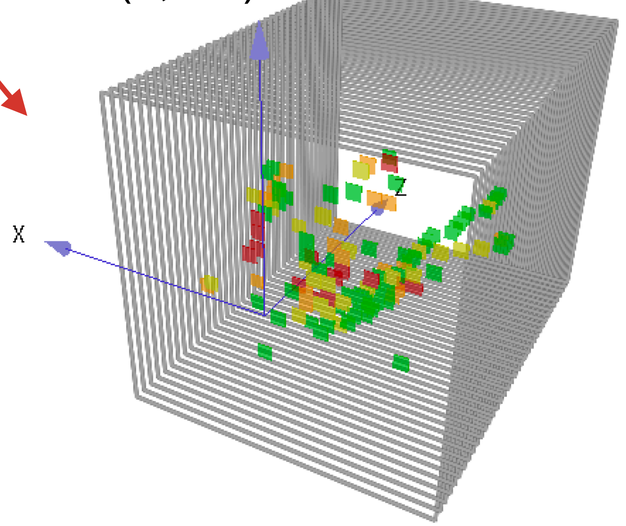
TBMay18 10GeV pion run:

2298 (**3,5%**) out of 64884 events

TBMay18 10GeV MC pion:

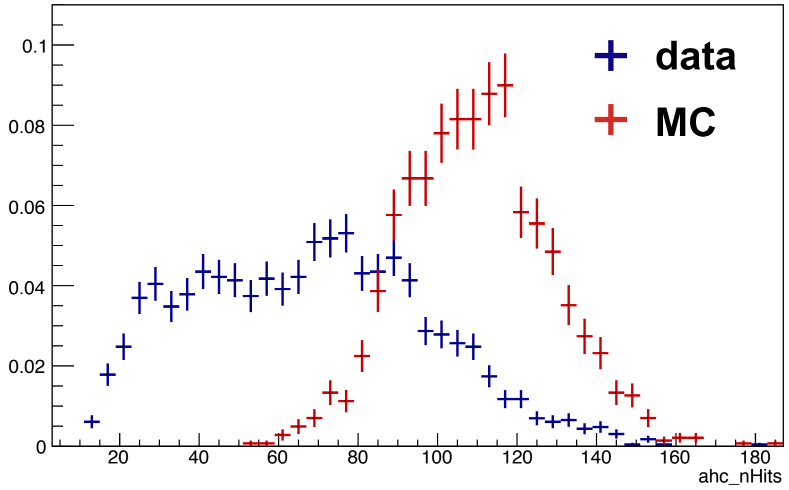
QGSP_BERT_HP physics list

1423 (**3,3%**) out of 41000 events

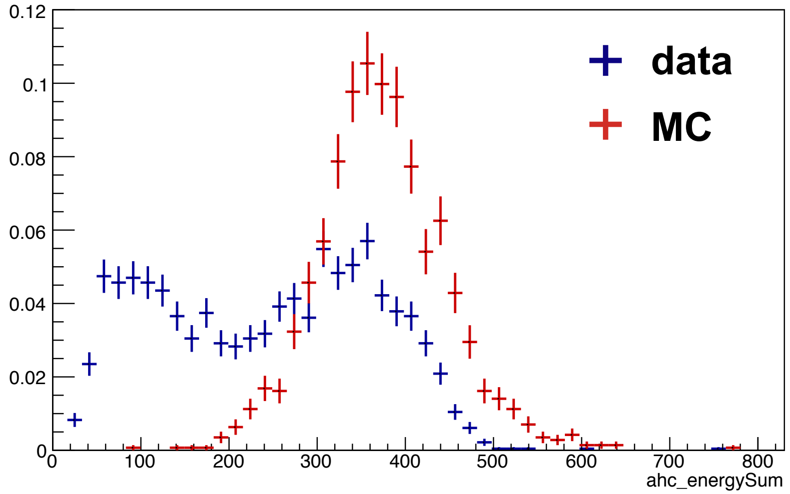


Same fraction of events for MC and data, **different distributions => too much early showering events simulated.**

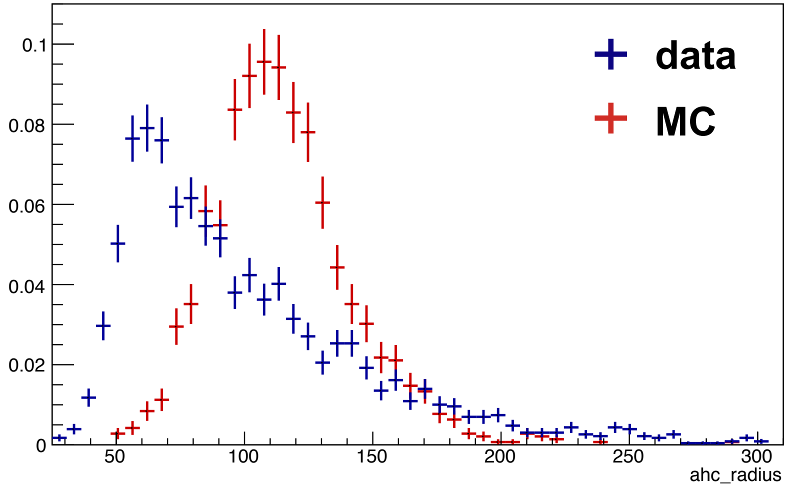
Number of hits



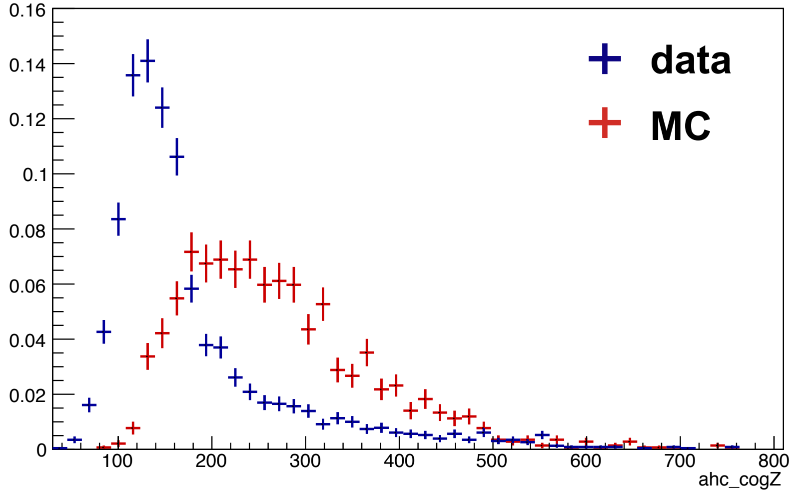
Energy sum, [MIP]



Shower radius, [mm]



Center of gravity in z, [mm]



Summary and outlook

Particle ID & Event filtering

- ☑ **Cut-based method** gives satisfying identification
  - Signal/background tuning is complicated
- ☑ **BDT method** gives promising results
  - ☑ Higher quality of identification
  - ☑ One parameter tuning will be useful for further calibration/physics analyses
  - ☐ Train BDT with all TB energies, implement to CaliceSoft
- ☑ **Event filtering** is implemented
  - ☑ - moves data to better agreement with MC
  - ☐ - should be checked with other TB runs (muons, other energies)
  - ☐ Early showering MC pions need further investigation
  - ☐ Clustering can be used for other studies
- ☐ Improve agreement of data and simulations (work in progress…)

Backup

Analysis flow

Of filtering studies

Reco + StdVariables calculation → Hit classification → 2D Clistering → Cluster analysis (in ***TBParticleID***)

Analysis flow

Of filtering studies

Reco + StdVariables calculation → Hit classification → 2D Clistering → Cluster analysis (in ***TBParticleID***)

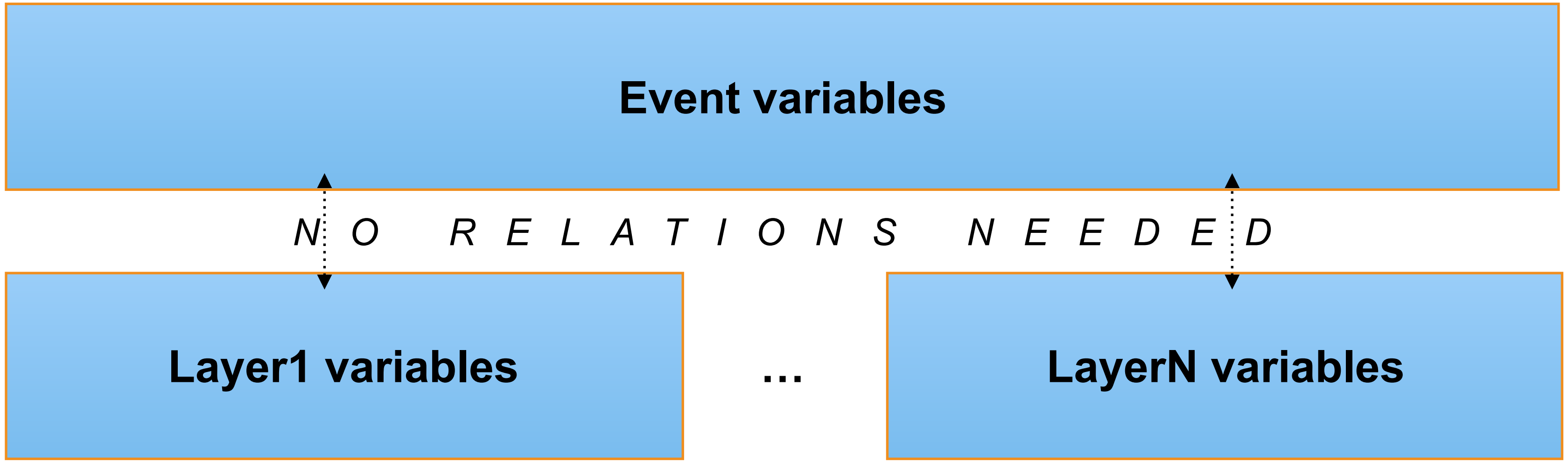
In *StdVariablesProcessor:*

- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

Analysis flow

Of filtering studies

Reco + StdVariables calculation



In *StdVariablesProcessor:*

- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

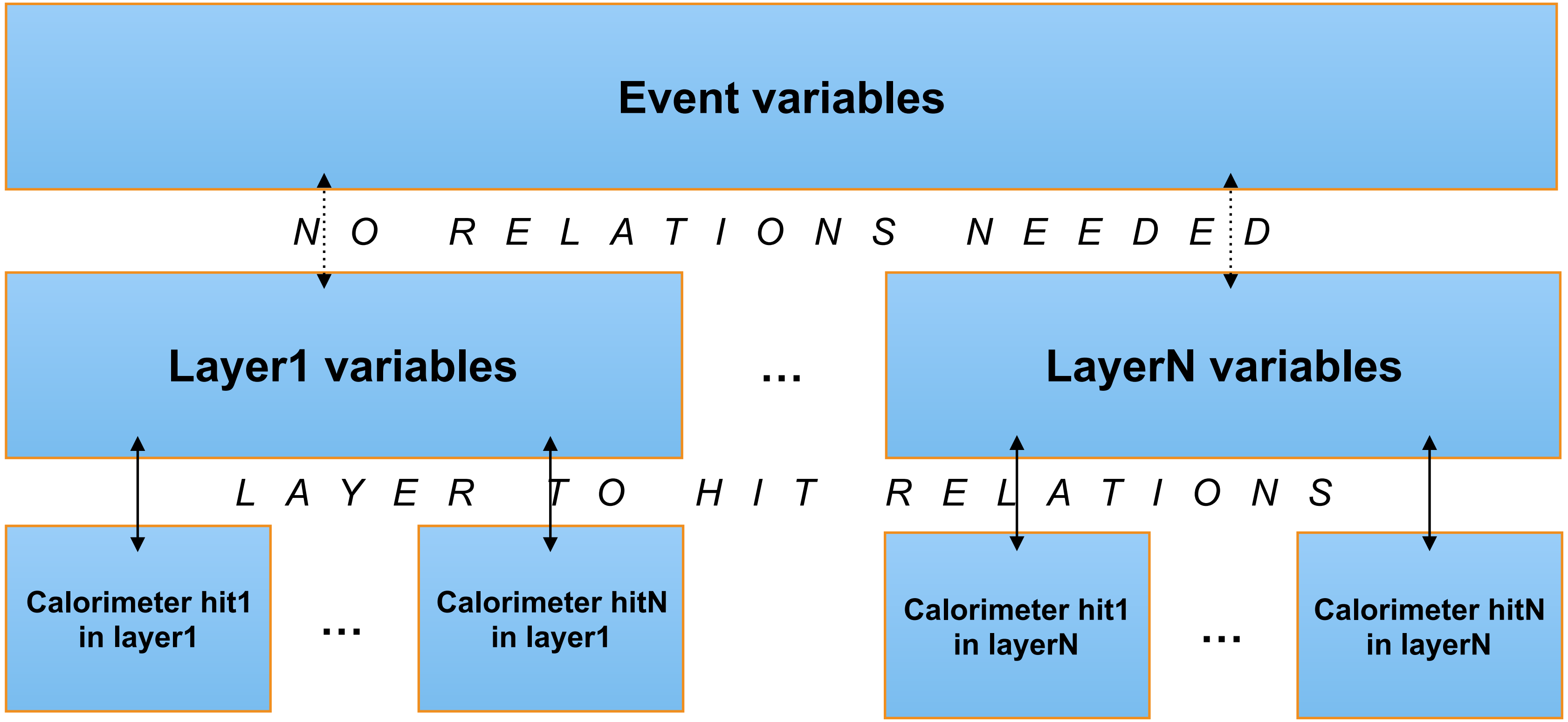
Analysis flow

Of filtering studies

Reco + StdVariables calculation

In *StdVariablesProcessor:*

- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

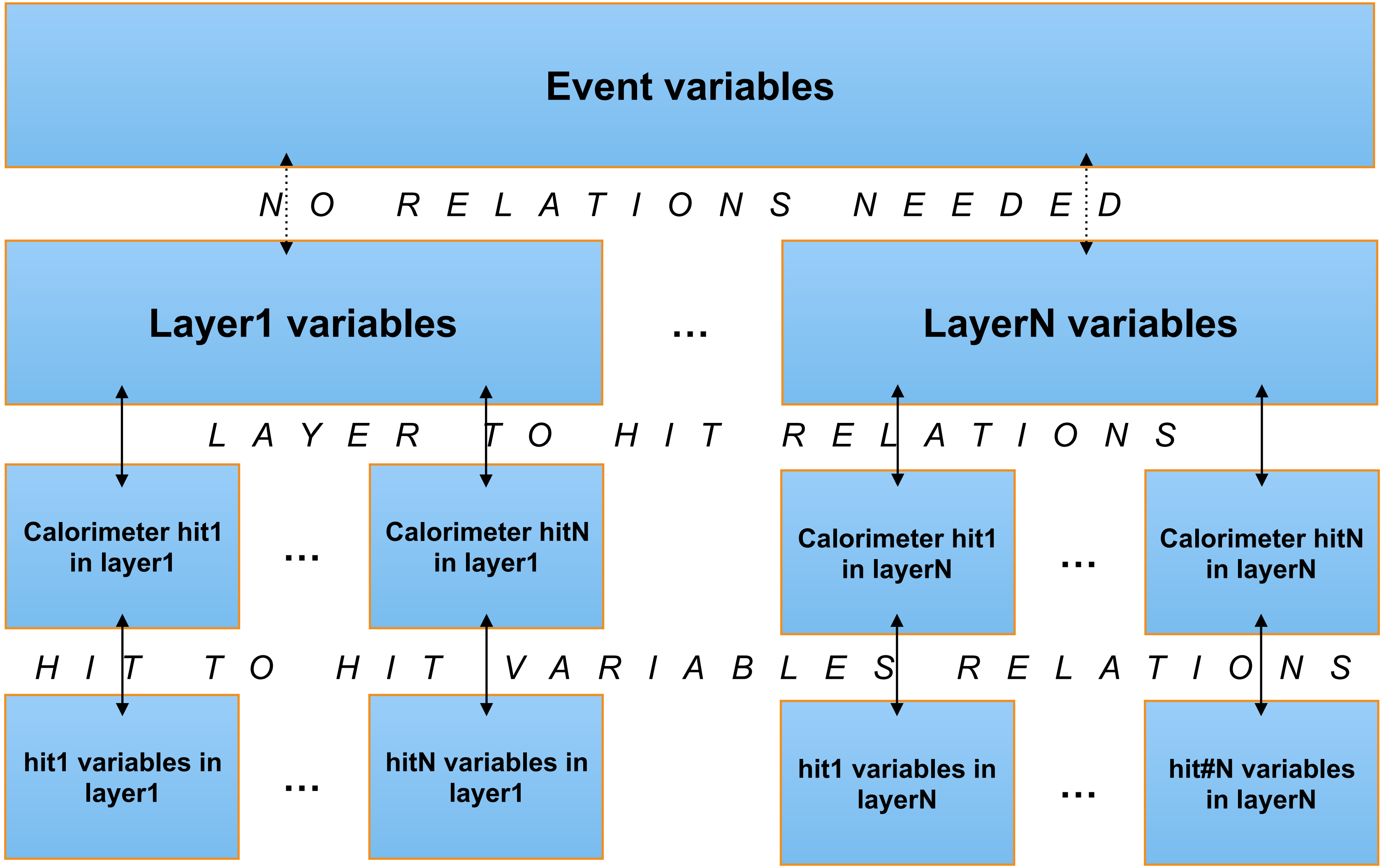


Analysis flow

Of filtering studies

Reco + StdVariables calculation

In *StdVariablesProcessor:*
- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables



Analysis flow

Of filtering studies

Reco + StdVariables calculation → **Hit classification** → 2D Clistering → Cluster analysis (in ***TBParticleID***)

In *StdVariablesProcessor:*
- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

In HitclassificarionProcessor*:*
- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

Analysis flow

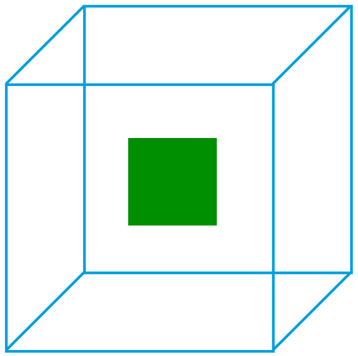
Of filtering studies

Hit classification

In HitclassificarionProcessor*:*

- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

Detached hits



Limited number of neighbours in volume (Calculated by {I,J,K} position)

Analysis flow

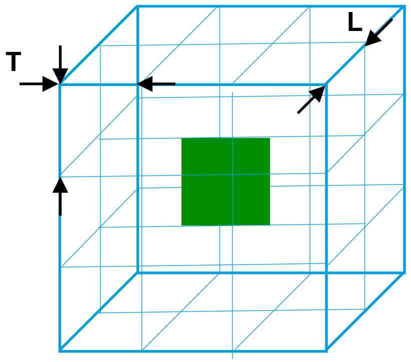
Of filtering studies

Hit classification

In HitclassificarionProcessor*:*

- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

Detached hits



Limited number of neighbours in volume (Calculated by {I,J,K} position)
Steering parameters:

- Size of volume (T and L)

Analysis flow

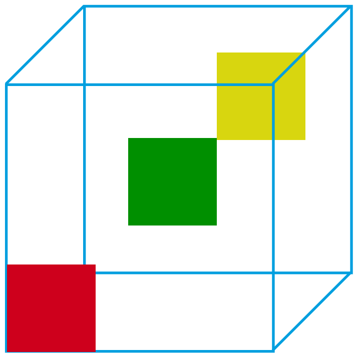
Of filtering studies

Hit classification

In HitclassificarionProcessor*:*

- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

Detached hits



Limited number of neighbours in volume
(Calculated by {I,J,K} position)
Steering parameters:

- Size of volume (T and L)
- Maximum number of neighbours

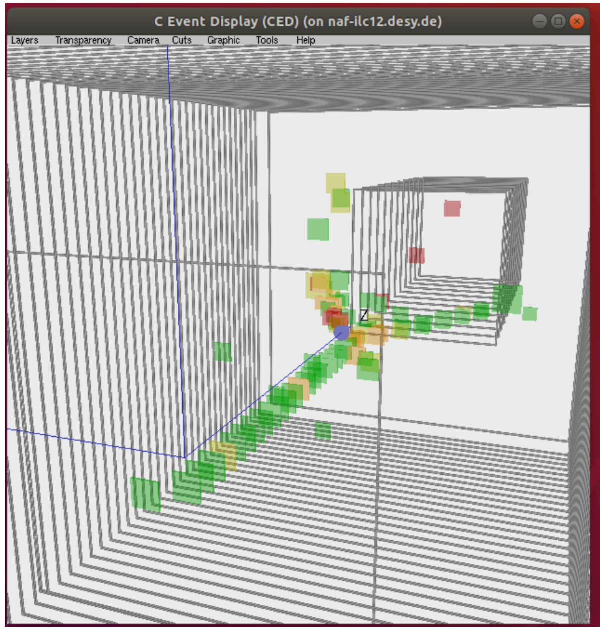
Analysis flow

Of filtering studies

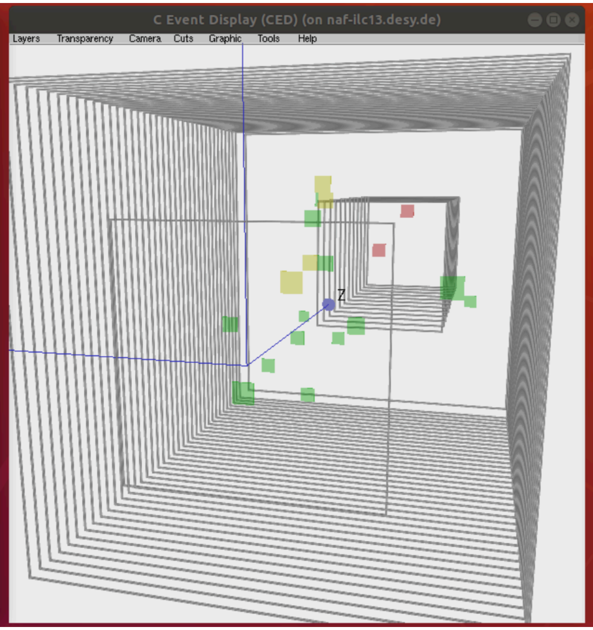
Hit classification

In HitclassificarionProcessor:

- Hits are divided into **two** groups:
    - **Detached** hits
    - **Non-detached** hits:
        - Track hit candidates (isolated)
        - Hits from cluster

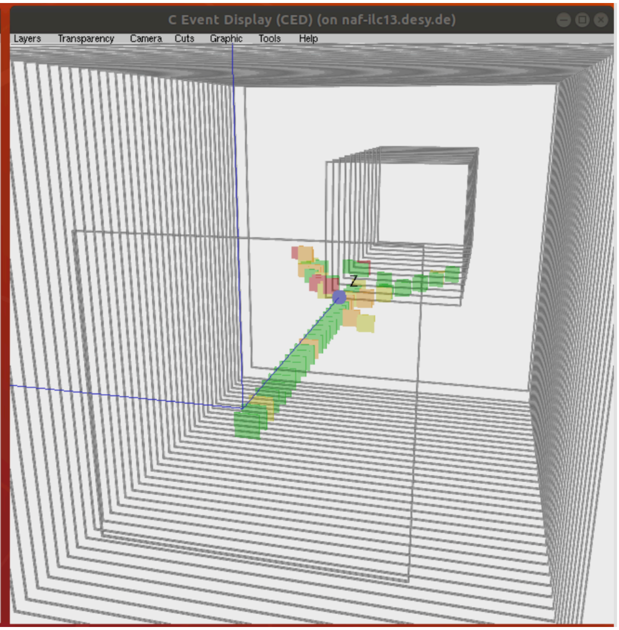


Full event



Only detached hits

No preshower hits



Only non-detached hits

No preshower hits

Parameters:

T = 1

L = 2

Nmax = 2

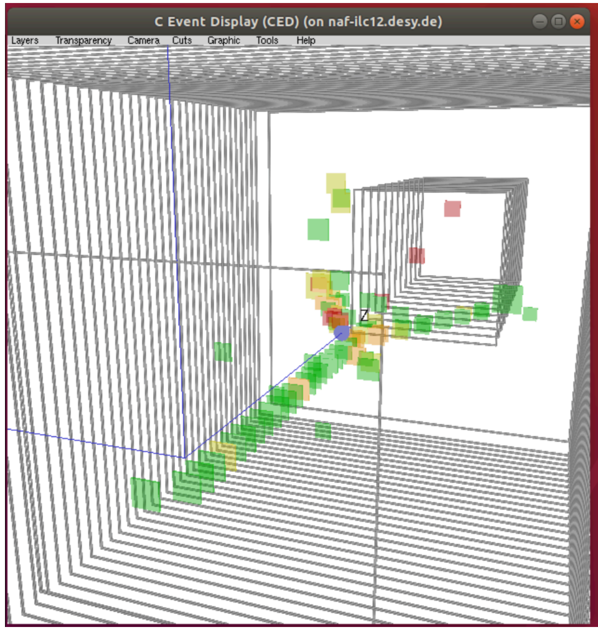
Analysis flow

Of filtering studies

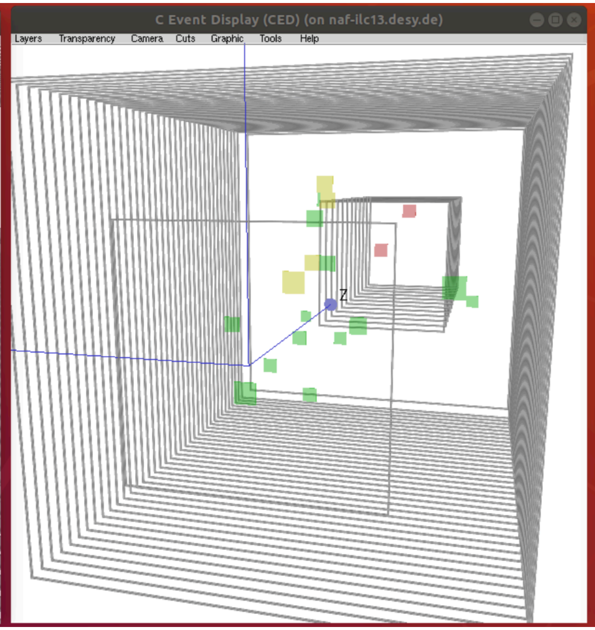
Hit classification

In HitclassificarionProcessor:

- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

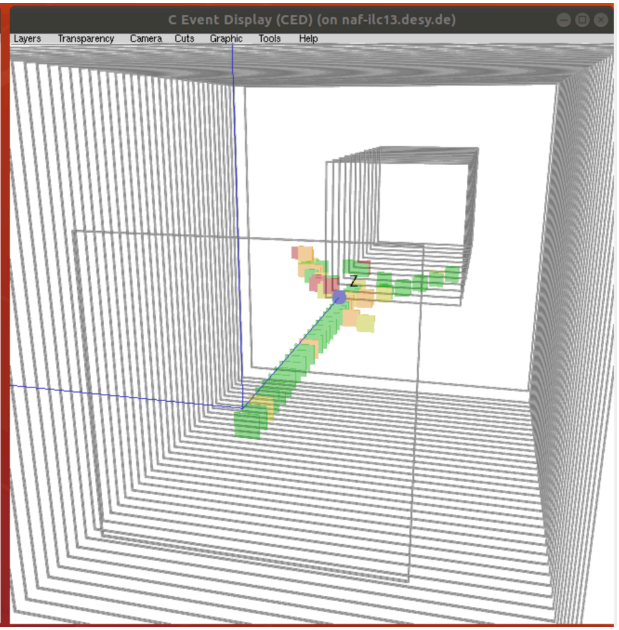


Full event



Only detached hits

No preshower hits



Only non-detached hits

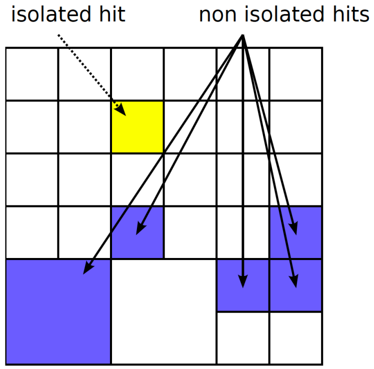
No preshower hits

Parameters:

T = 1

L = 2

Nmax = 2



CAN22

Analysis flow

Of filtering studies

Reco + StdVariables calculation → Hit classification → **2D Clistering** → Cluster analysis (in ***TBParticleID***)

In *StdVariablesProcessor:*

- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

In HitclassificarionProcessor*:*

- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

In ClusterCounterProcessor*:*

- **Group** all **non-detached hits** to clusters
- **Calculate CoG** in *x* and *y* for each cluster
- **Set relations** from layer variables to clusters

Analysis flow

Of filtering studies

Reco + StdVariables calculation → Hit classification → 2D Clistering → Cluster analysis (in ***TBParticleID***)

In *StdVariablesProcessor:*
- **Calculate** event, layer and hit **variables**
- **Set LCRelations** between sets of variables

In HitclassificarionProcessor*:*
- Hits are divided into **two** groups:
  - **Detached** hits
  - **Non-detached** hits:
    - Track hit candidates (isolated)
    - Hits from cluster

In ClusterCounterProcessor*:*
- **Group** all **non-detached hits** to clusters
- **Calculate CoG** in *x* and *y* for each cluster
- **Set relations** from layer variables to clusters

In *TBParticleID:*
- **Calculate average spread of clusters** in **first** *nFirstLayers* **layers**

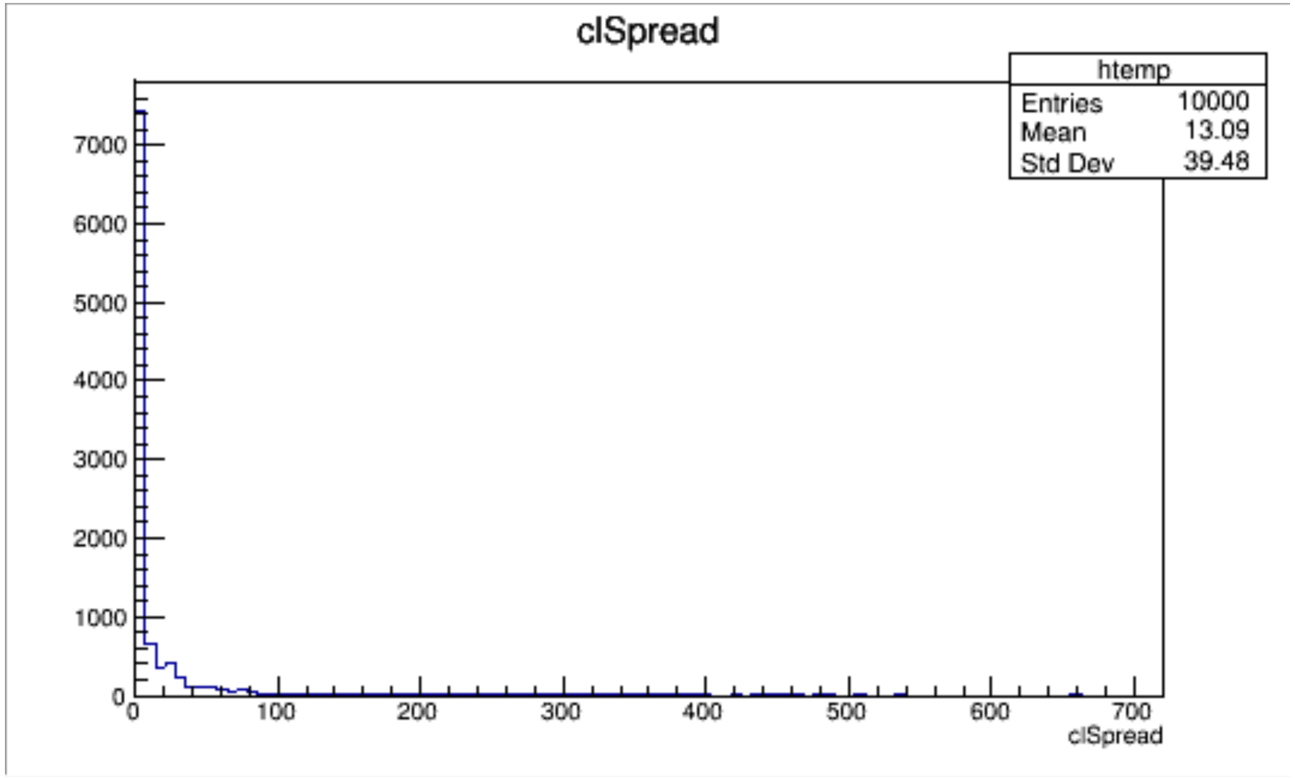
Analysis flow

Of filtering studies

Cluster analysis
(in ***TBParticleID***)

In *TBParticleID:*

- **Calculate average** transverse **spread of clusters** in **first** *nFirstLayers* **layers**



nFirstLayers = 5

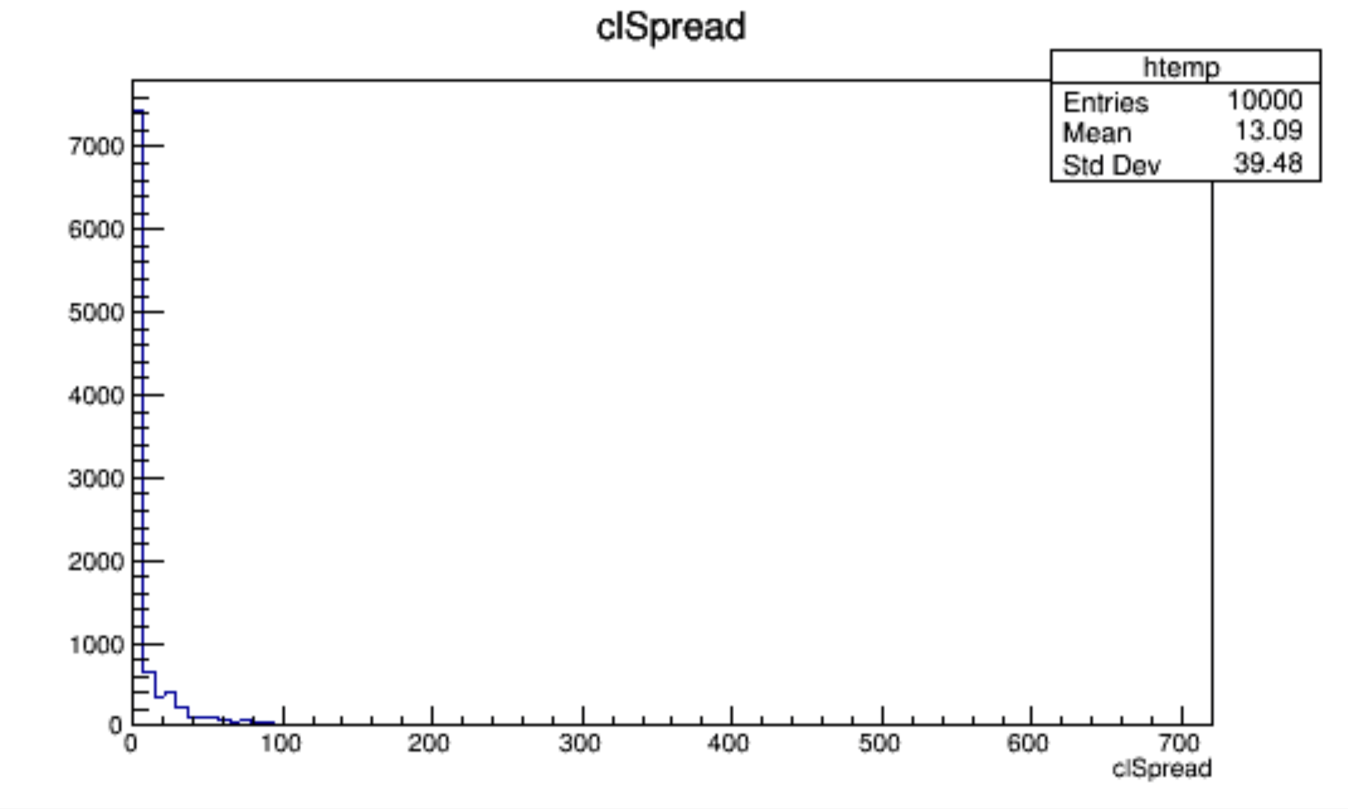
Analysis flow

Of filtering studies

Cluster analysis
(in ***TBParticleID***)

In *TBParticleID:*

- **Calculate average** transverse **spread of clusters** in **first** *nFirstLayers* **layers**



nFirstLayers = 3

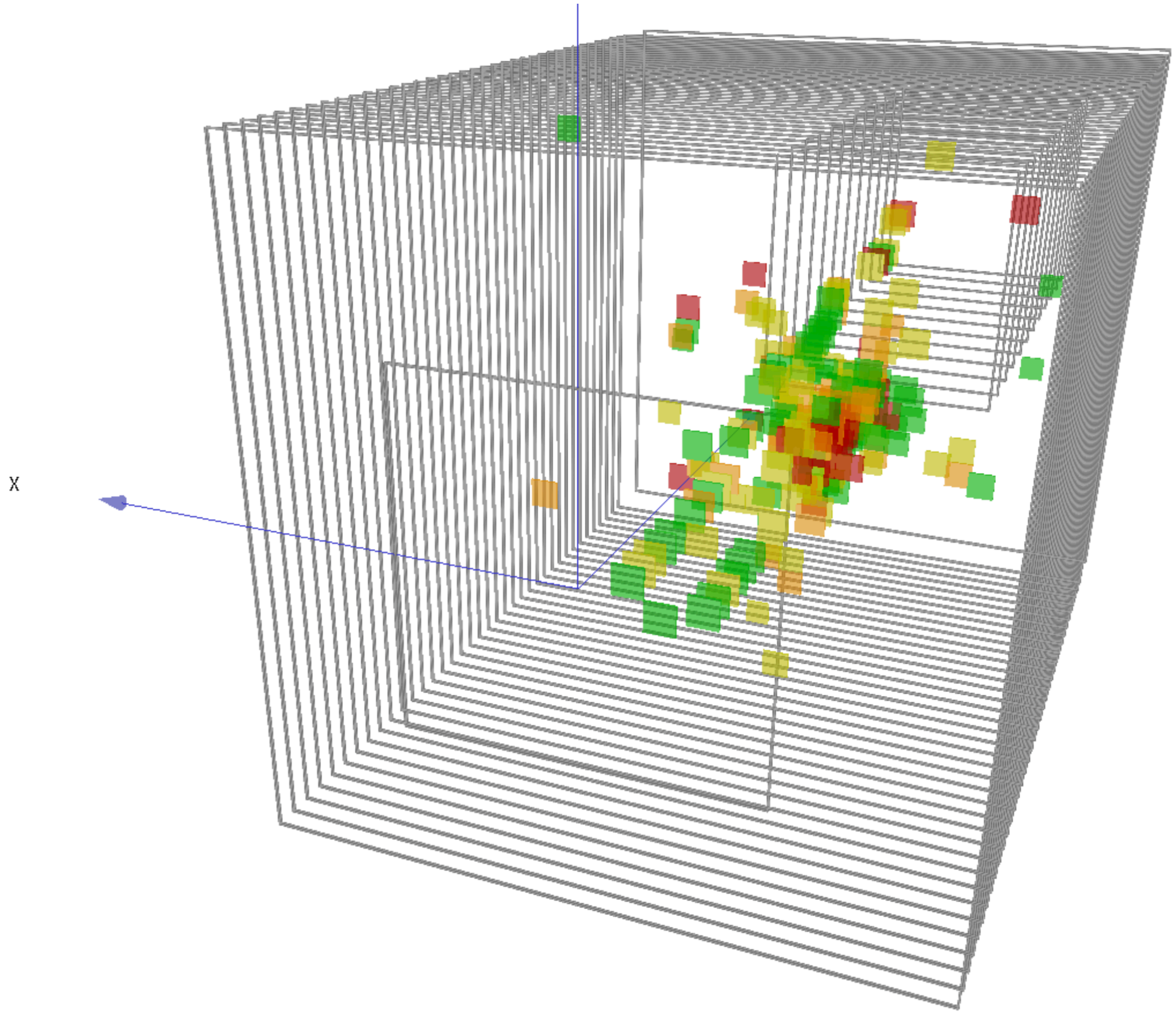
Analysis flow

Of filtering studies

Cluster analysis
(in ***TBParticleID***)

In *TBParticleID:*

- **Calculate average** transverse **spread of clusters** in **first** *nFirstLayers* **layers**



nFirstLayers = 3, minSpread(for rejection) = 30mm

Analysis flow

Of filtering studies

Cluster analysis
(in ***TBParticleID***)

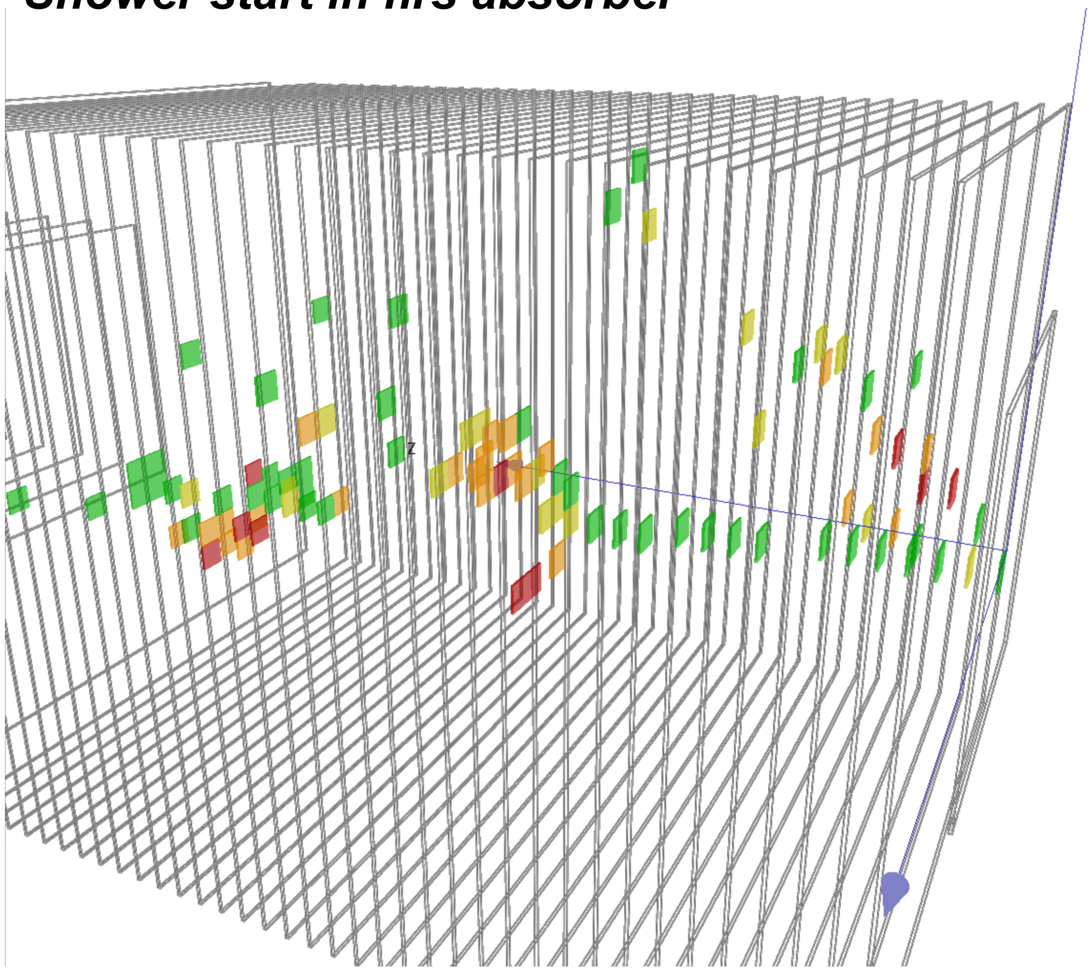
In *TBParticleID:*

- **Calculate average** transverse **spread of clusters** in **first** *nFirstLayers* **layers**

Problems:

- Shower start in first absorber

Shower start in firs absorber



nFirstLayers = 3

TODO: Set weights to layers

Analysis flow

Of filtering studies

Cluster analysis
(in ***TBParticleID***)

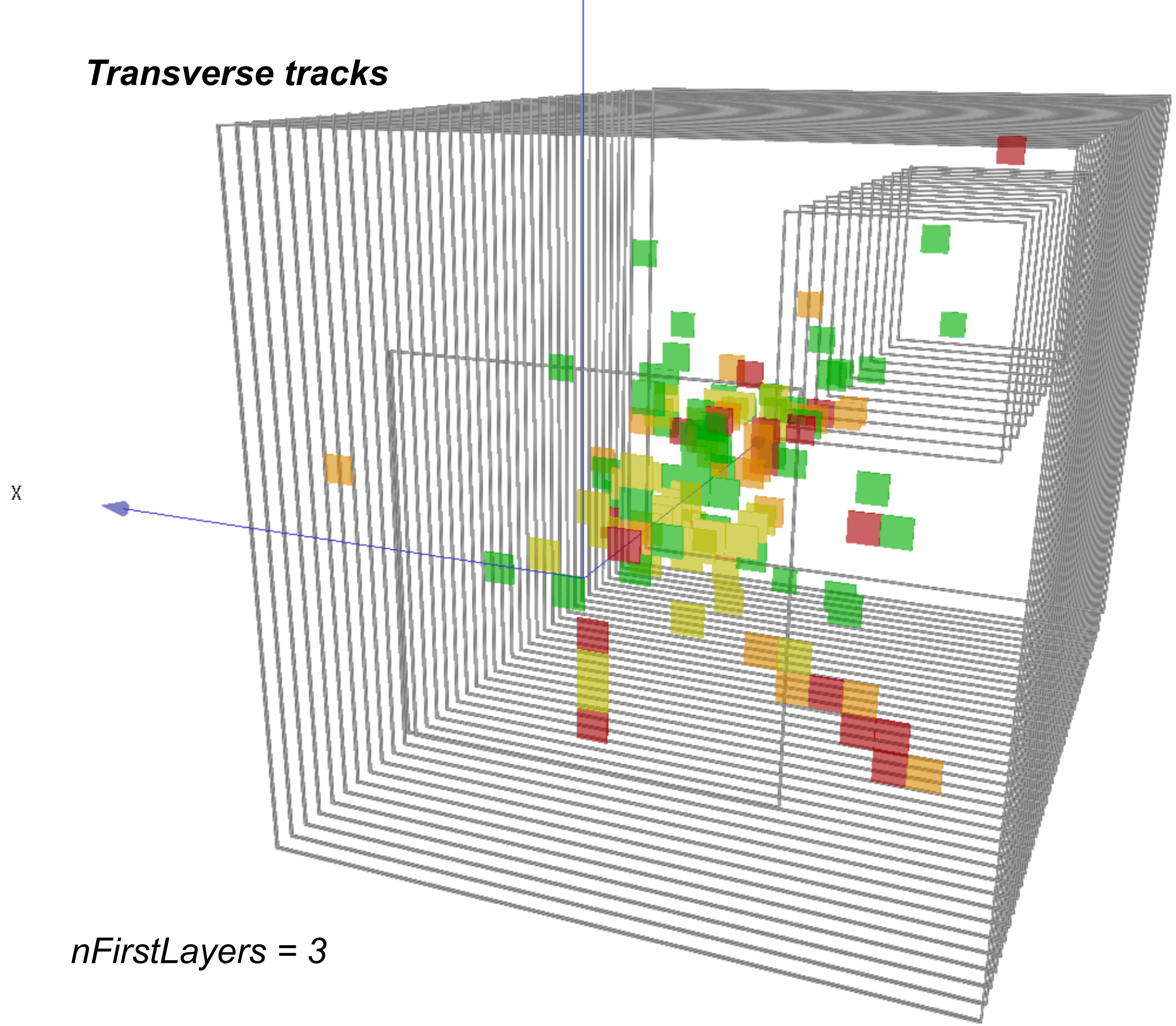
In *TBParticleID:*

- **Calculate average** transverse **spread of clusters** in **first** *nFirstLayers* **layers**

Problems:

- Shower start in first absorber
- Transverse and back-scattering tracks

Transverse tracks



nFirstLayers = 3