Test beam particle ID

Event filtering

Vladimir Bocharnikov
AHCAL analysis workshop, 13-19 Dec 2018

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

CALICE Calorimeter for ILC

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

NATIONAL RESEARCH NUCLEAR UNIVERSITY MEPhI NRNU MEPhI

DESY.

Particle ID

Reminder from AHCAL main meeting

Method* developed for physics prototype by Marina Chadeeva was taken as basement:

Calorimeter-based id by topology and cluster properties

Clustering and shower start finding precedes particle identification.
Set of cut-based topological discriminants is based on published CANs and papers.

- **Multi-particle events**: OR-based combination of discriminants
  - several clusters in first layers
  - parallel tracks
  - too high deposition
- **Muons or muon-like**: complex combination of discriminants
  - number of hits
  - energy comparison in calorimeters
  - track finding
- **Electrons**: AND-based combination of discriminants
  - cluster radius <37 mm
  - CoG depth <10·X0
  - 90% of measured energy deposited in first 24·X0
  - shower start before 6·X0
- **Hadrons**:
  - remaining species after above rejection
  - Cherenkov-based discrimination of hadron types

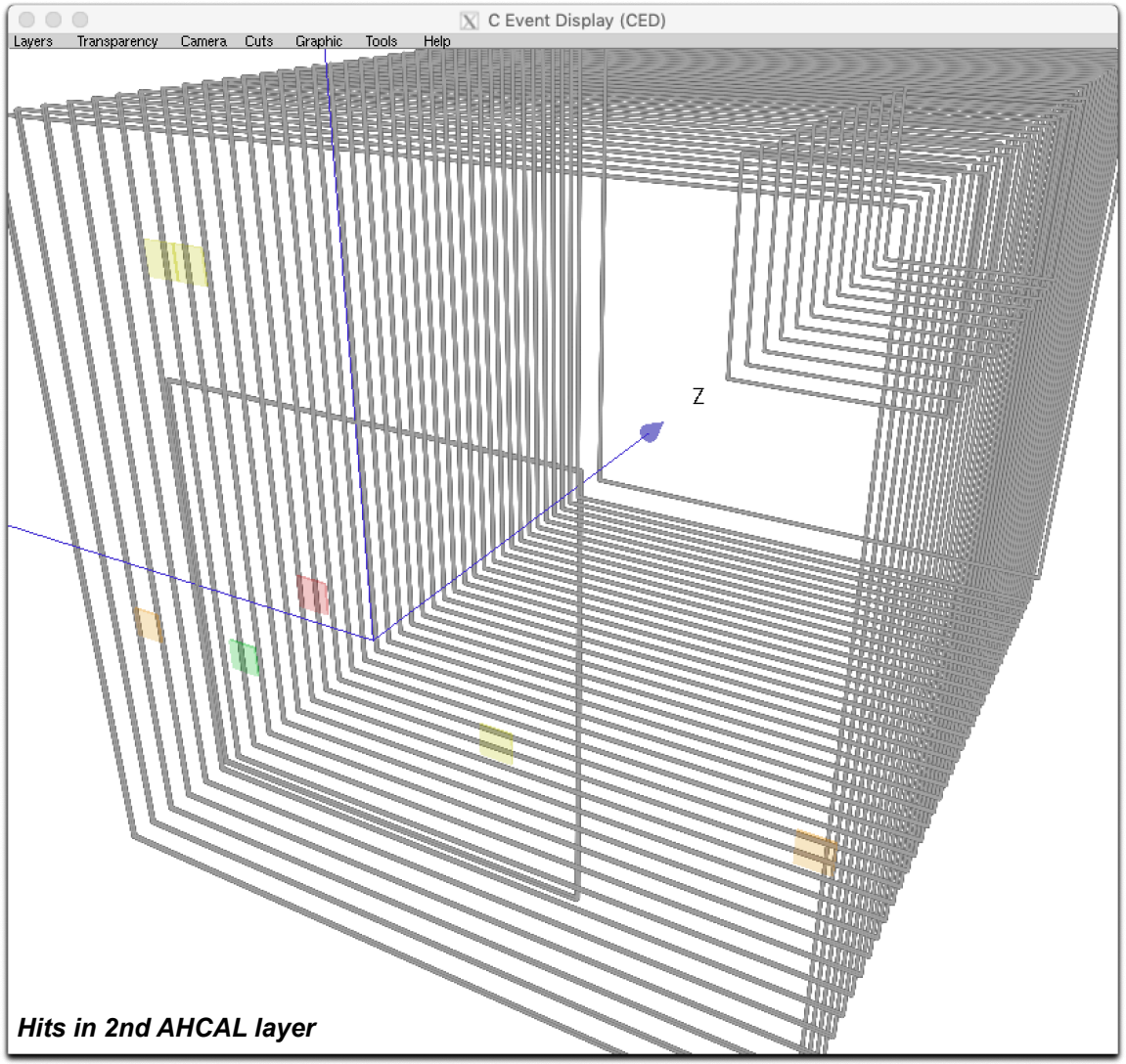
*https://agenda.linearcollider.org/event/7454/contributions/38731/attachments/31380/47203/chadeeva_Fe-W_LLR2017.pdf

- **Same approach**
- **Same variables are used**
- First **cut adjustment** was done using MC data sets simulated and fully reconstructed according to **TBMay18** and **TBJune18** setups using QGSP_BERT_HP physics list
- **CALICE software** was modified to **centralise** the **calculation** of all common **variables**
- Clustering and tracking analyses are not implemented yet
- **Only AHCAL hits** are taken into account

Clustering

For event filtering

- Algorithm rejects events that have too much clusters in first layers



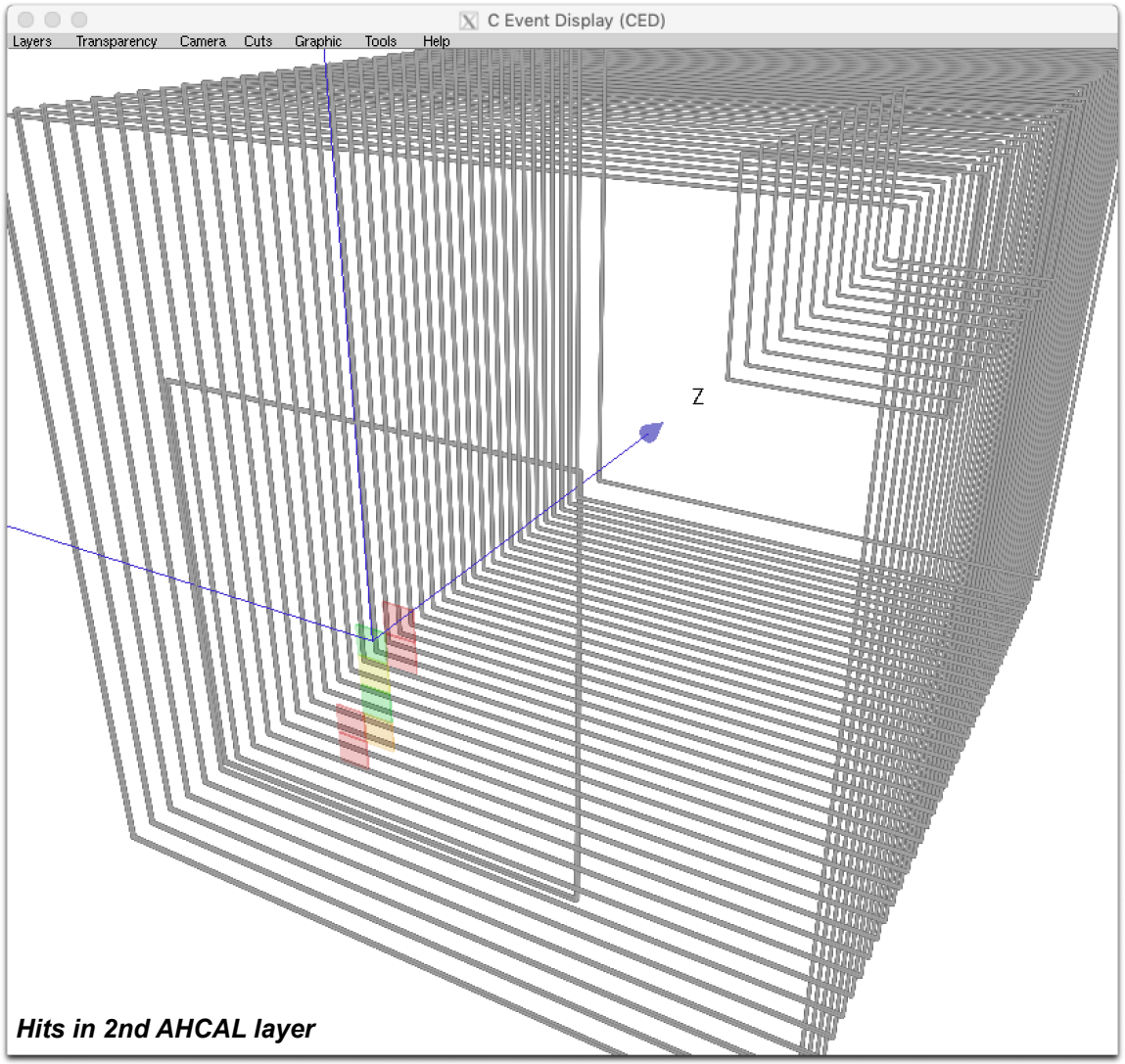
Hits in 2nd AHCAL layer

Clustering

For event filtering

- Algorithm rejects events that have too much clusters in first layers
- For such as events output is 1 cluster

⇒ Energy weighted method is needed



Hits in 2nd AHCAL layer

Clustering

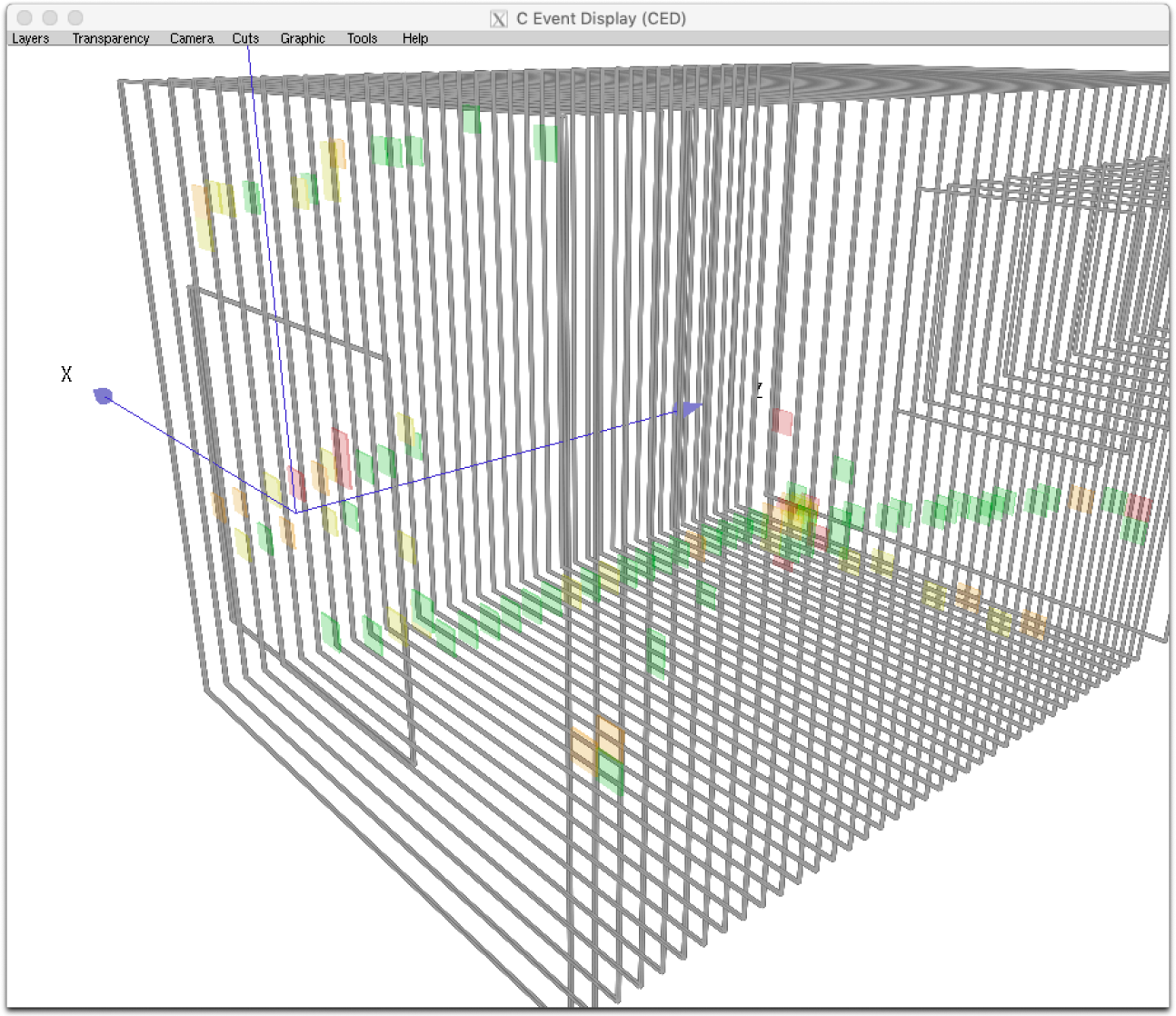
For event filtering

- Algorithm rejects events that have too much clusters in first layers
- For such as events output is 1 cluster

⇒ Energy weighted method is needed

- Can reject multi-particle events

```
nClusters in 1st layer: 5
nClusters in 2nd layer: 6
```



Clustering

For event filtering

- Algorithm rejects events that have too much clusters in first layers
- For such as events output is 1 cluster

⇒ Energy weighted method is needed

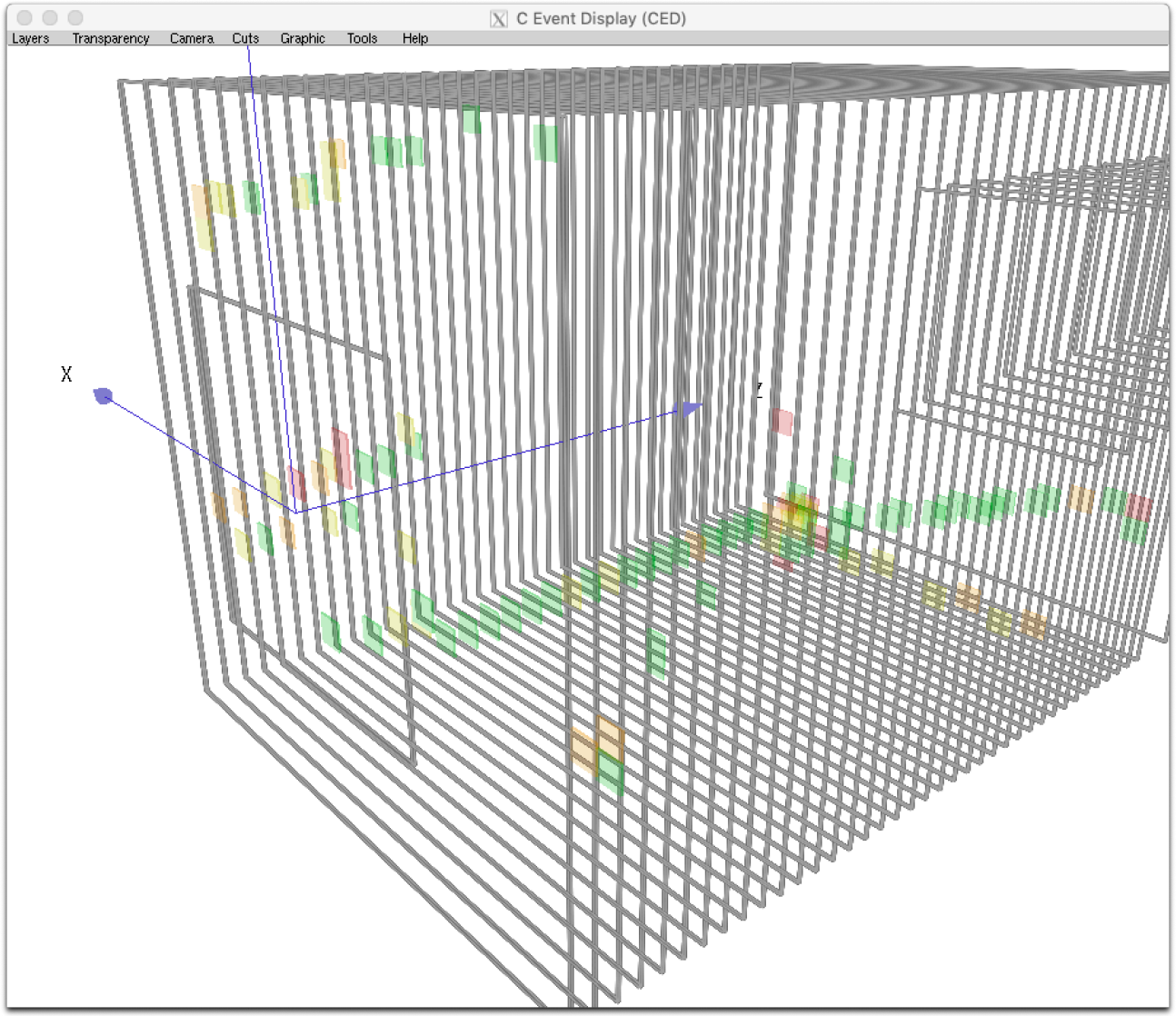
- Can reject multi-particle events

```
nClusters in 1st layer: 5
nClusters in 2nd layer: 6
```

- Problem in AHCAL_hits collection:

For TBMay18 “1st layer” - 1st AHCAL layer (K=1)

For TBJune18 “1st layer” - pre-shower layer



Clustering

For event filtering

- Algorithm rejects events that have too much clusters in first layers
- For such as events output is 1 cluster

⇒ Energy weighted method is needed

- Can reject multi-particle events

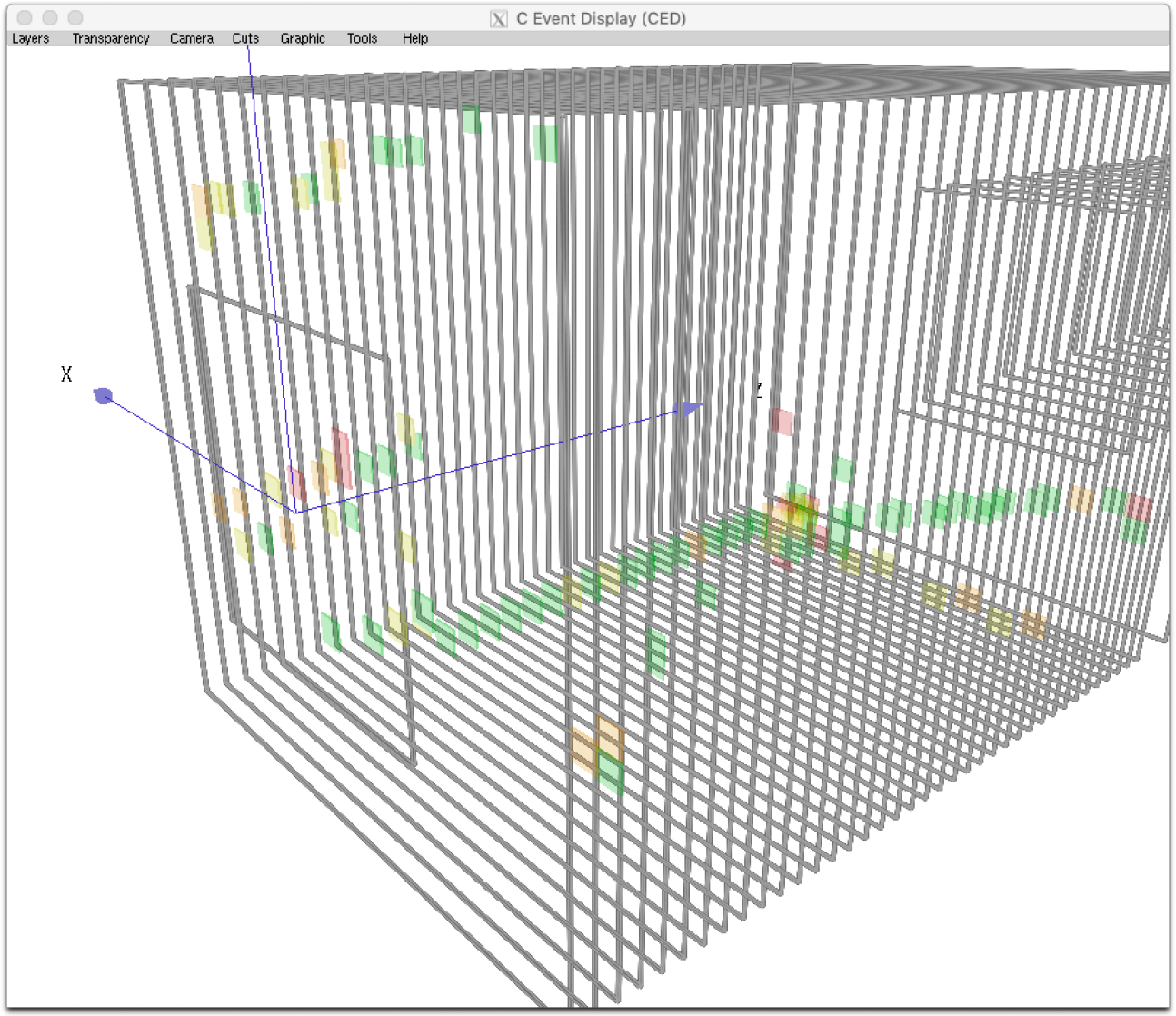
```
nClusters in 1st layer: 5
nClusters in 2nd layer: 6
```

- Problem in AHCAL_hits collection:

For TBMay18 “1st layer” - 1st AHCAL layer (K=1)

For TBJune18 “1st layer” - pre-shower layer

Additional parameter in the IDprocessor
or set K = 0 for the pre-shower layer?



Conclusion & ToDo list

Particle ID

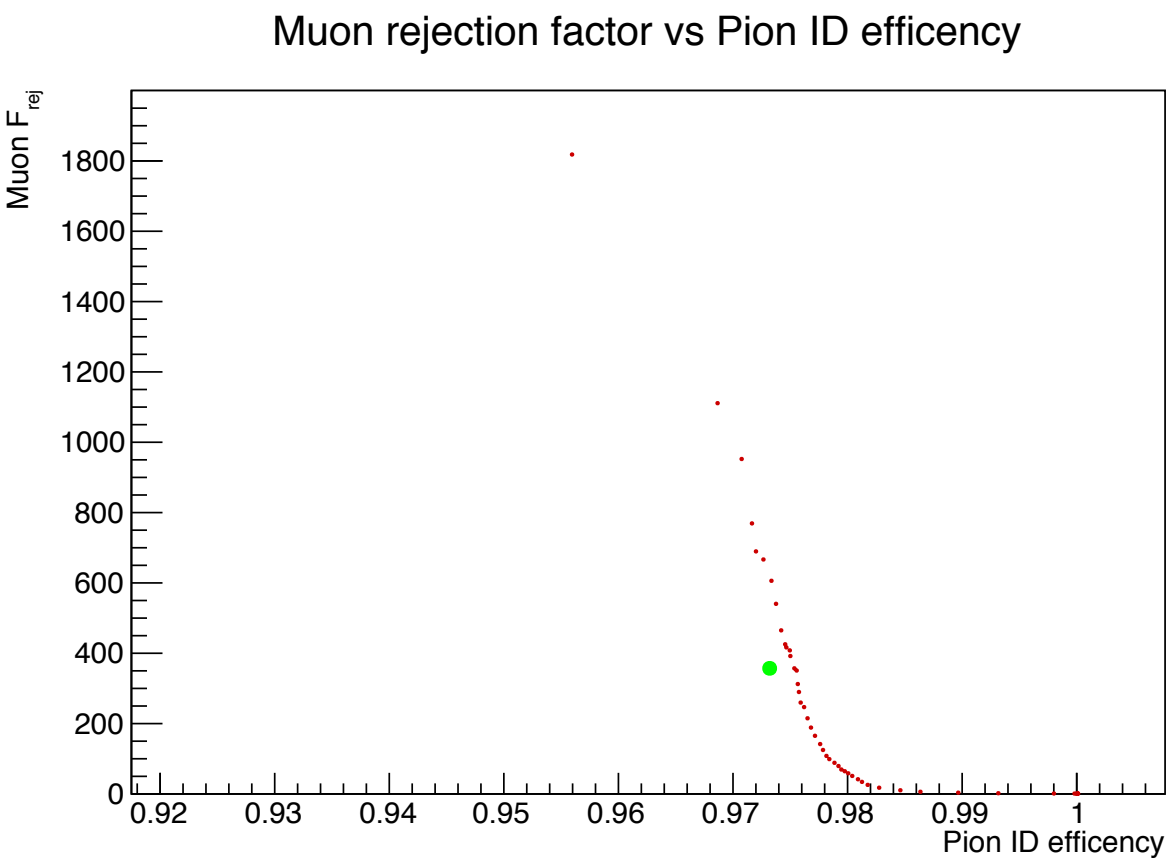
- ☑ At 10GeV we have satisfying cut-based identification
  - ☑ Software infrastructure is ready for more precise studies
  - ☑ First attempt of implementation by Amine and Anna
  - ☐ MC check of energy dependent cuts is needed
  - ☐ Investigate problem with energy and number of hits in simulations
- ☐ We need to implement event quality selection for data:
  - ☐ Multi-particle events
  - ☐ Early showering particles

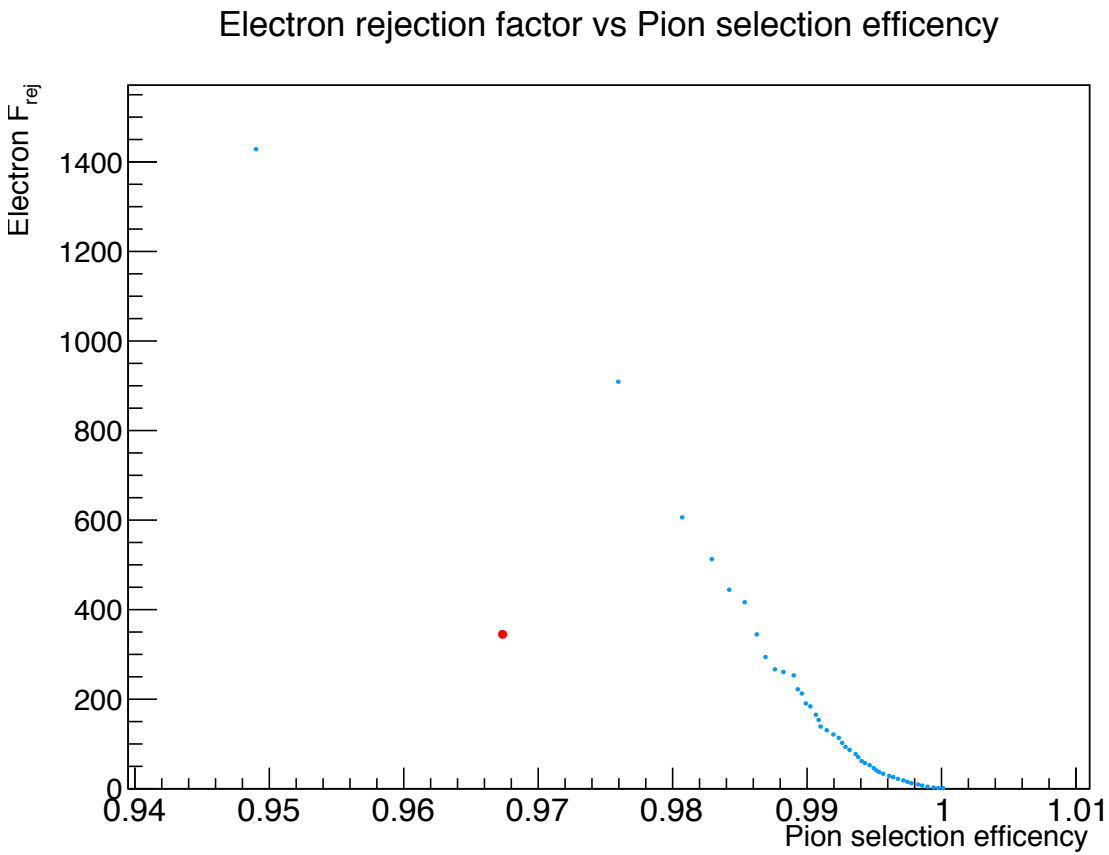
  } ☑ Naive clustering can partly reject this events
  - ☐ …
- ☑ BDT method gives promising results
  - ☐ Cuts for training samples
  - ☐ Train BDT with all TB energies
  - ☐ …
  - ☐ After better agreement MC with data there will be more BDT studies

Backup

Background rejection vs signal efficiency

Another representation of performance.

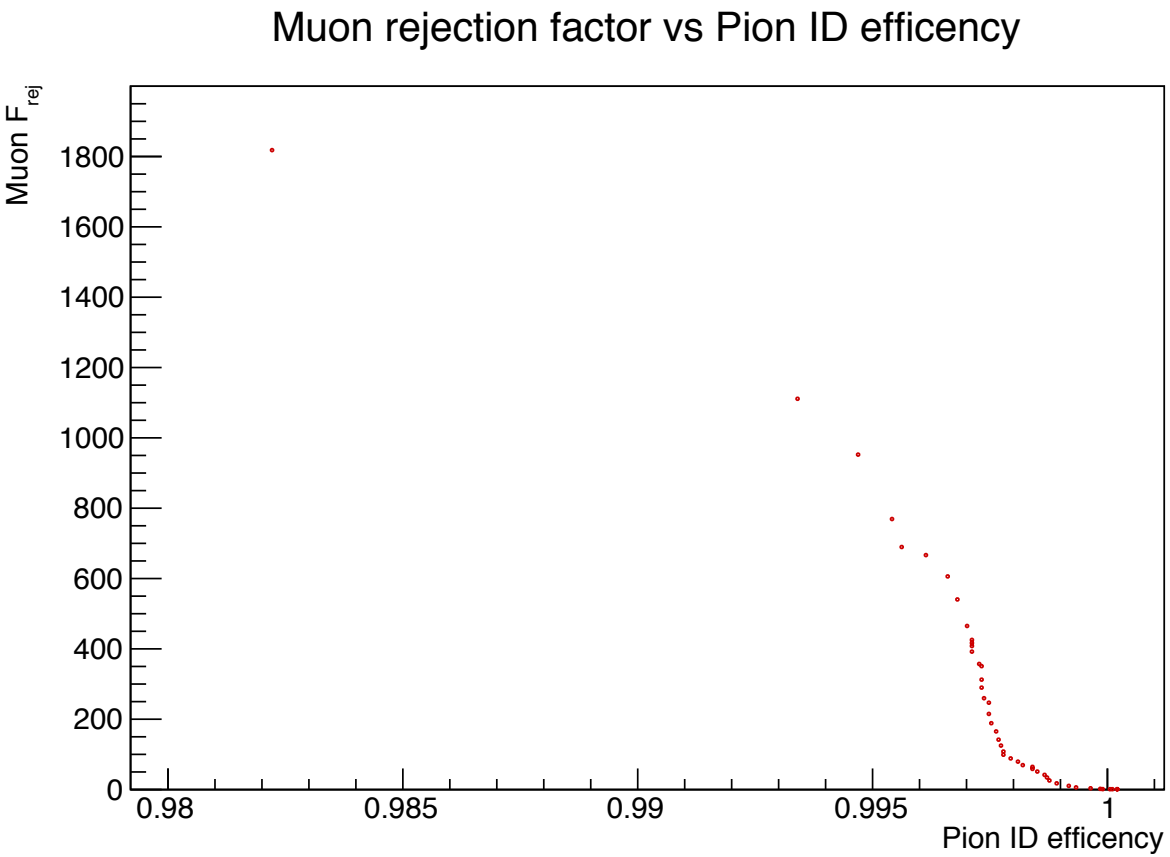


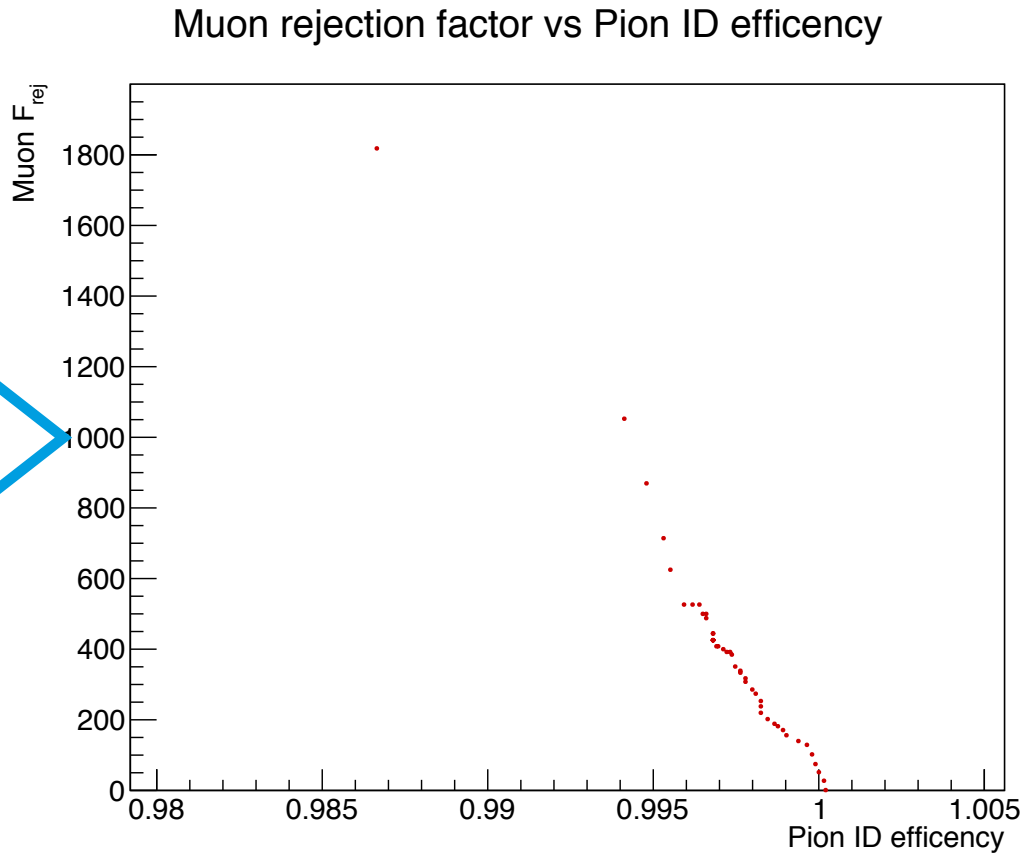


New cut for BDT input

On shower start layer. Muon vs pion separation

Cut: Pion events w/o shower are thrown away.





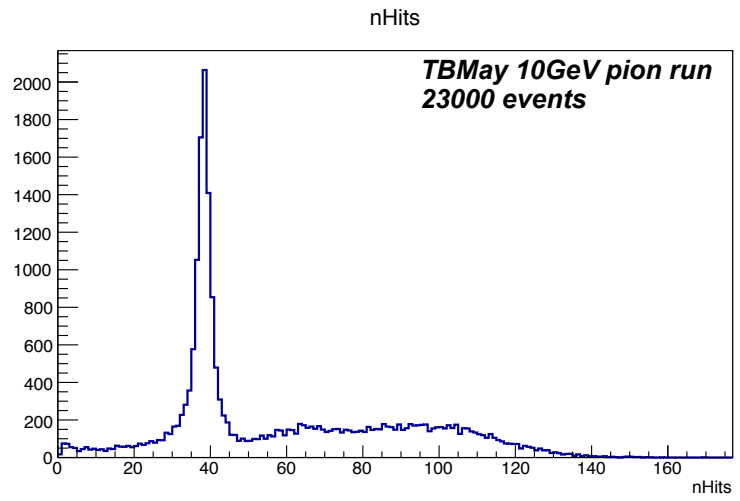
Observables

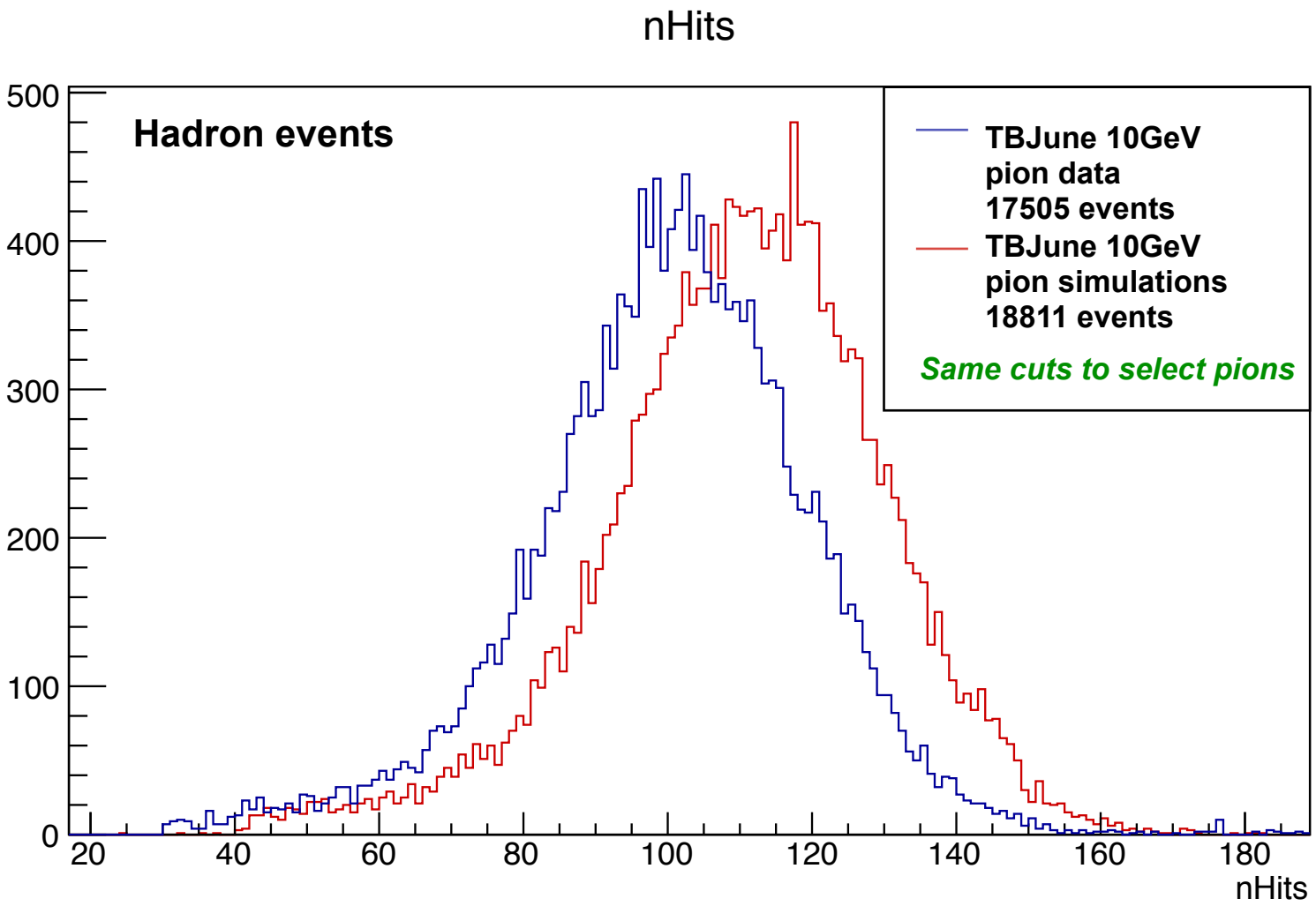
For particle ID

Number of hits per event

- "clean" μ^- : $N_{hits} \approx N_{layers}$

Probably, 0.5MIP cut gives more hits for MC



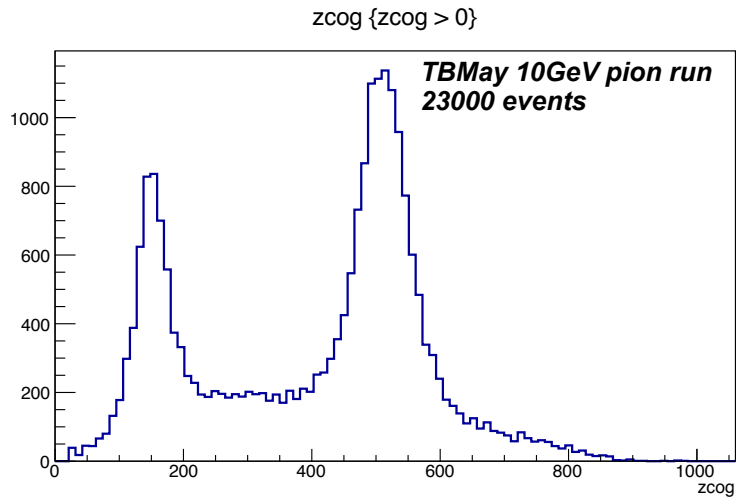


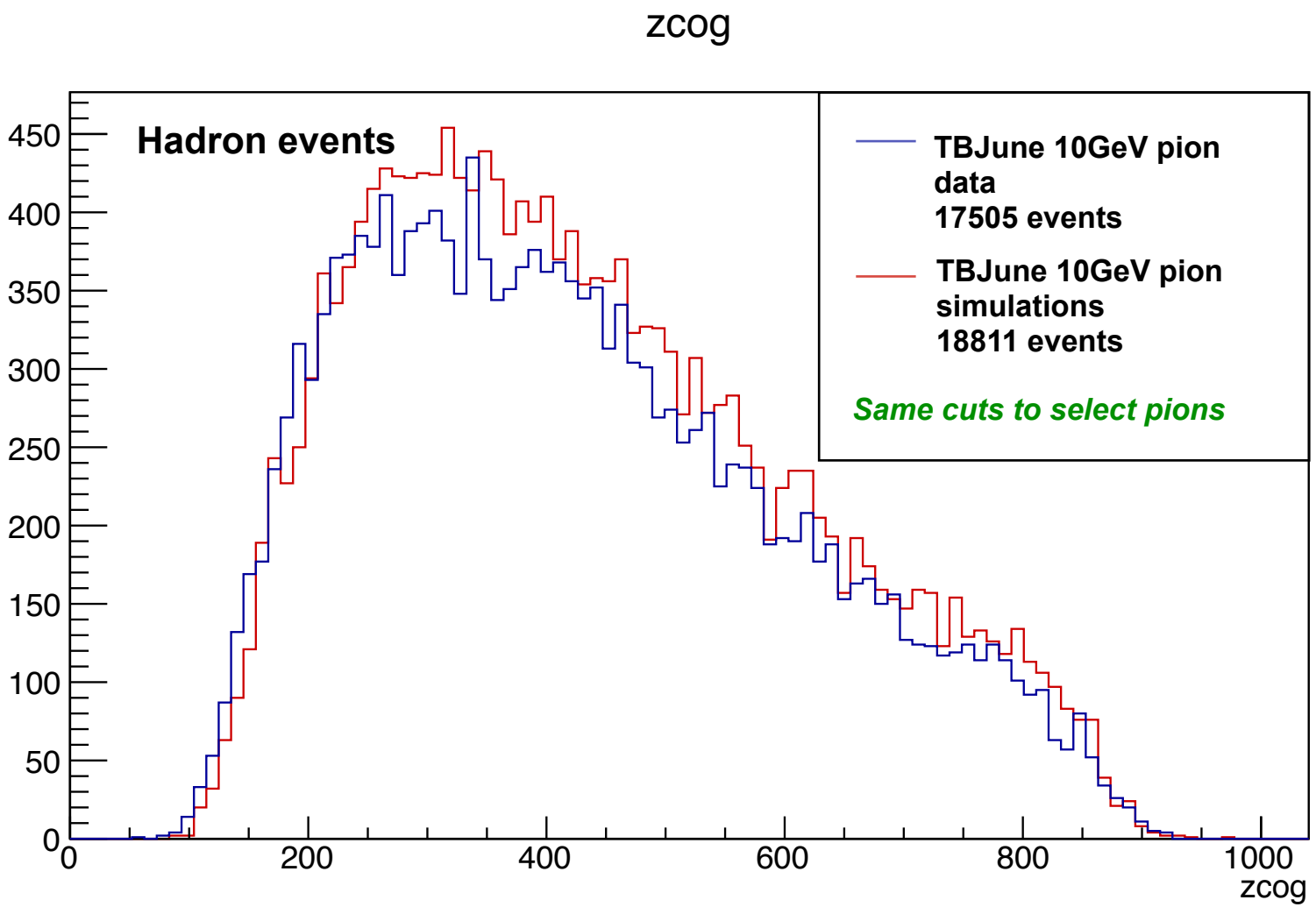
Observables

For particle ID

Center of gravity in z

- μ^- : z_{cog} peaks in the middle of detector
- e^- : z_{cog} peaks in the first half of detector
- π^- : z_{cog} more-less spread





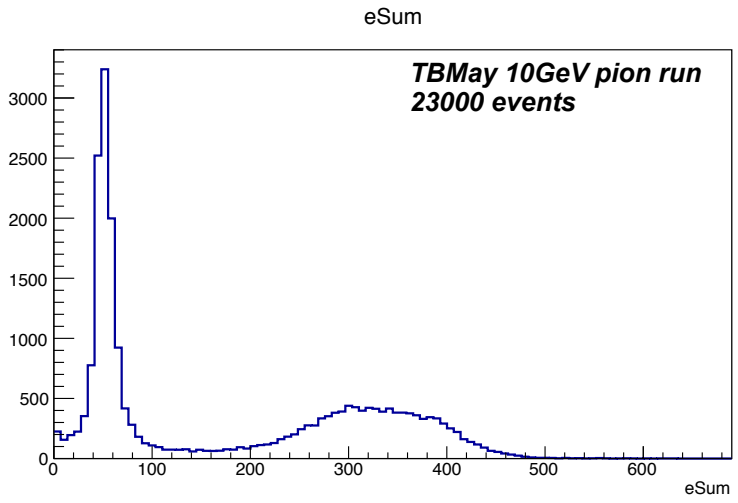
Observables

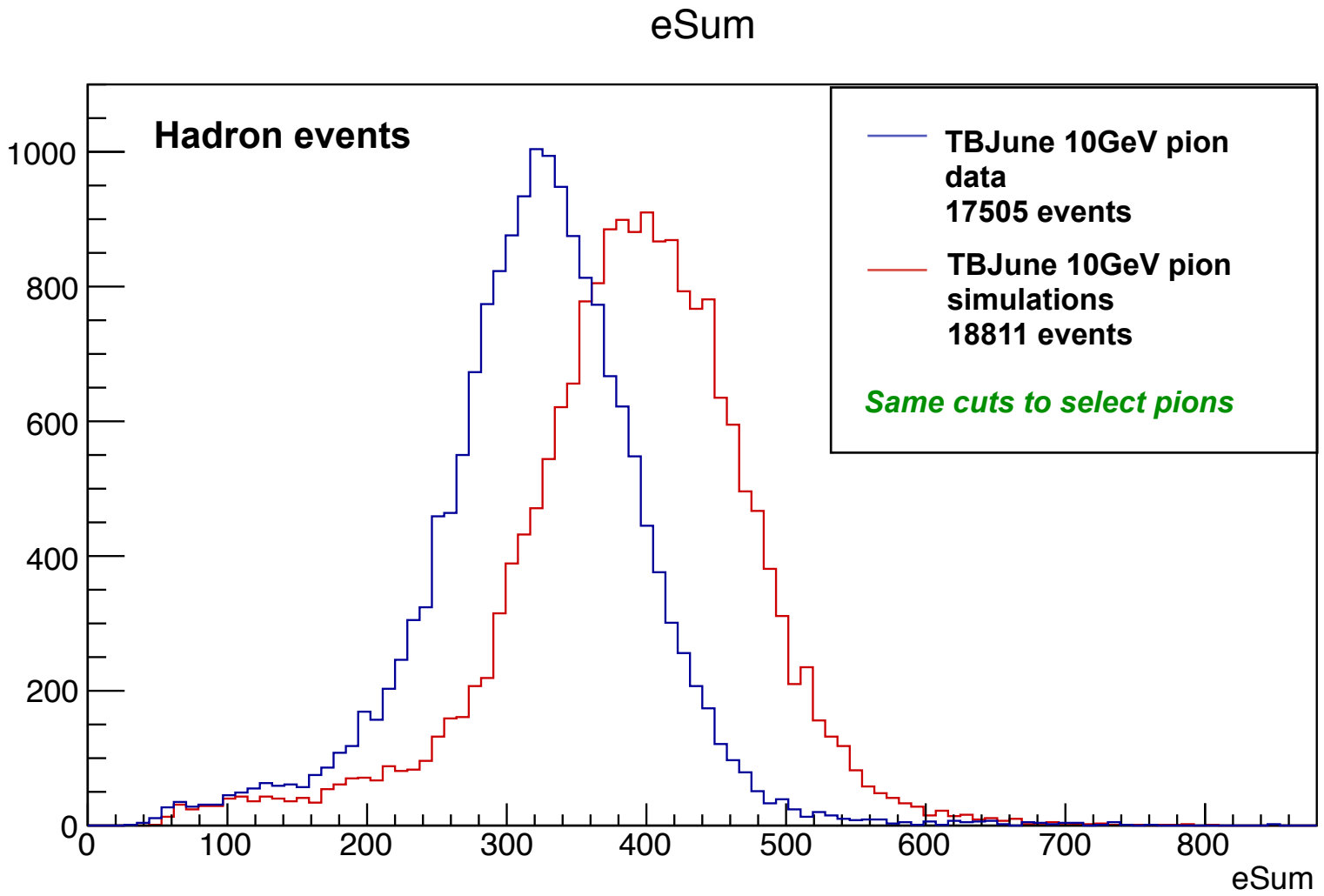
For particle ID

Energy sum per event

- "clean" μ^- : $E_{sum} \approx N_{layers} * MIP$

Bad agreement MC with data for the moment (not used).
Probably, IC problem





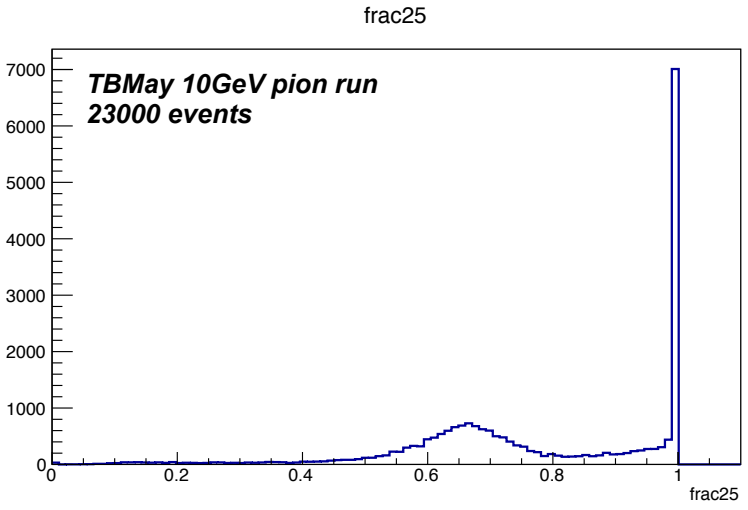
Observables

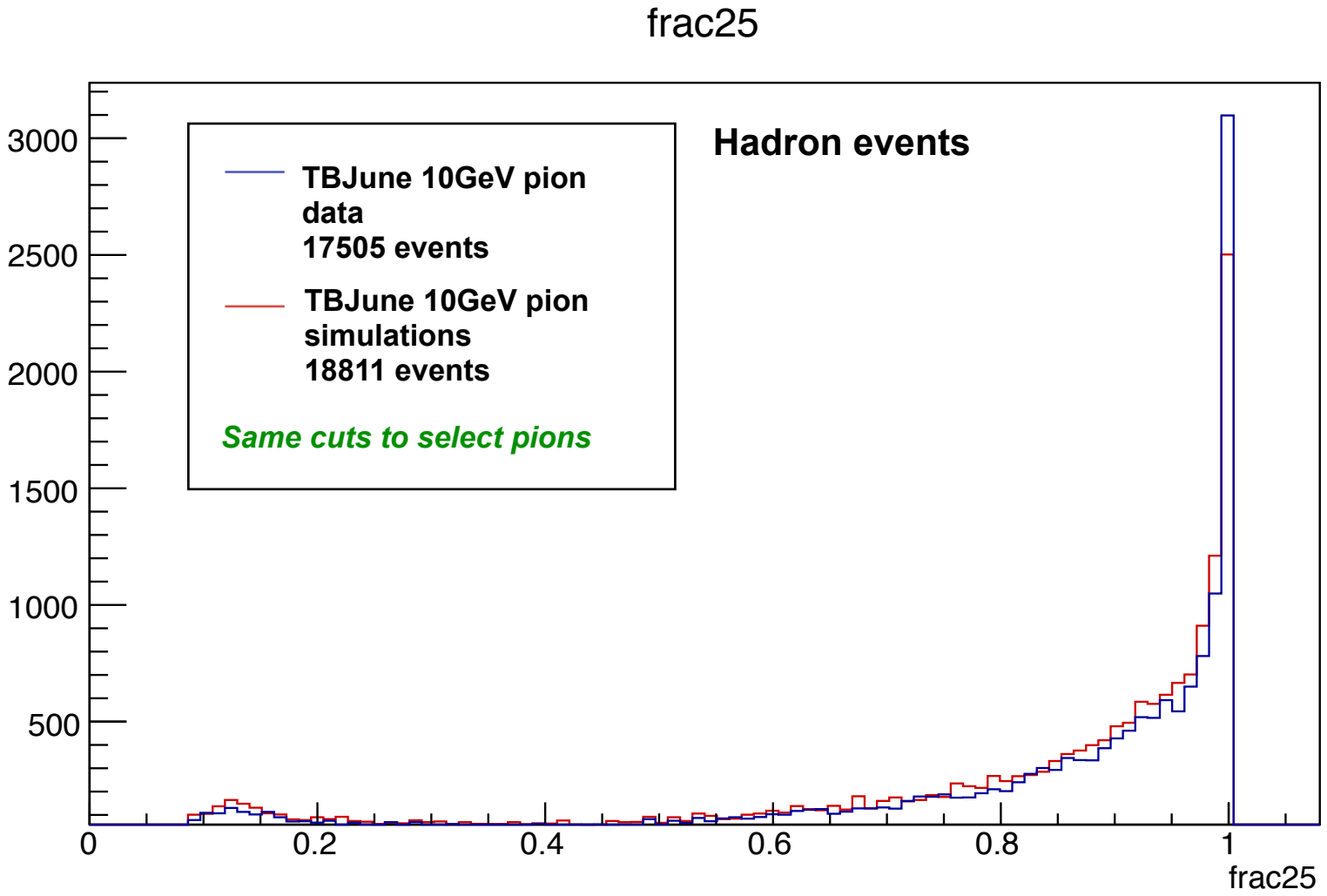
For particle ID

Fraction of energy in first 25 layers

$$(E_{25l}/E_{total})$$

- μ^- : fraction peaks at $25/N_{layers}$
- e^- : fraction ≈ 1
- π^- : fraction more-less spread



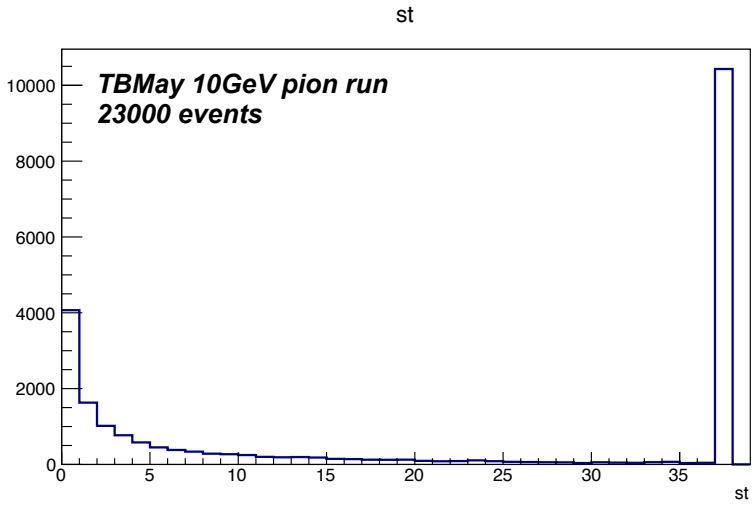


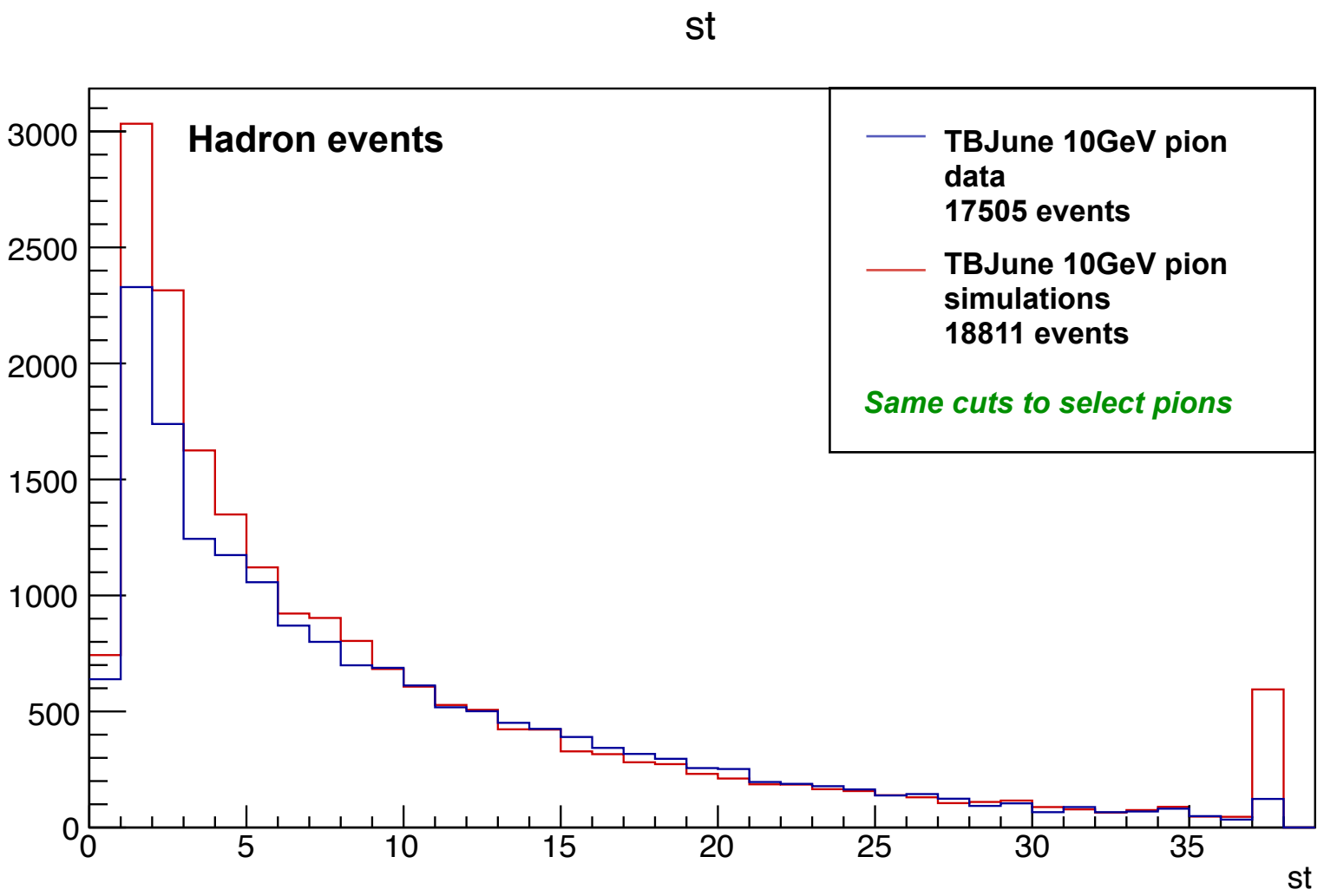
Observables

For particle ID

Shower start layer number

- "clean" μ^- : no shower
- e^- : in the first half of detector
- π^- : more-less spread



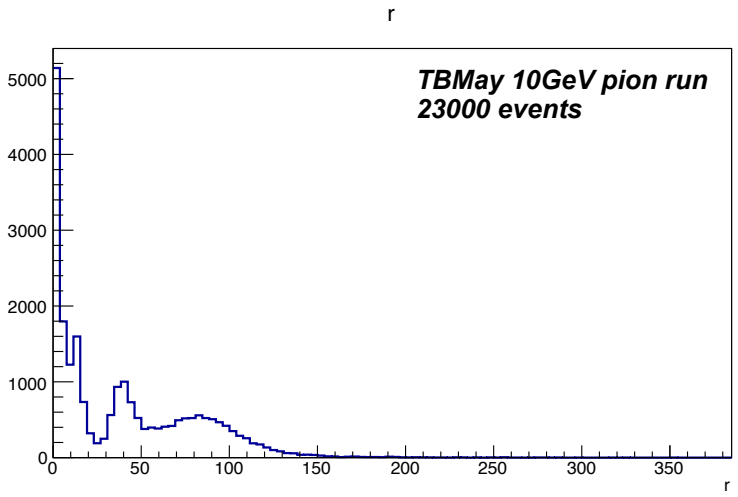


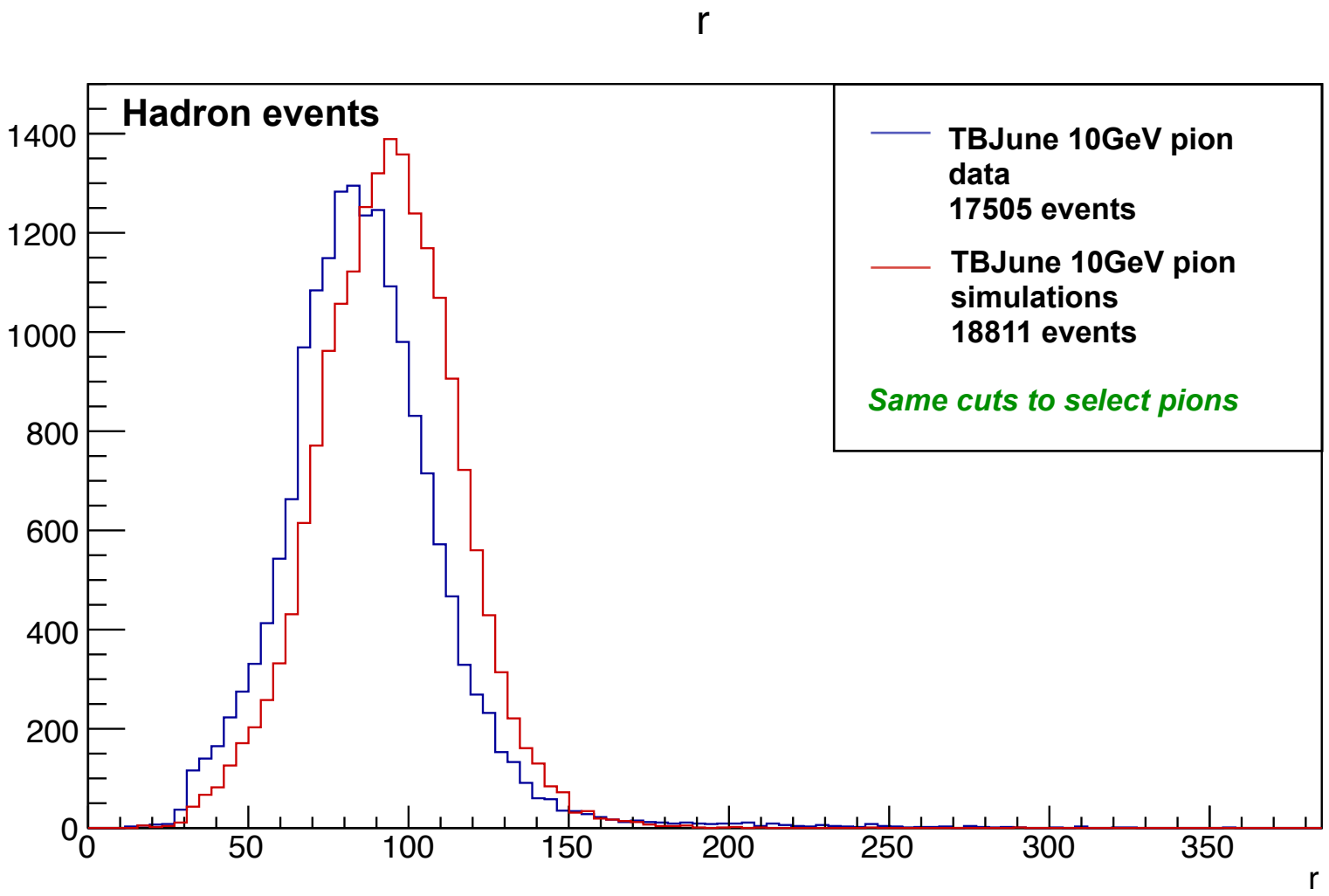
Observables

For particle ID

Shower radius

- "clean" μ^- : no shower ($R = 0$)
- e^- : R has a peak
- π^- : more-less spread





Particle ID

Cut-based method

Method* developed for physics prototype by Marina Chadeeva was taken as basement:

Calorimeter-based id by topology and cluster properties

Clustering and shower start finding precedes particle identification.
Set of cut-based topological discriminants is based on published CANs and papers.

- **Multi-particle events**: OR-based combination of discriminants
  - several clusters in first layers
  - parallel tracks
  - too high deposition
- **Muons or muon-like**: complex combination of discriminants
  - number of hits
  - energy comparison in calorimeters
  - track finding
- **Electrons**: AND-based combination of discriminants
  - cluster radius <37 mm
  - CoG depth $<10 \cdot X0$
  - 90% of measured energy deposited in first $24 \cdot X0$
  - shower start before $6 \cdot X0$
- **Hadrons**:
  - remaining species after above rejection
  - Cherenkov-based discrimination of hadron types

*https://agenda.linearcollider.org/event/7454/contributions/38731/attachments/31380/47203/chadeeva_Fe-W_LLR2017.pdf

Current cuts

For particle ID (10 GeV particles)

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

Empty events(optional):

- **Hits**: nHits < nHits_min

Muon (muon-like) events:

- **Hits**: 0 < nHits < 70
- **(&)Energy:** 0 < E < 5 GeV (will be used after proper calibration)
- **(&)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**.

Note: multi-particle event filtering not implemented yet (in todo list).

Performance

Of cut-based ParticleID method for 10GeV particles.

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

>12 main cut values making tuning of selection efficiency complicated

BDT method.

BDT input

10GeV pion selection.

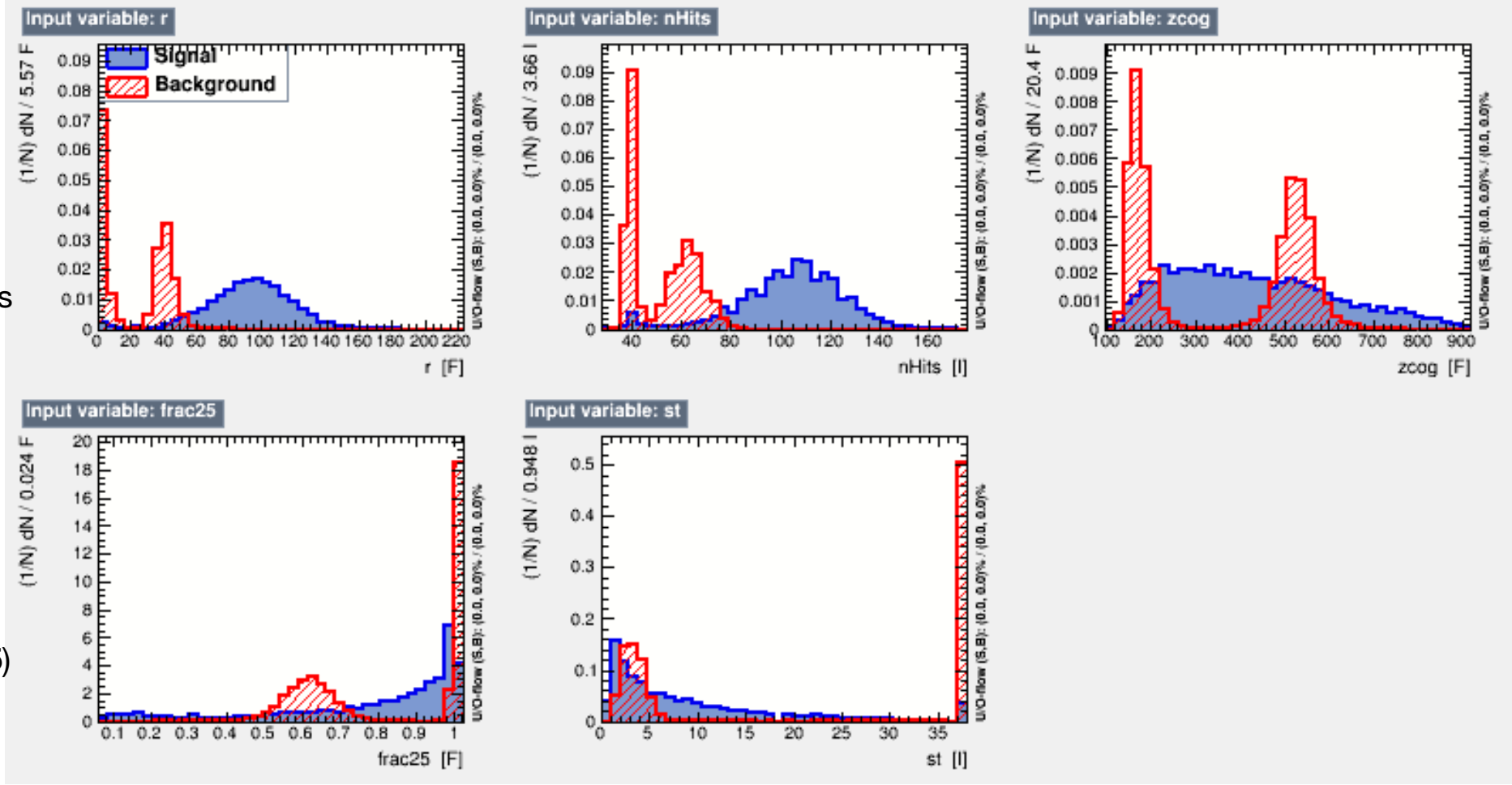
Training data:

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

Variables:

- Shower radius (**st**)
- Number of hits (**nHits**)
- COGz (**zcog**)
- Fraction in first 25 layers (**frac25**)
- Shower start layer (**st**)

No cuts

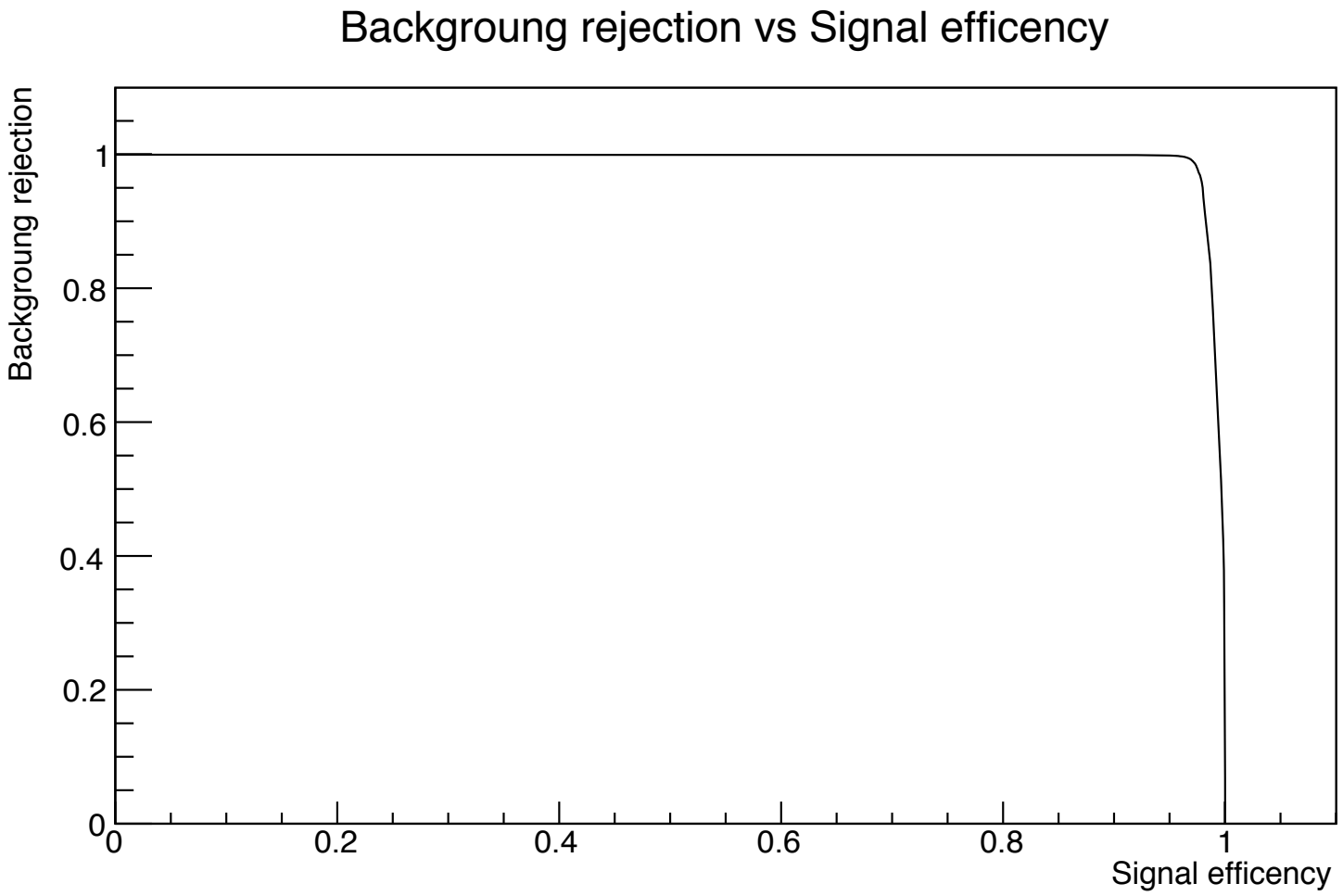


BDT output

10GeV pion selection.

Test data:

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

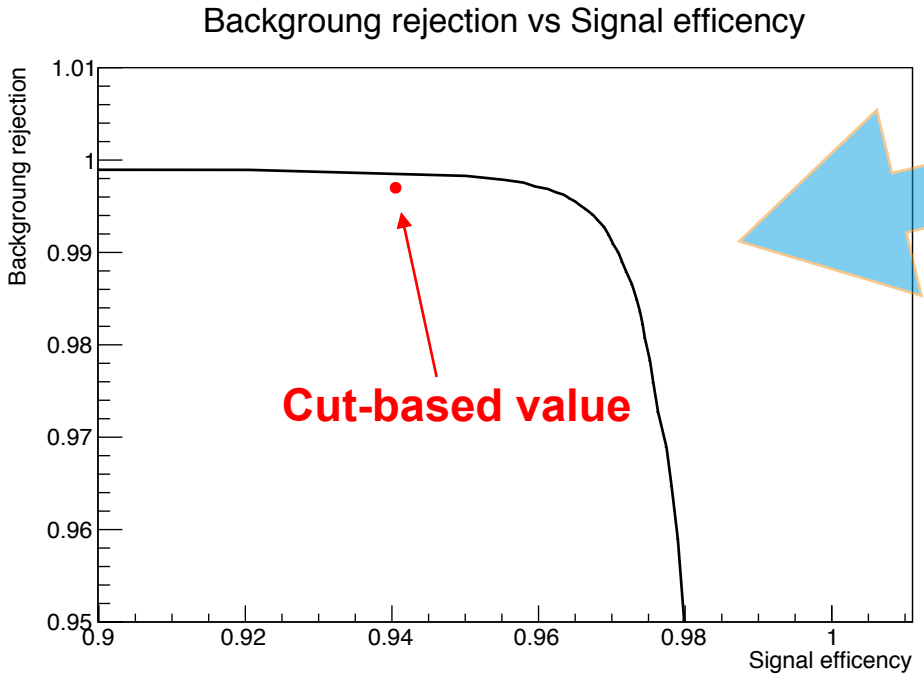


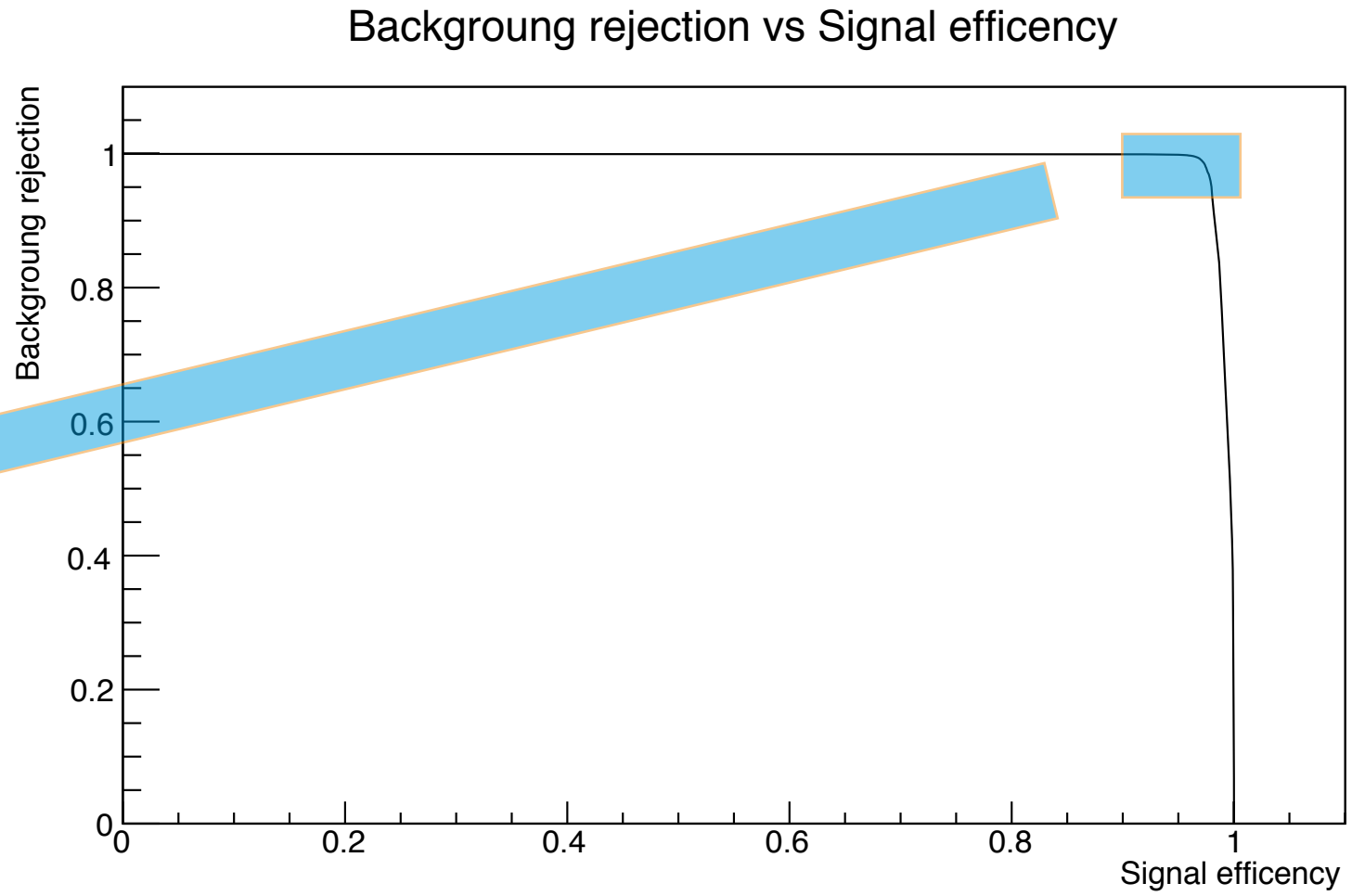
BDT output

10GeV pion selection.

Test data:

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





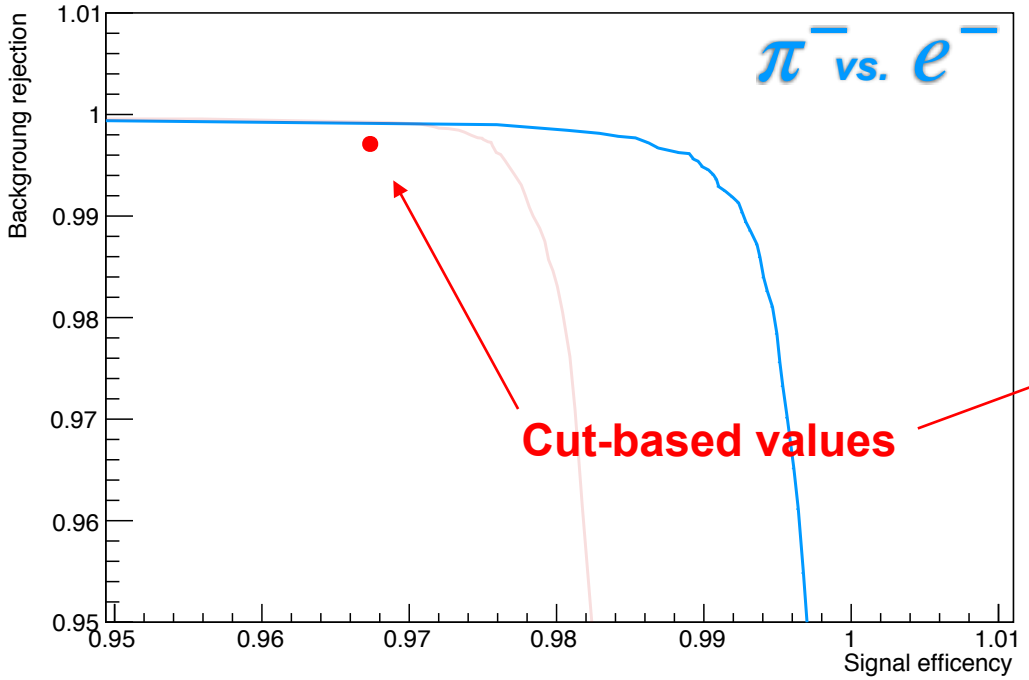
Pions vs electrons & pions vs muons

BDT output.

Training/test data:

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

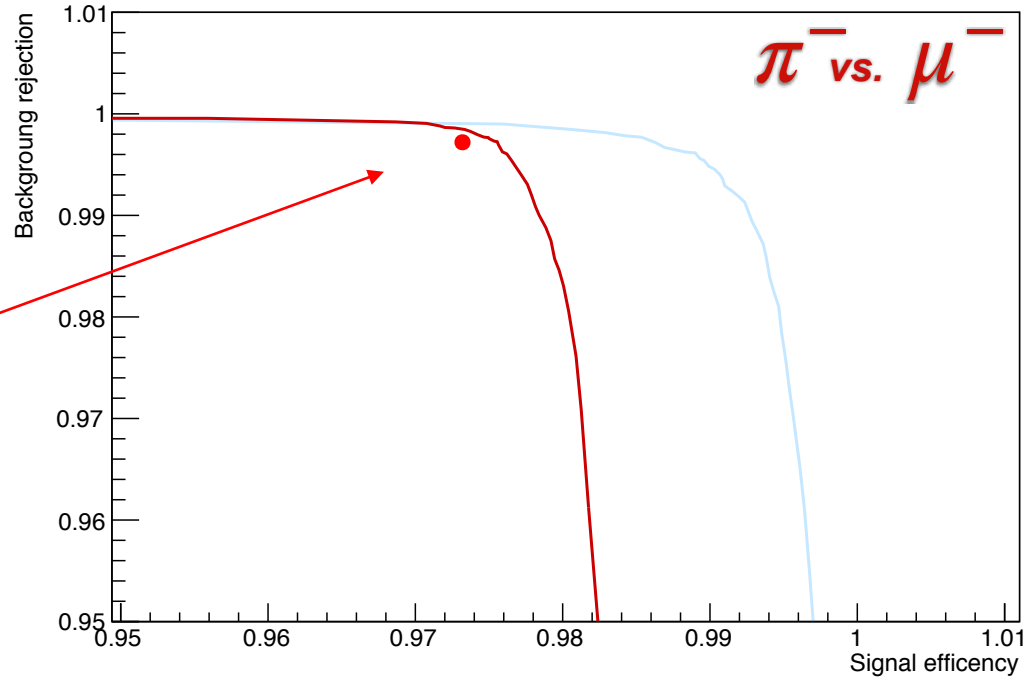
Backgroung rejection vs Signal efficency



Training/test data:

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

Backgroung rejection vs Signal efficency



Pions vs electrons & pions vs muons

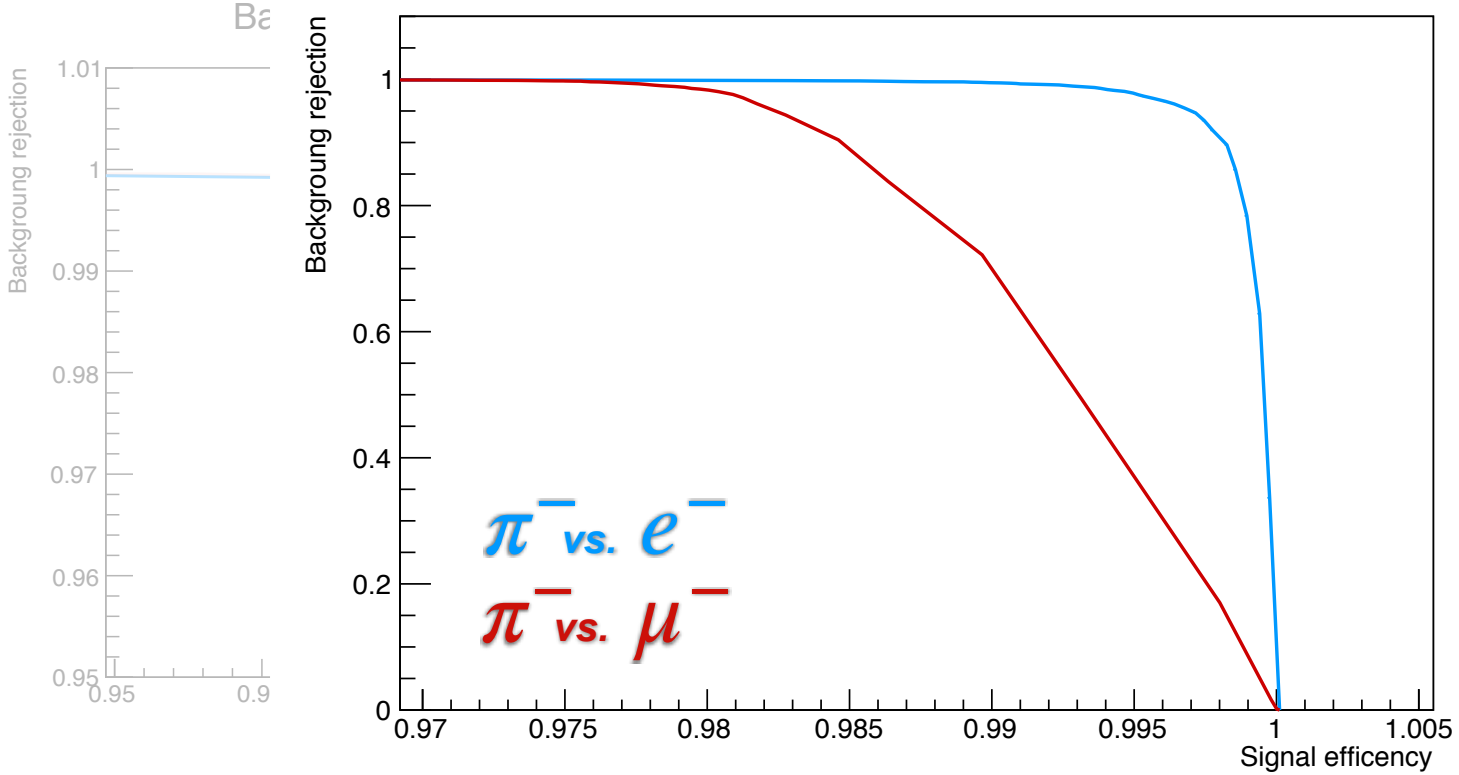
BDT output.

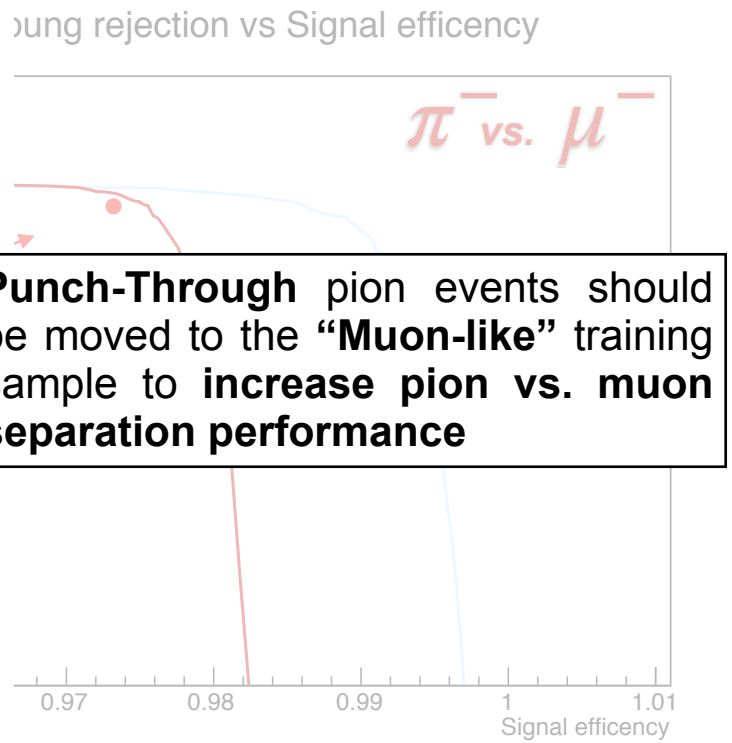
Training/test data:

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

Training/test data:

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

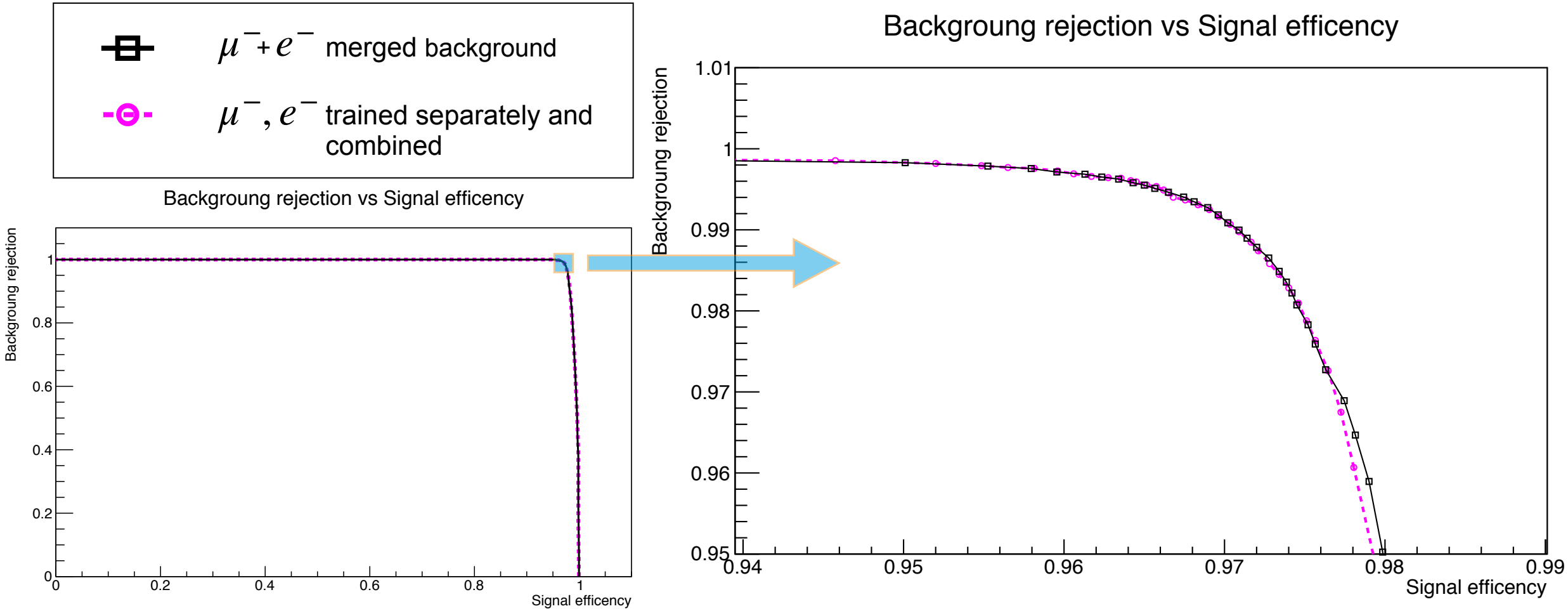




Punch-Through pion events should be moved to the **"Muon-like"** training sample to **increase pion vs. muon separation performance**

Separation of pion “background”

Comparison of different BDT output



Confluence page

TBParticleID processor

Content:

- Description including using variables
- Input
- Output
- Important parameters

link: https://confluence.desy.de/display/Calice/TBParticleID+processor

Important parameters:

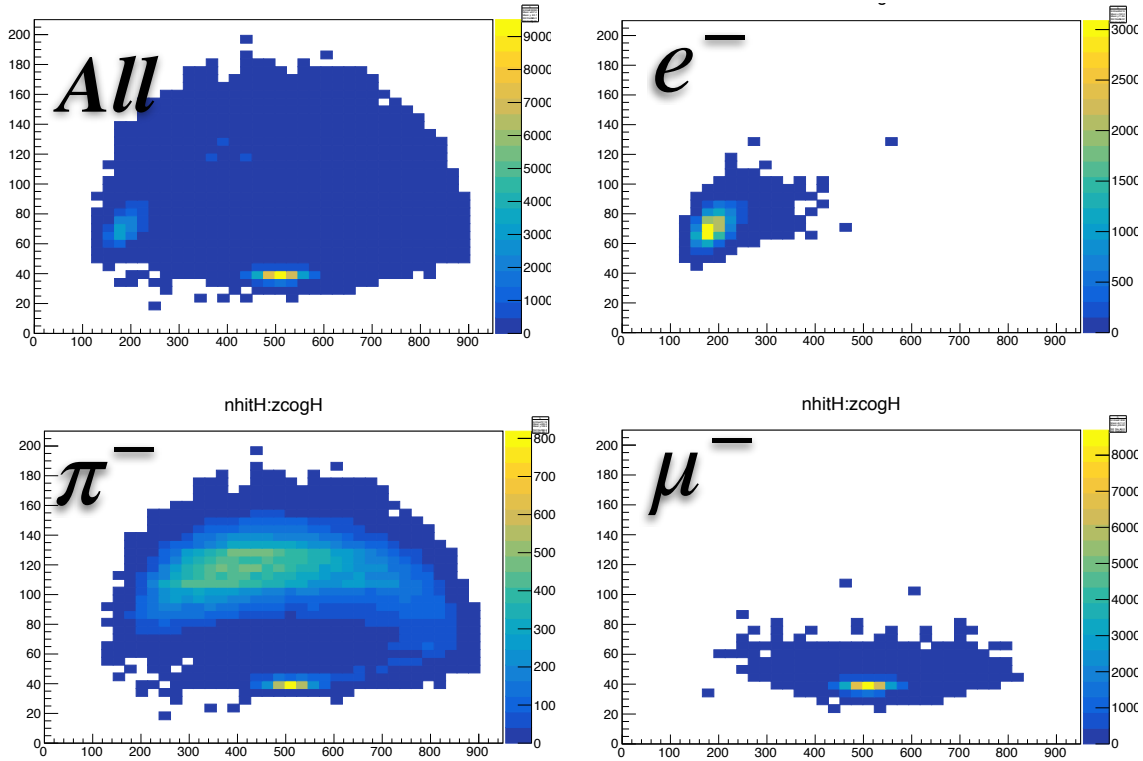
- Beam energy in GeV - using only to find shower start layer
- IDMode: 0 - keep all, 1 - only hadron events, 2 - only muon-like events, 3 - only electron events
- Minimum number of hits - to reject empty events
- Path to IDCuts.txt file

Adjustment of cuts

2D plots are useful to check the cuts

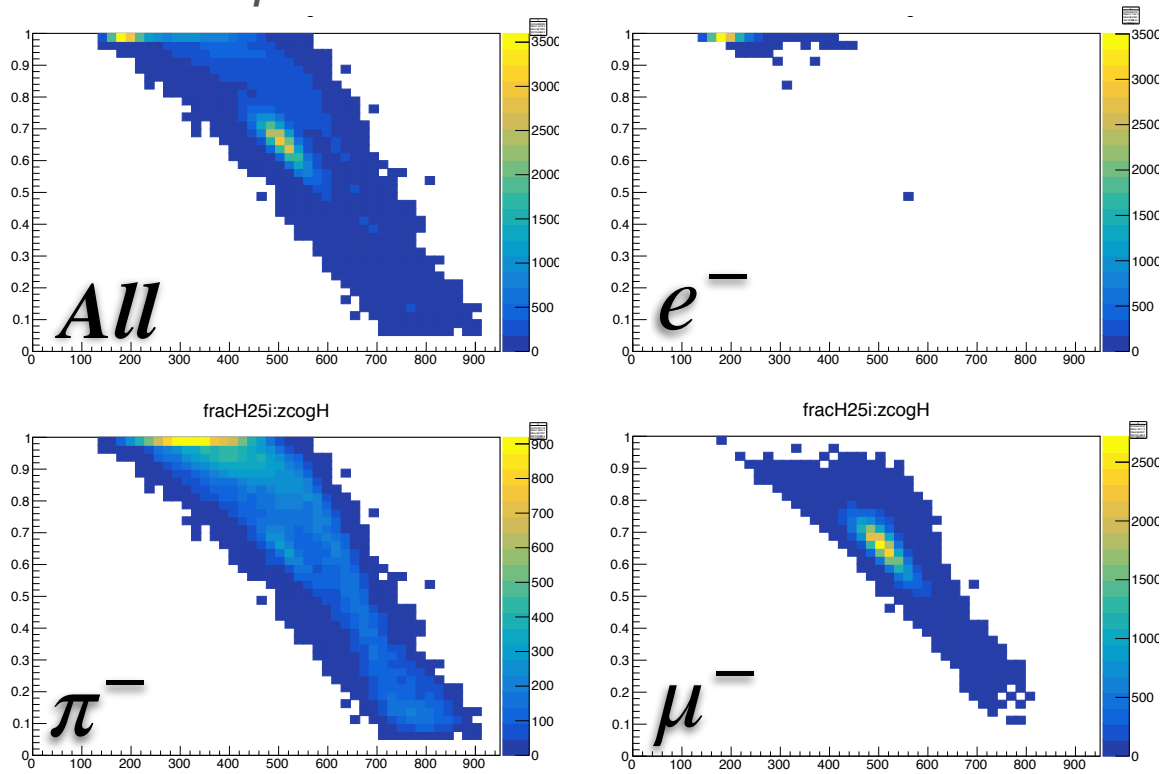
Number of hits vs CoG in z

10GeV MC particles. No cuts.



Fraction in first 25 layers vs CoG in z

10GeV MC particles. No cuts.



Root script for cut adjustment

MC 10GeV particles

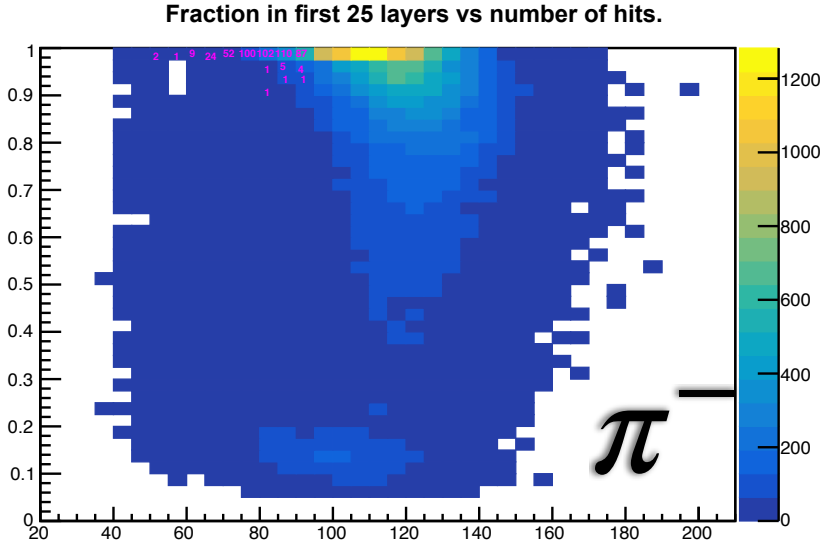
Before cuts:

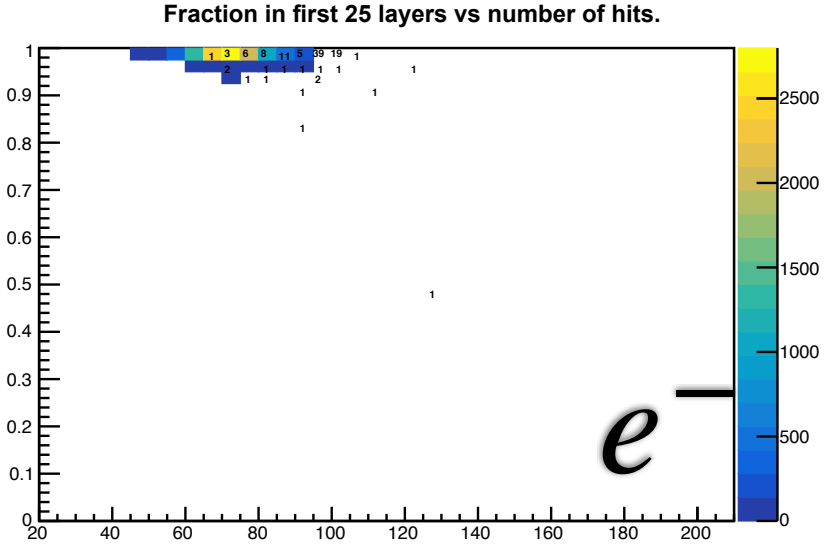
π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

$total$: 138395 events



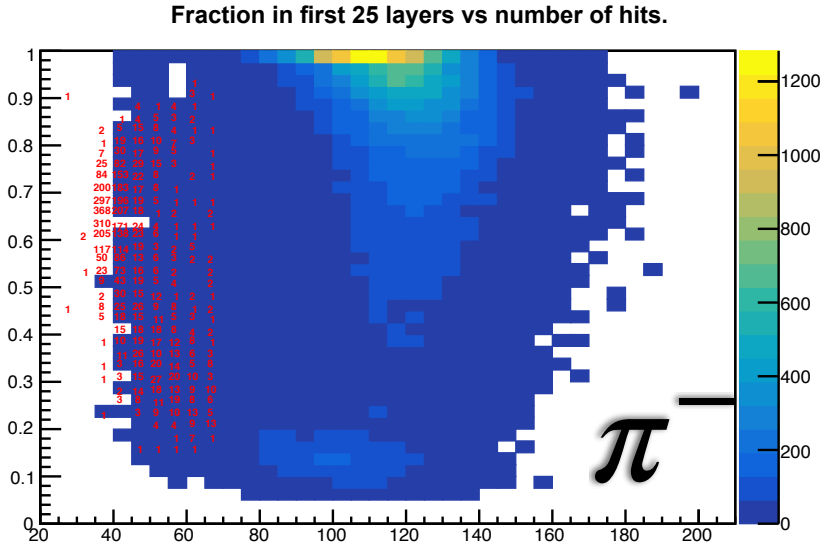


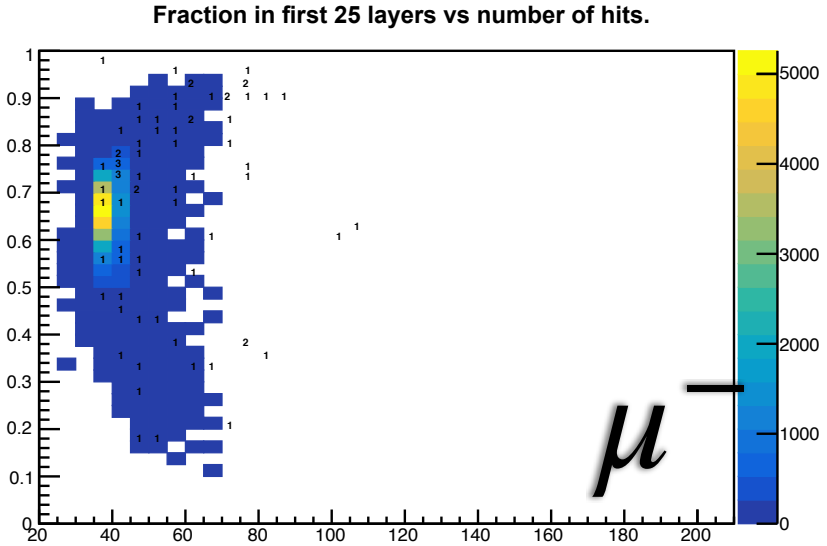
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

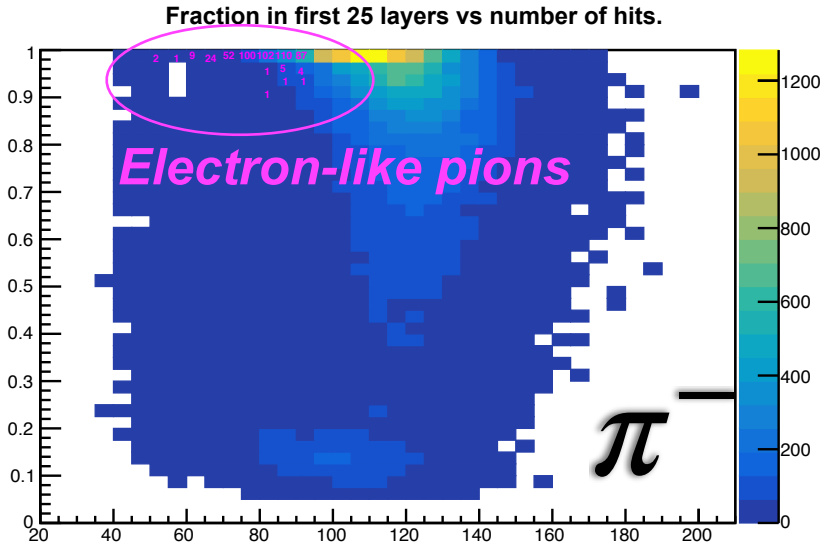
Before cuts:

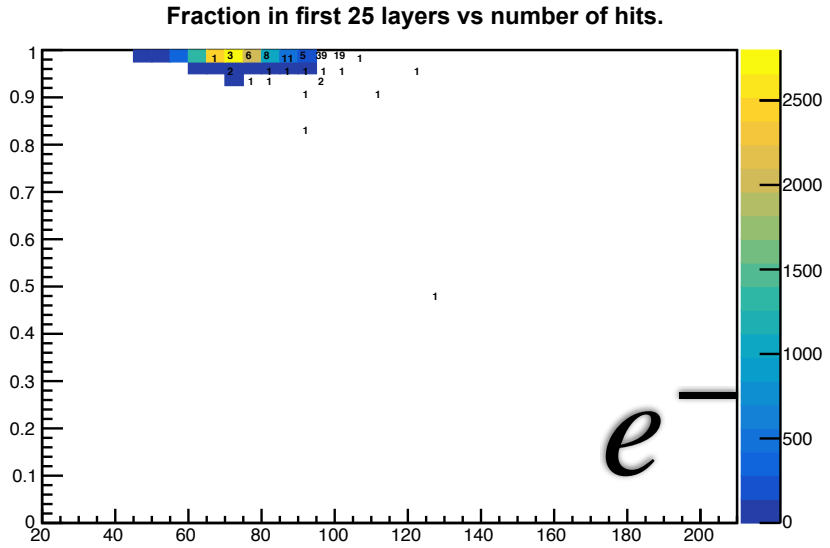
π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

$total$: 138395 events



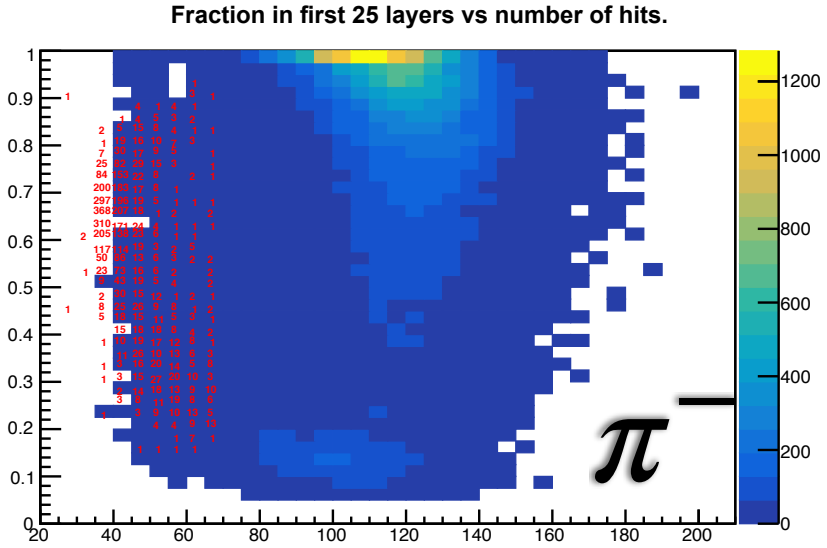


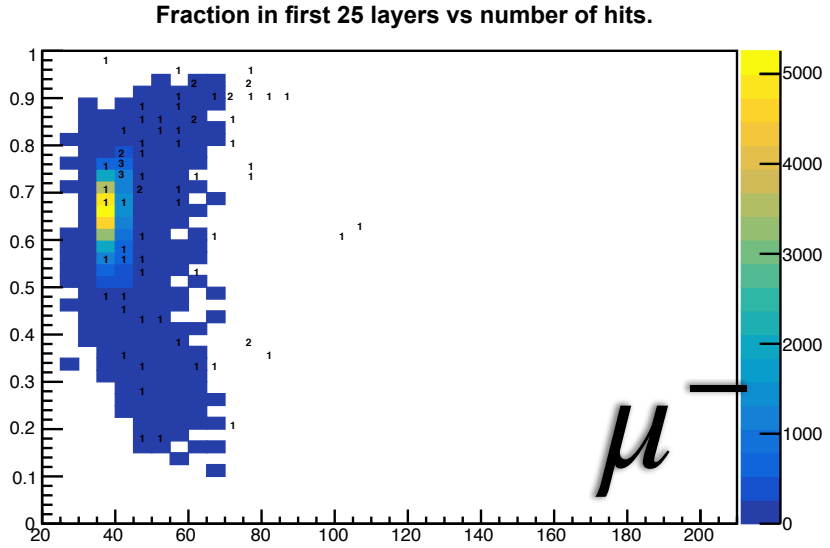
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

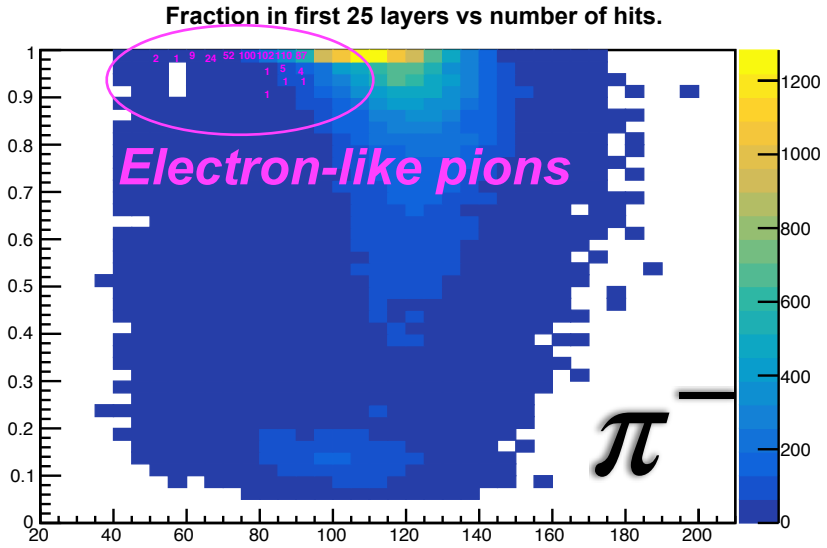
Before cuts:

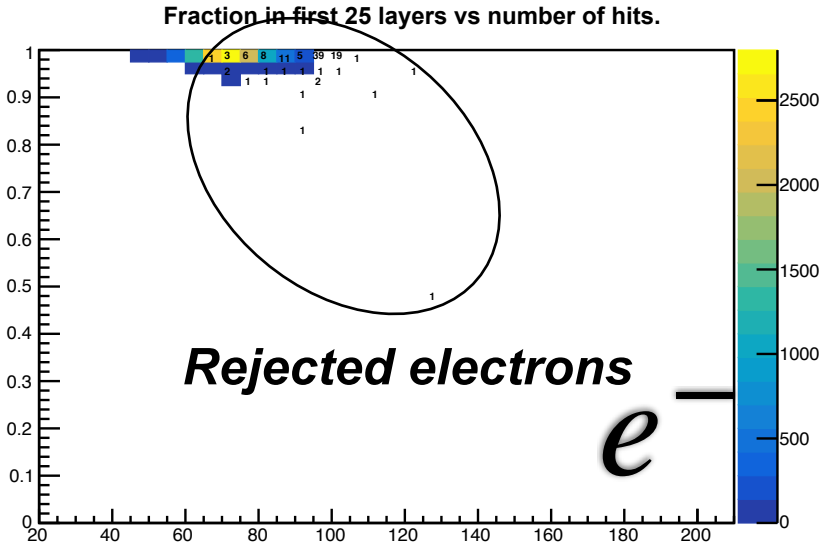
π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

$total$: 138395 events



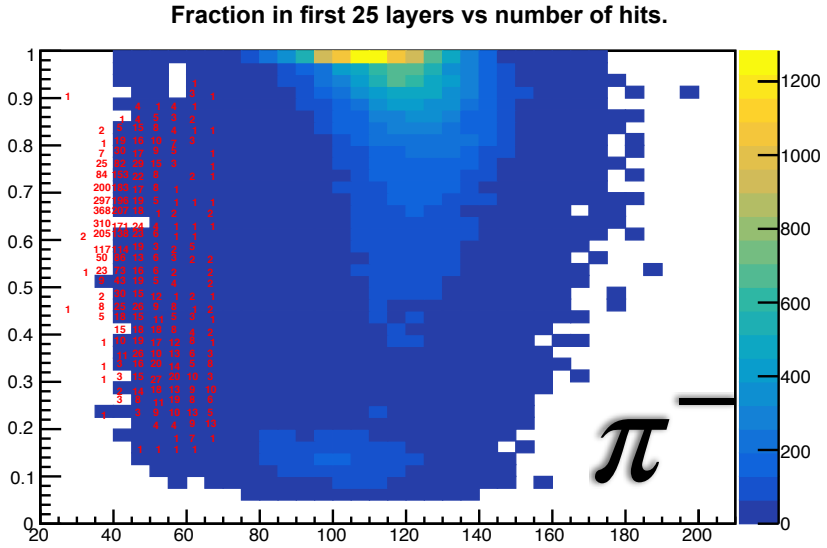


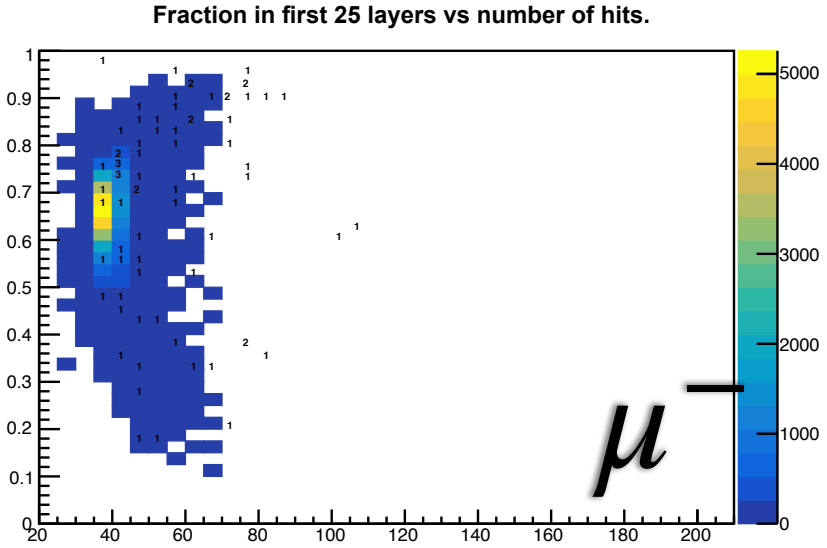
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

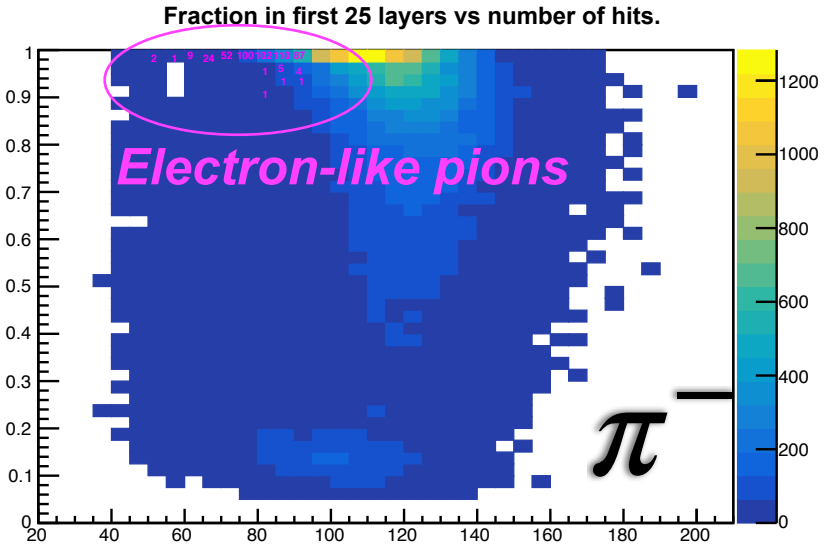
Before cuts:

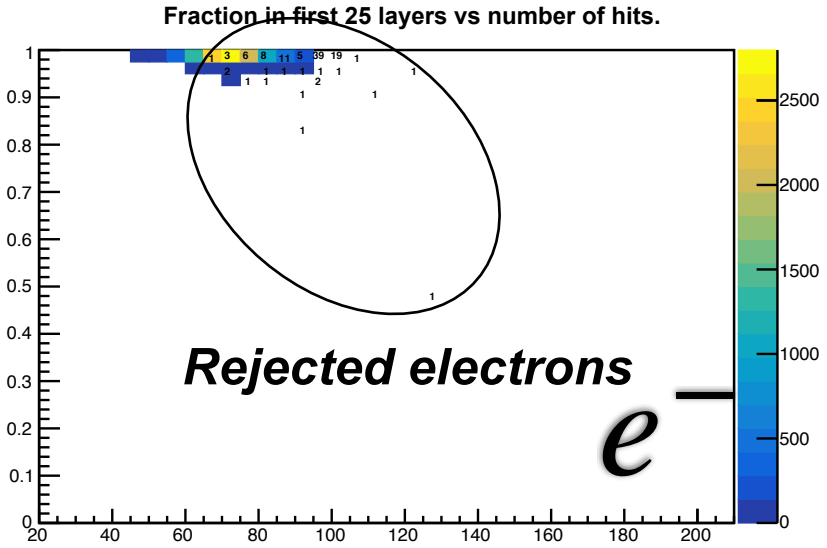
π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

$total$: 138395 events



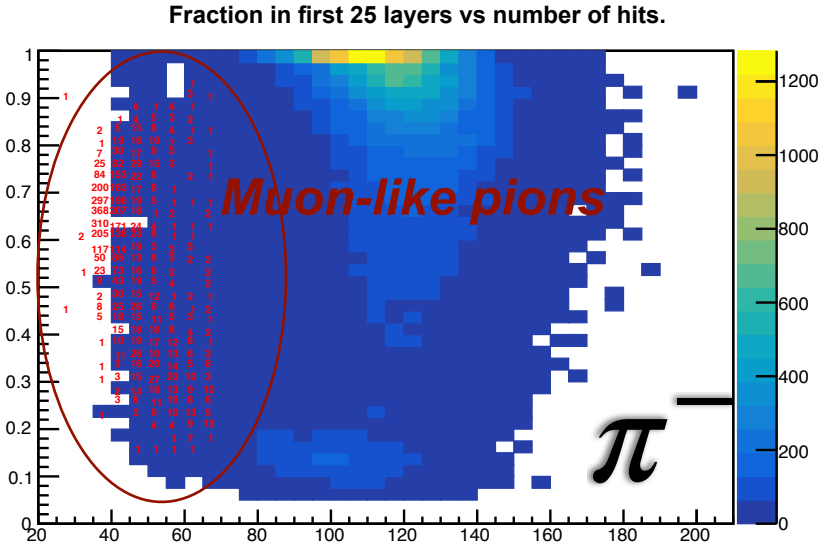


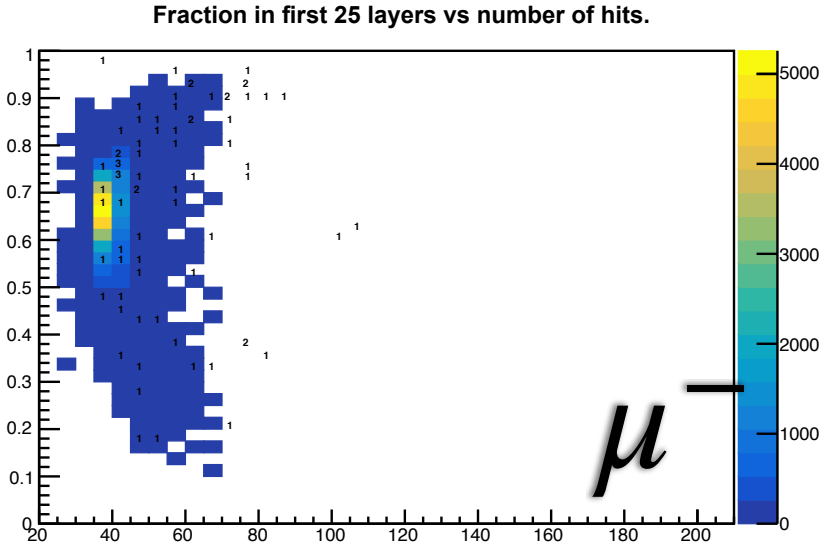
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

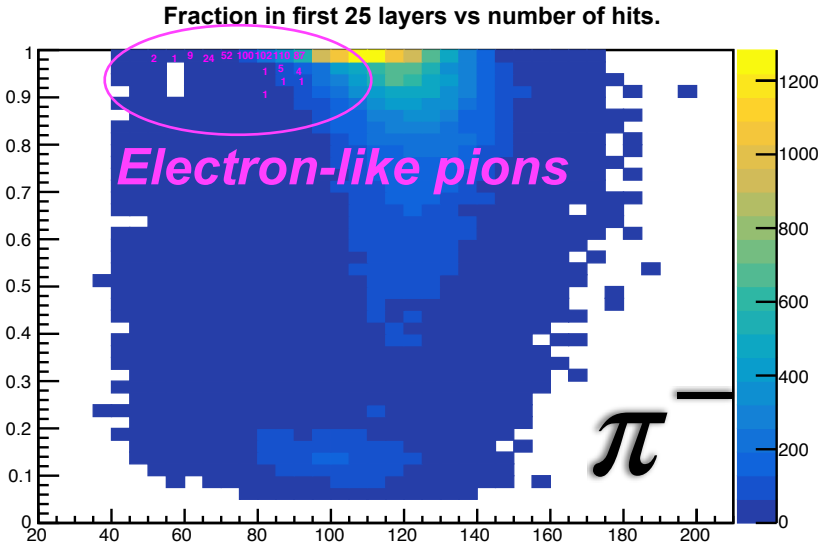
Before cuts:

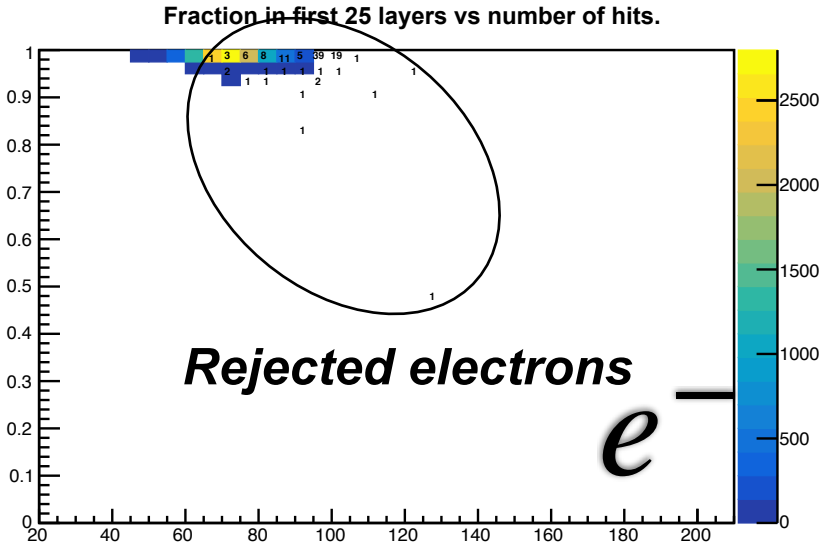
π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

$total$: 138395 events



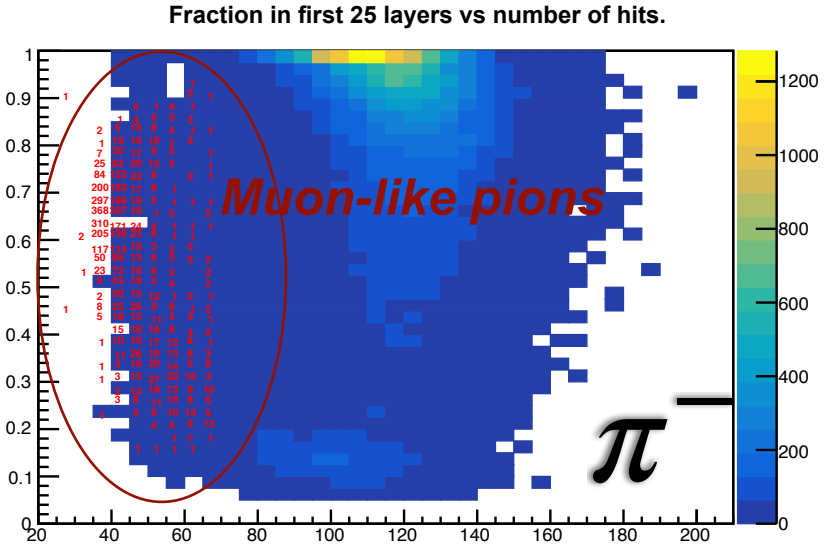


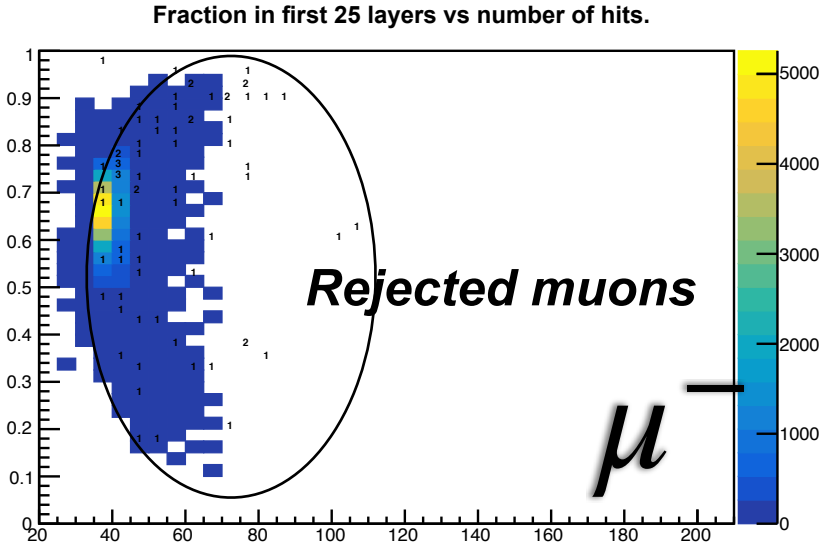
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

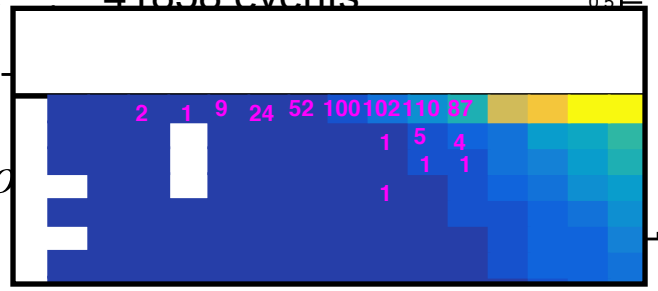
Before cuts:

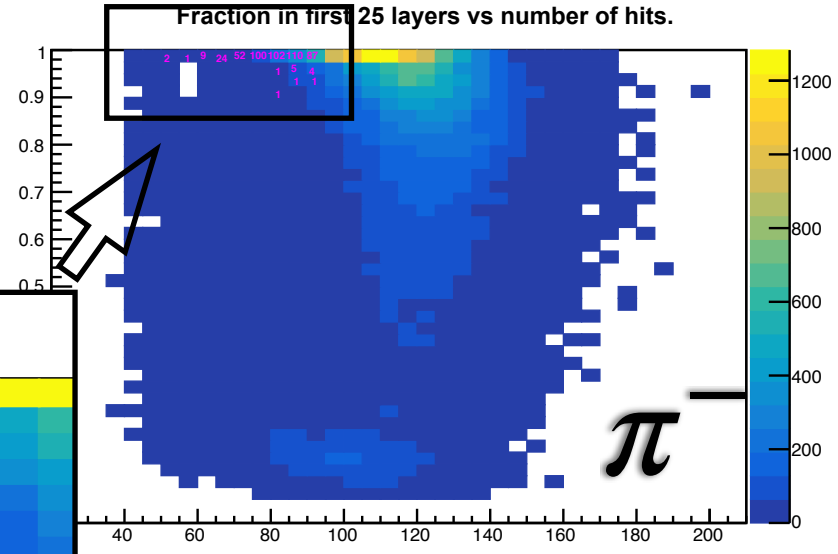
π^- : 70118 events

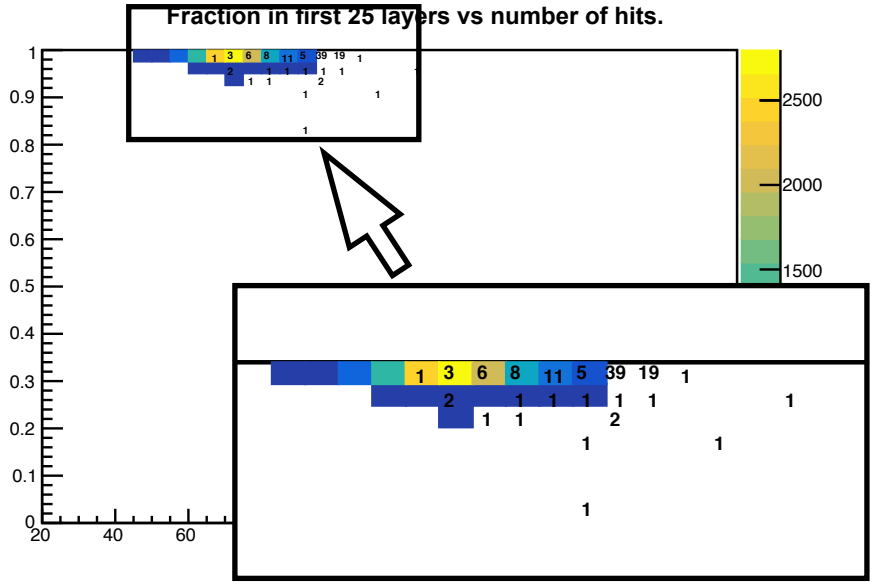
μ^- : 41858 events

e^- :

to





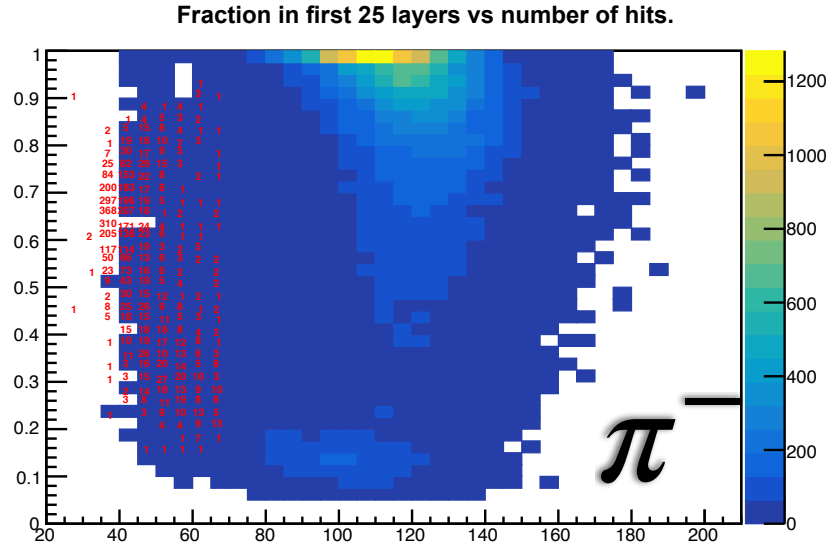


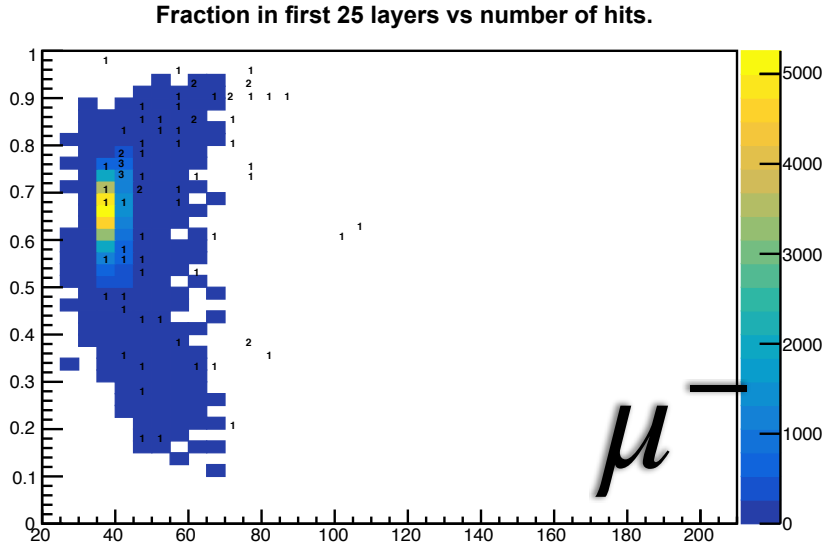
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

e^- : 26230 events (99.2%)





Root script for cut adjustment

MC 10GeV particles

Before cuts:

π^- : 70118 events

μ^- : 41858 events

e^- : 26419 events

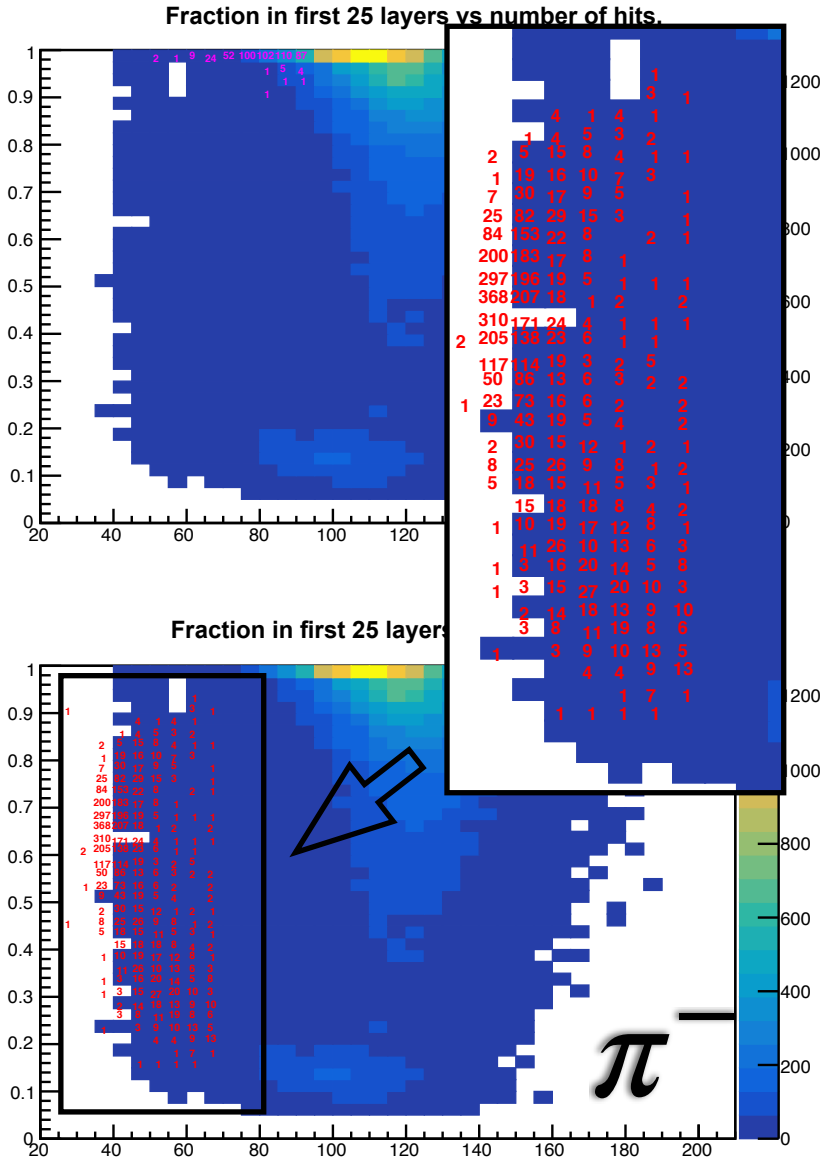
$total$: 138395 events

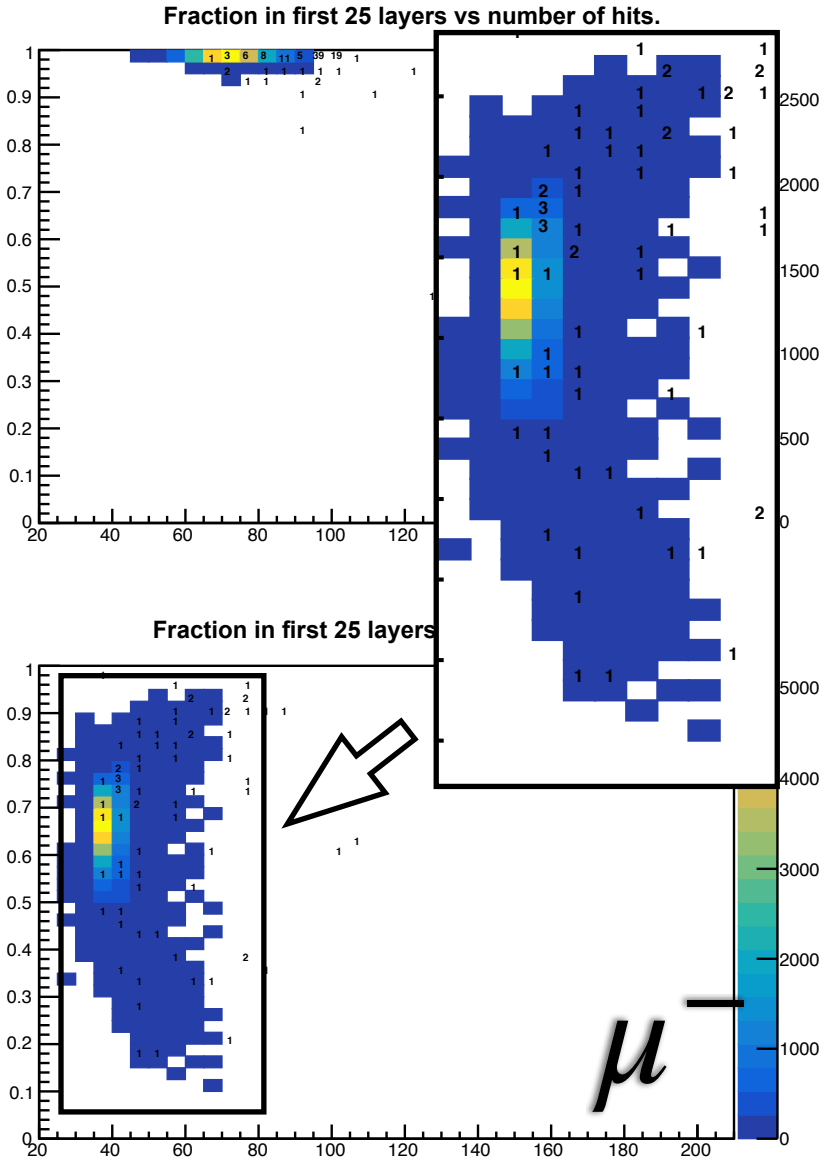
After cuts:

π^- : 64529 events (92%)

μ^- : 41783 events (99.7%)

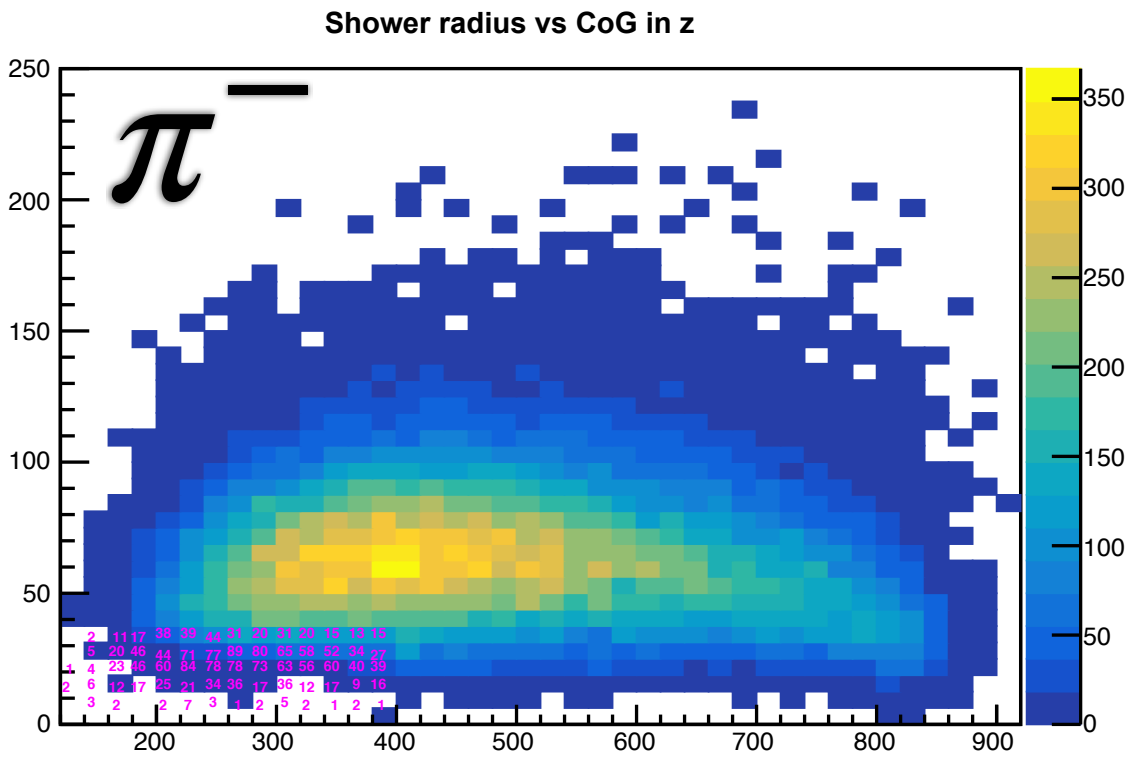
e^- : 26230 events (99.2%)

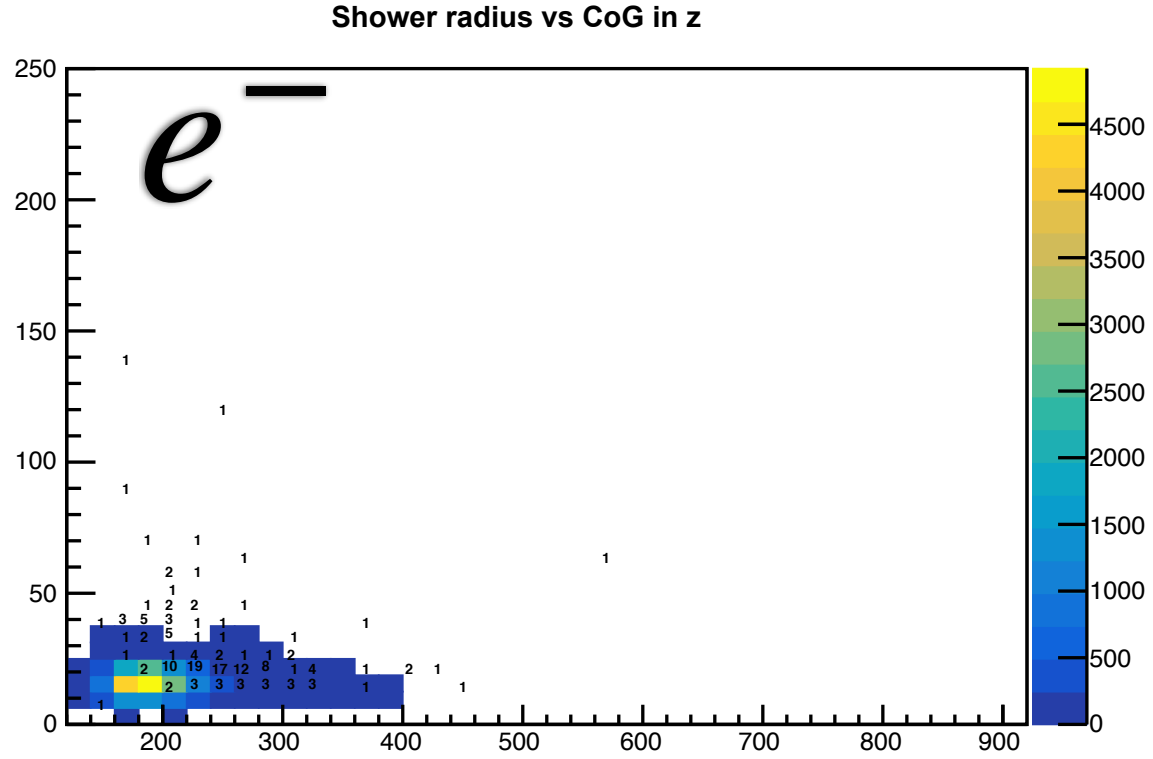




Additional cut for shower radius vs CoG in z.

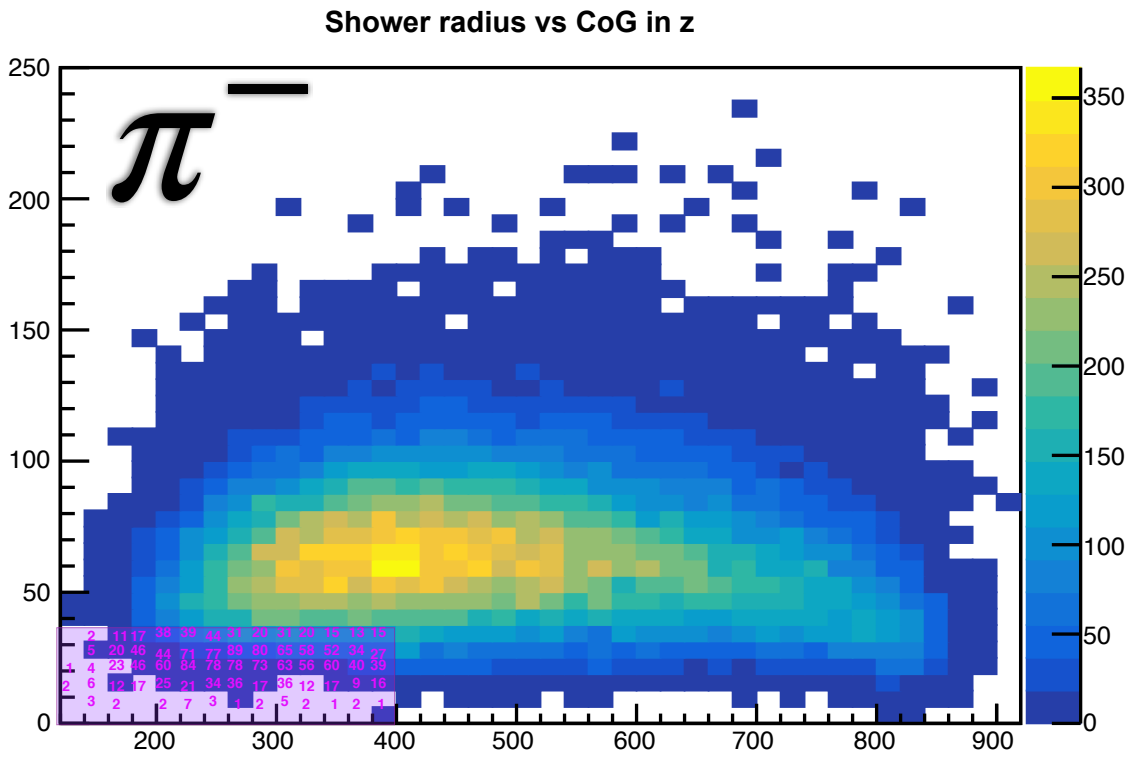
MC 10GeV particles

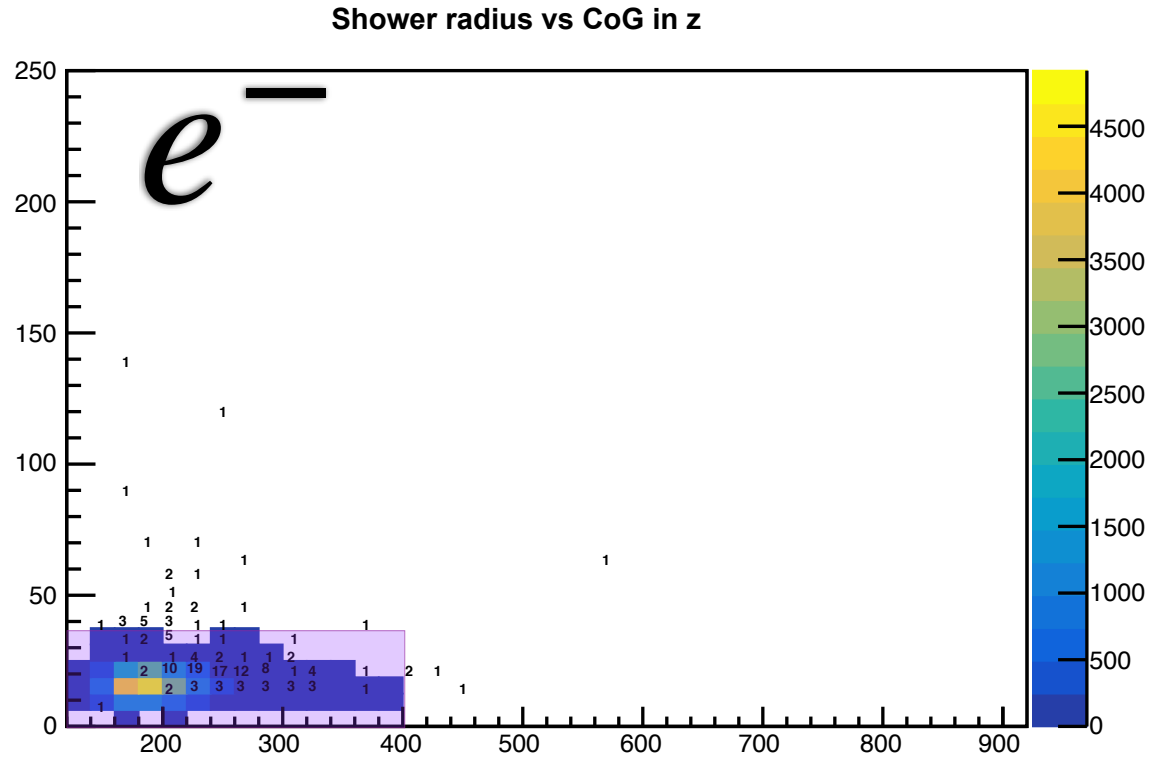




Additional cut for shower radius vs CoG in z.

MC 10GeV particles





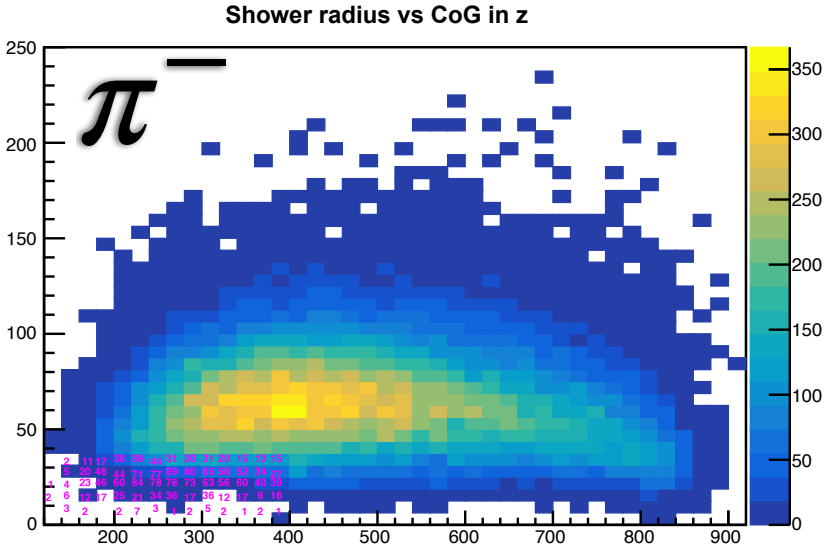
Additional cut for shower radius vs CoG in z.

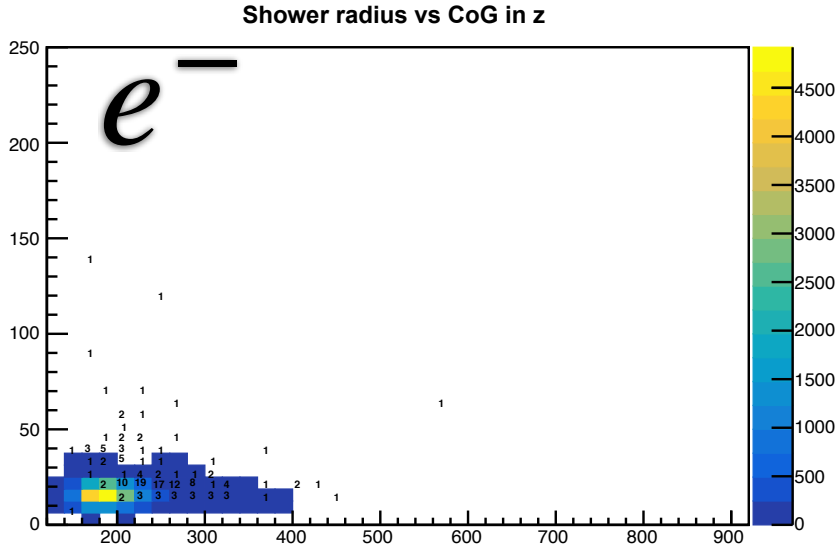
MC 10GeV particles

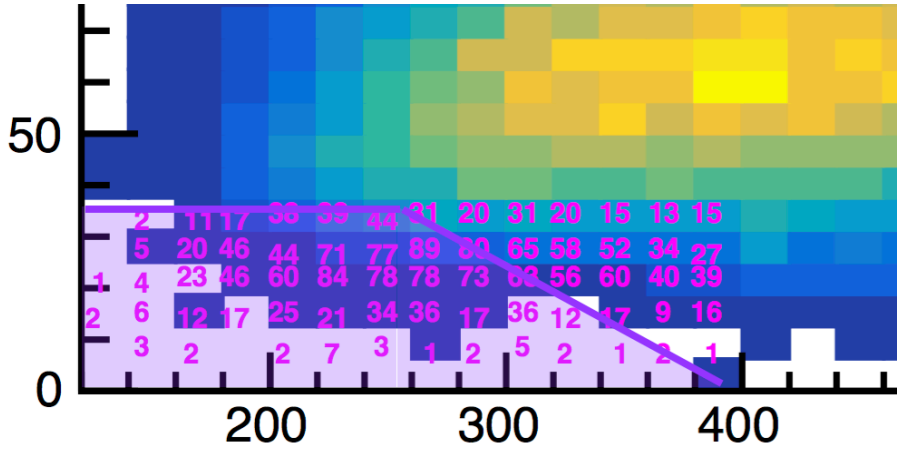
Before additional cut:

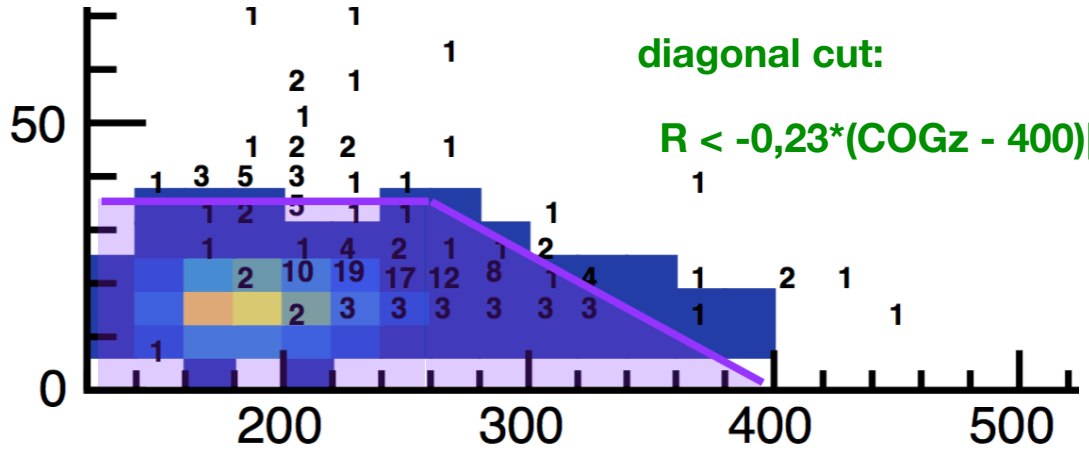
π^- : 63748 events (90,9%)

e^- : 26264 events (99.4%)









- **Additional shower radius diagonal cut:**

R < -0,23*(COGz - 400)[mm]

Additional cut for shower radius vs CoG in z.

MC 10GeV particles

Before additional cut:

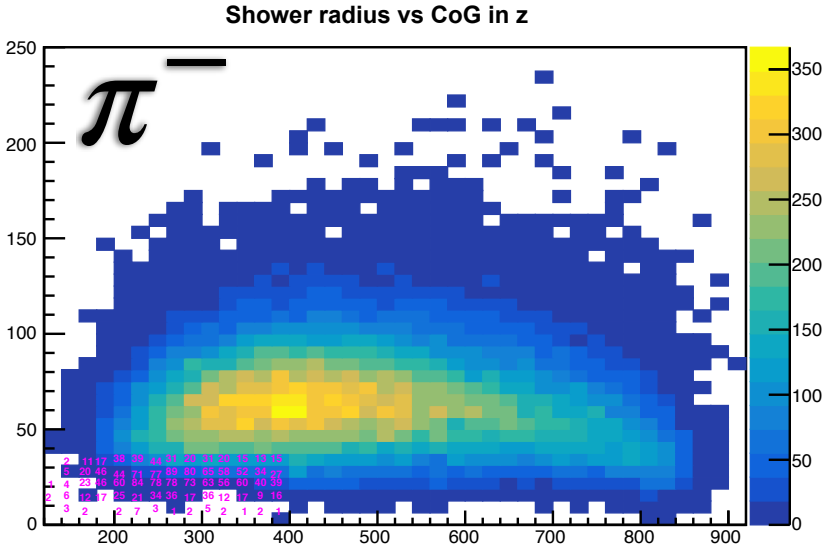
π^- : 63748 events (90,9%)

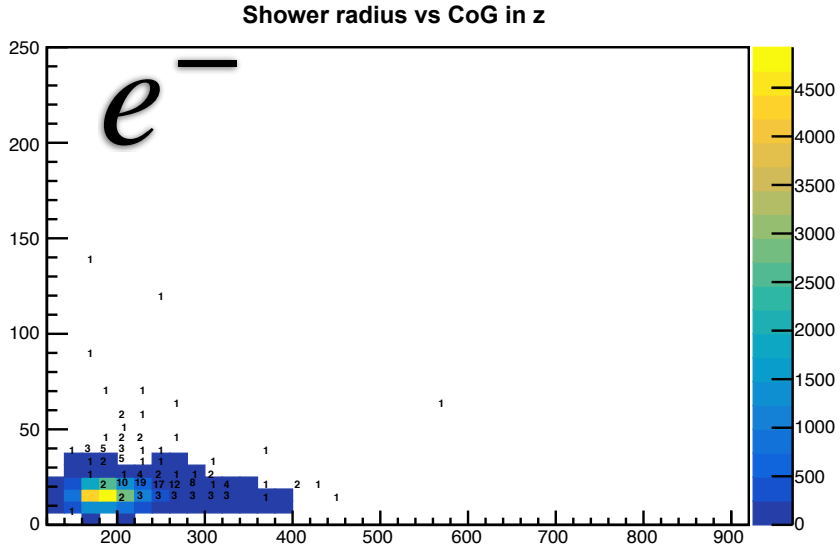
e^- : 26264 events (99.4%)

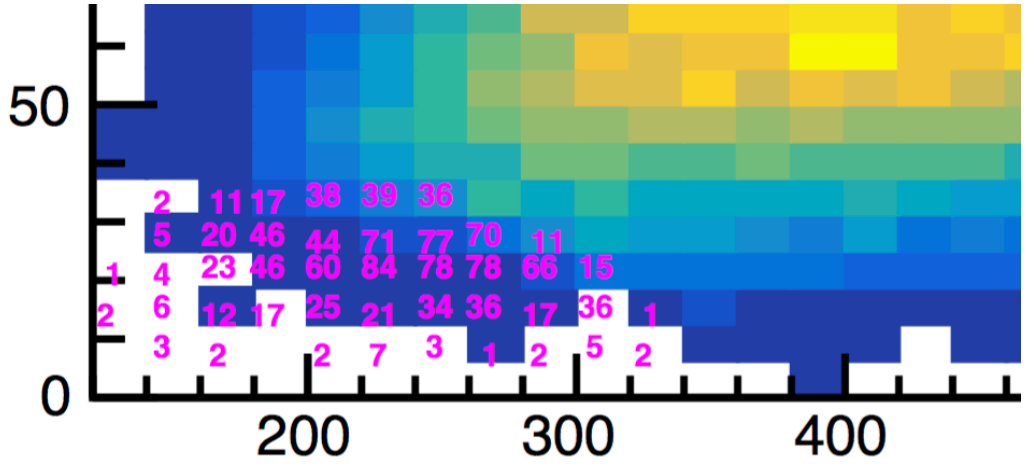
After additional diagonal cut:

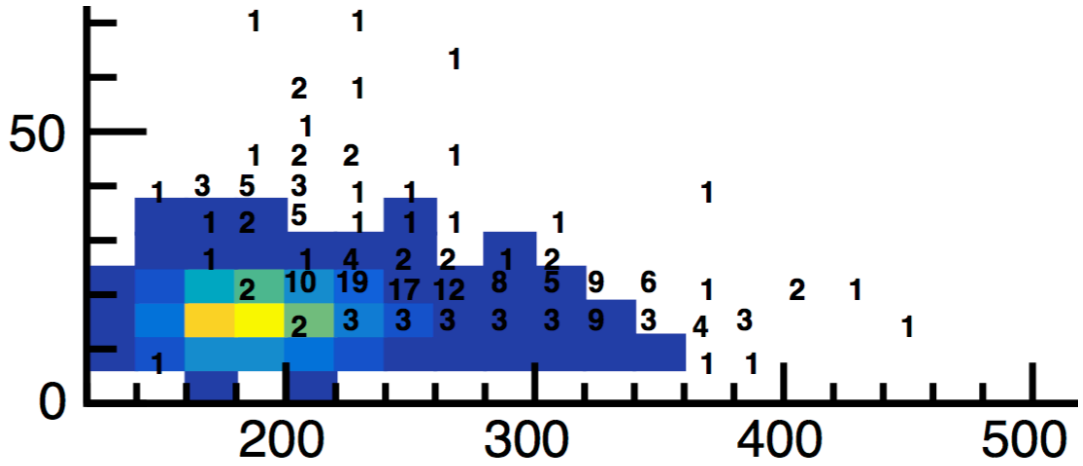
π^- : 64529 events (92%)

e^- : 26230 events (99.2%)







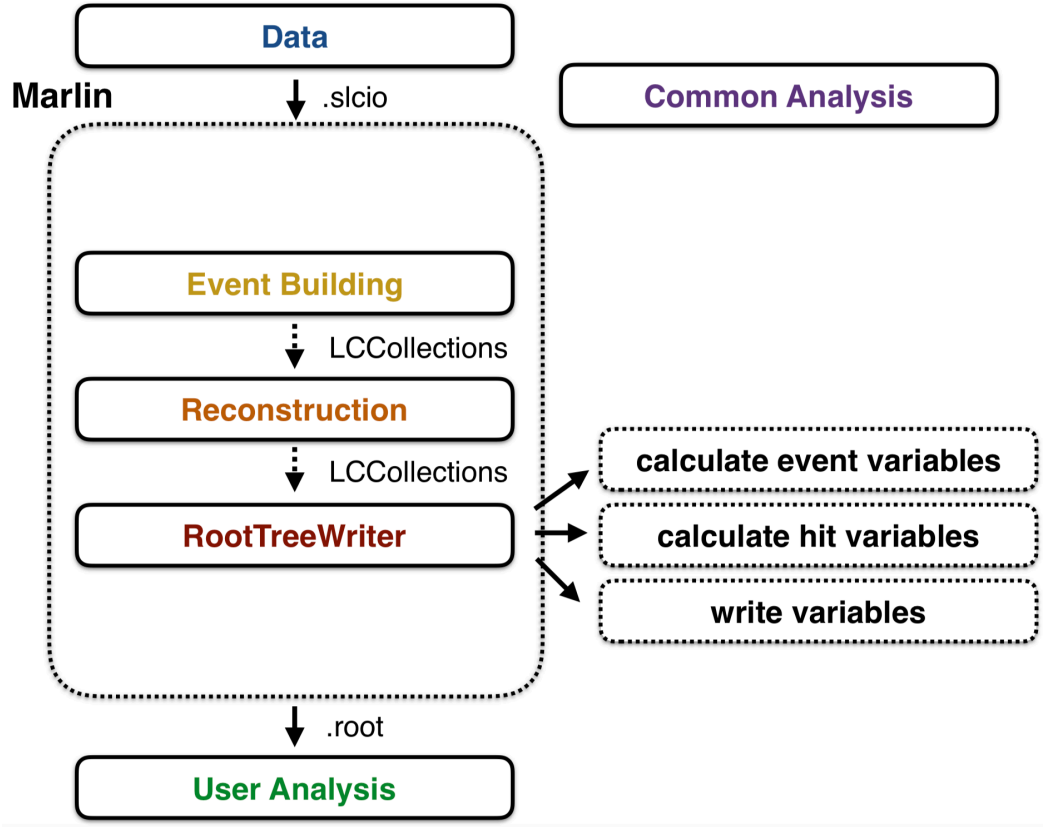


Software architecture studies

To implement particle ID as a standard tool

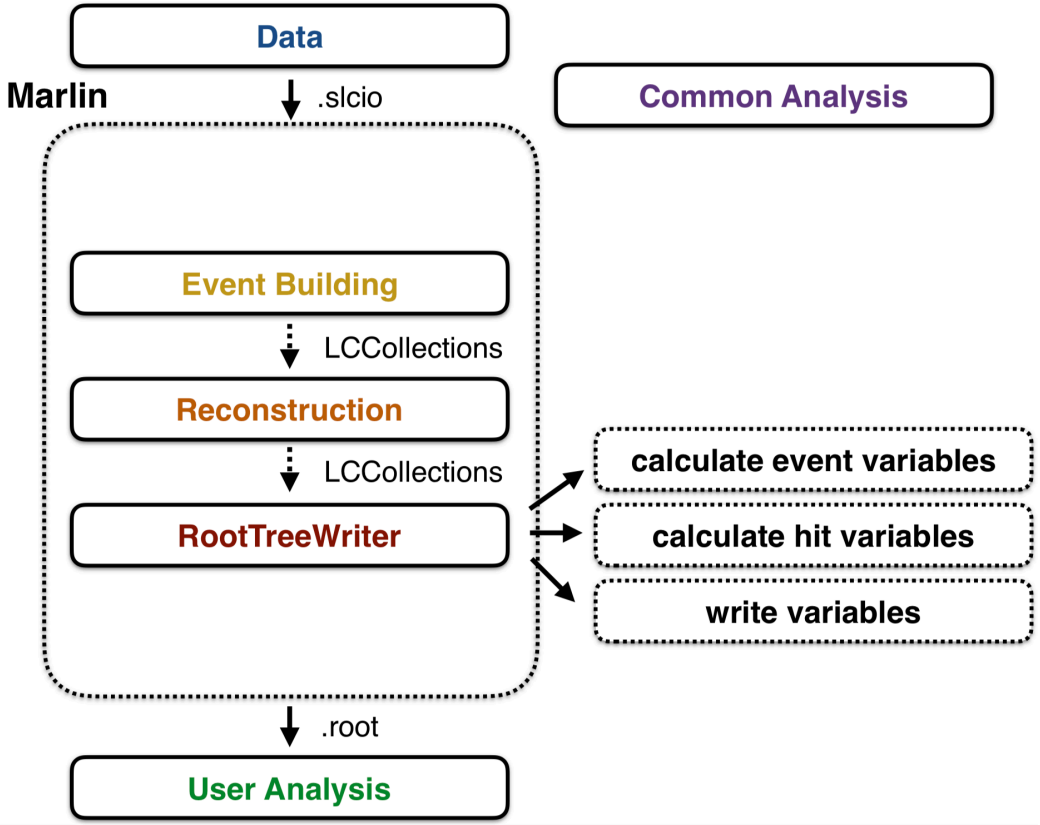
Software architecture

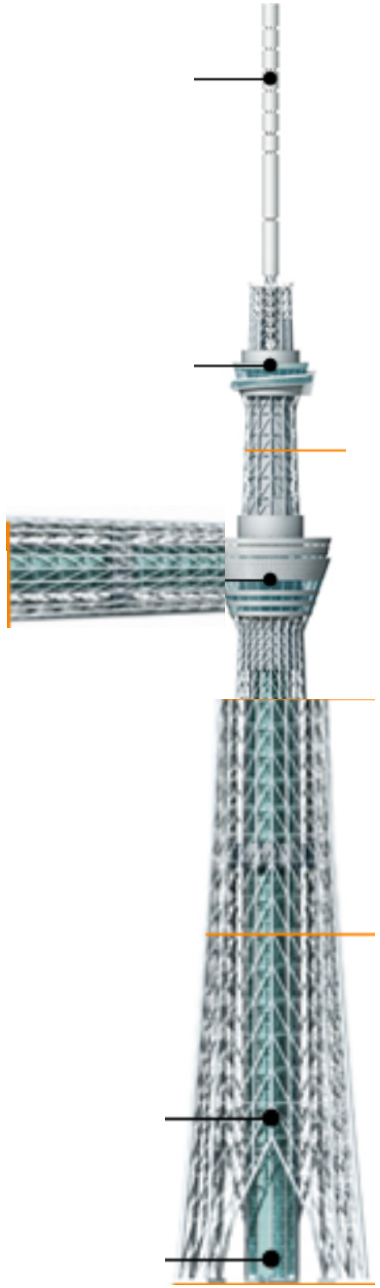
Current state



Software architecture

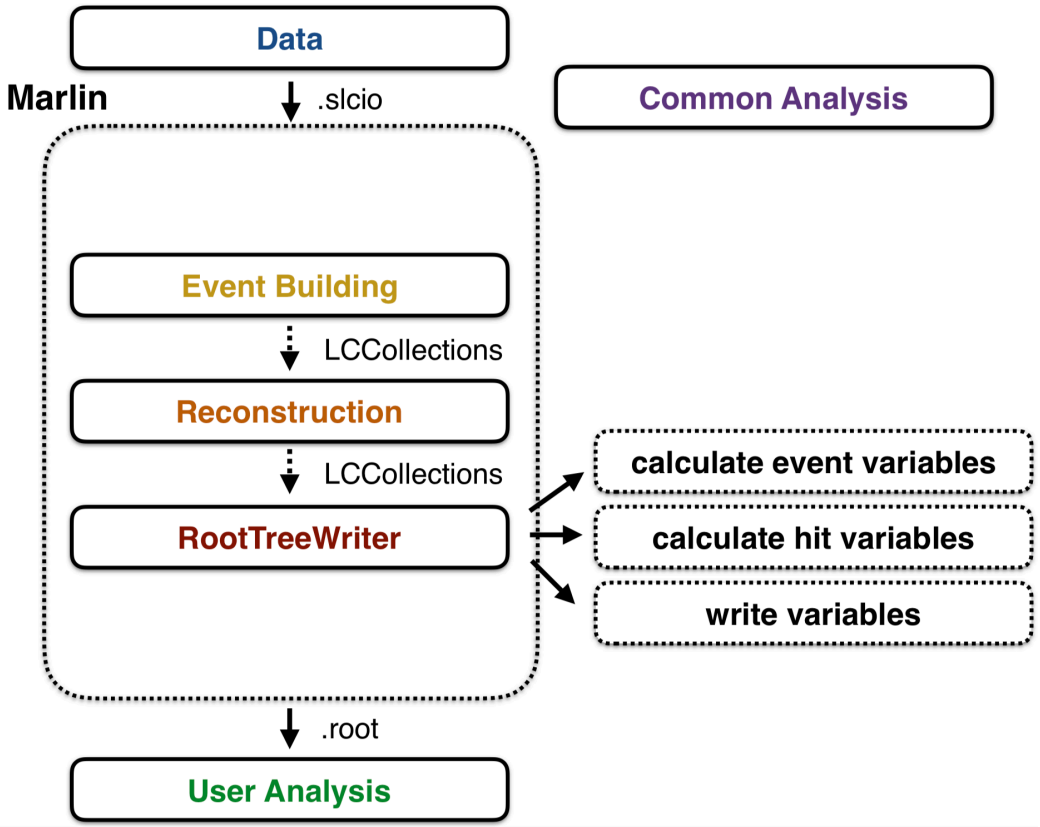
Current state

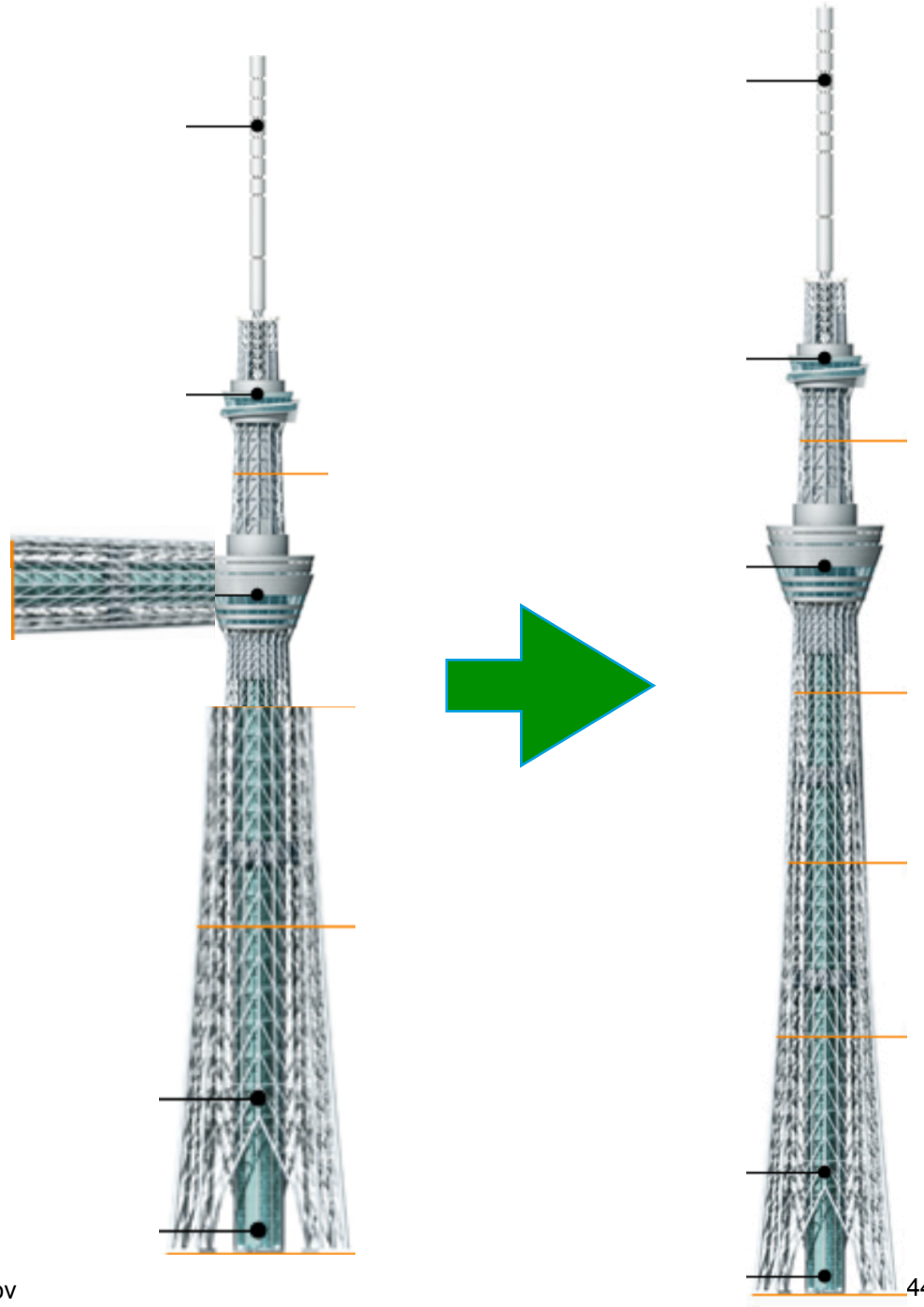




Software architecture

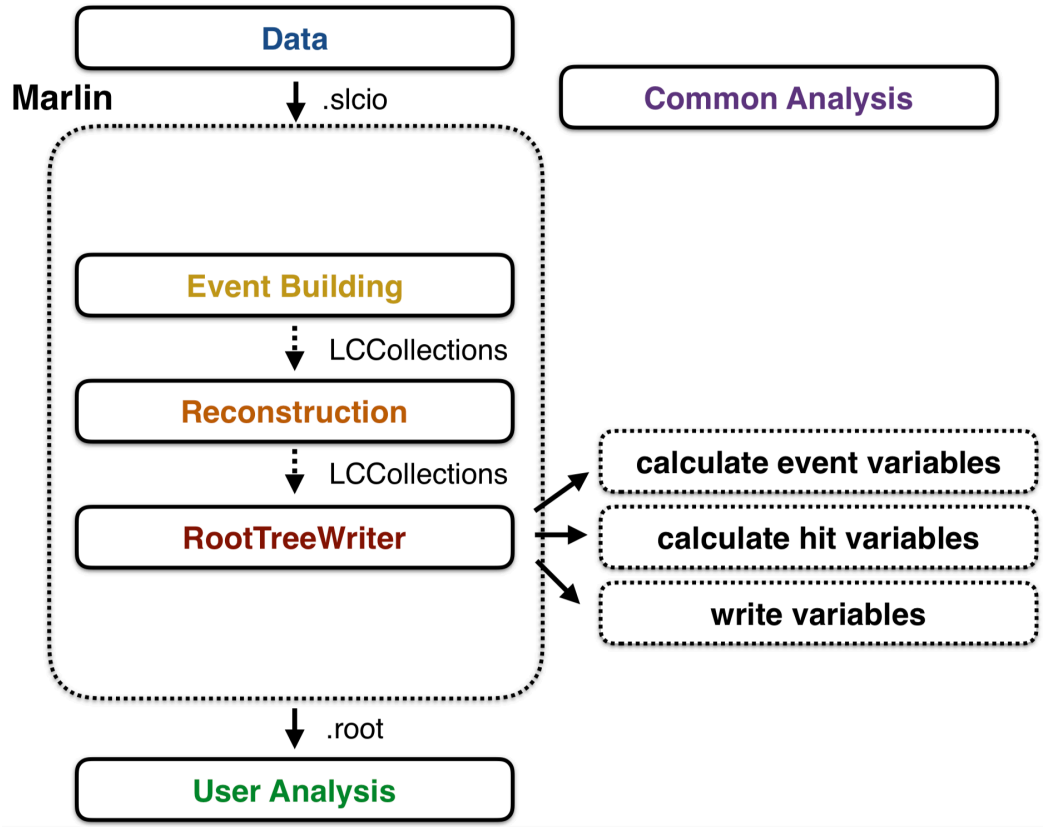
Current state

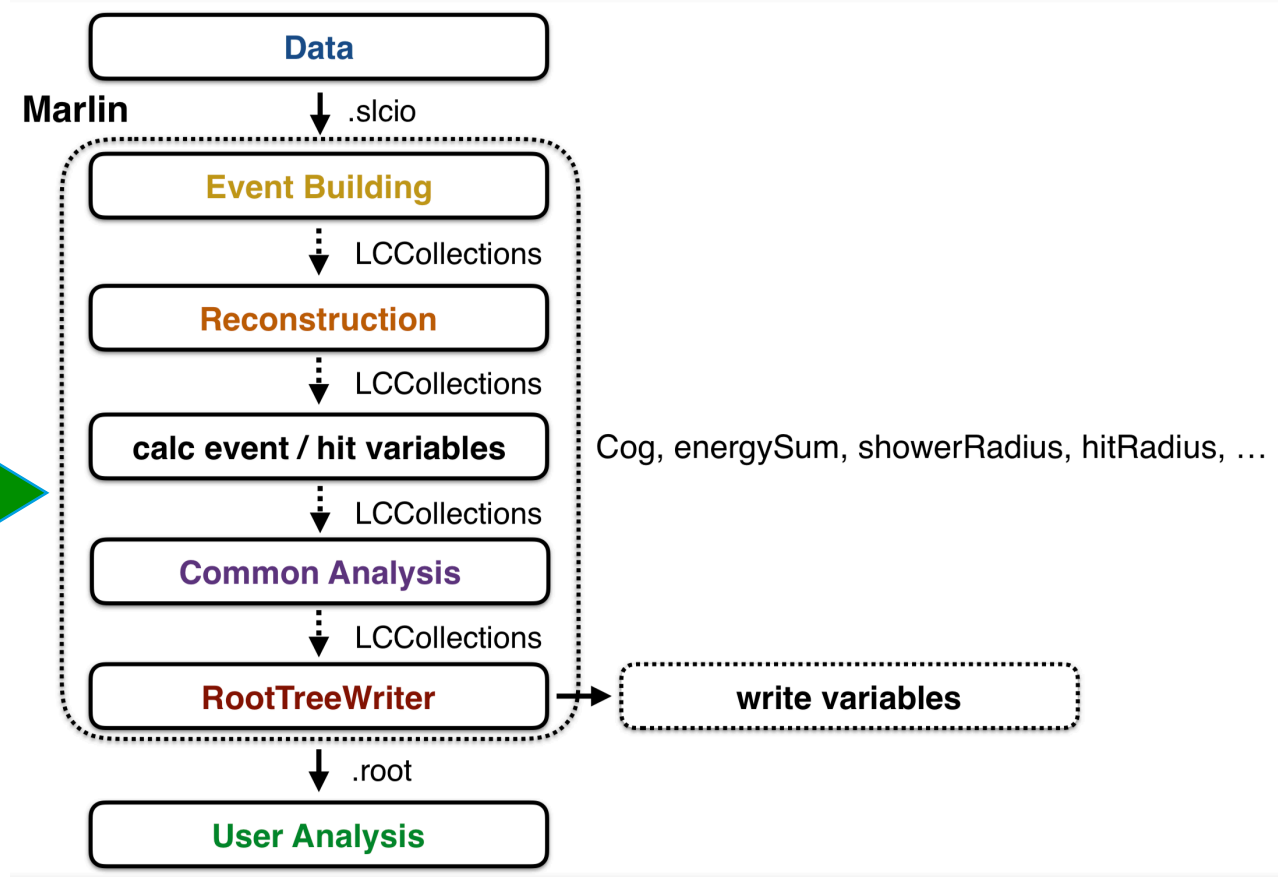




Software architecture

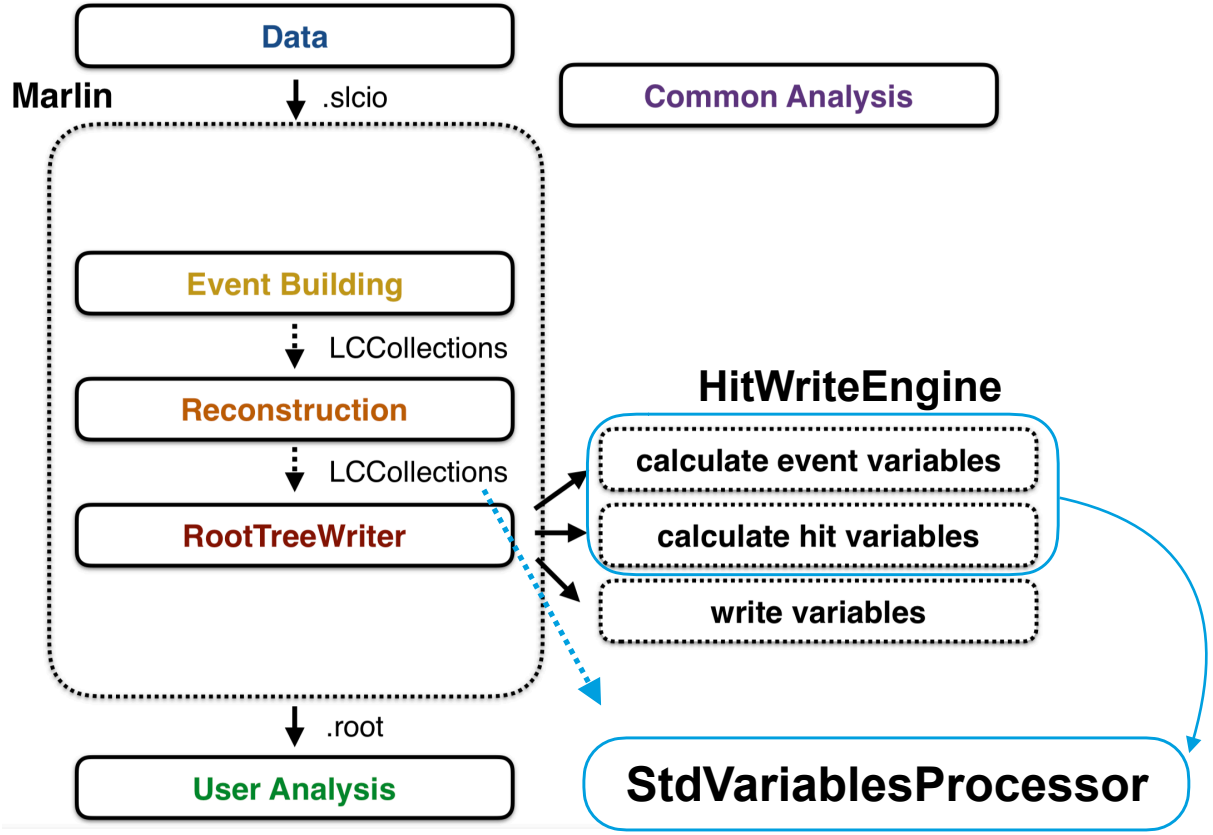
Current state & goal





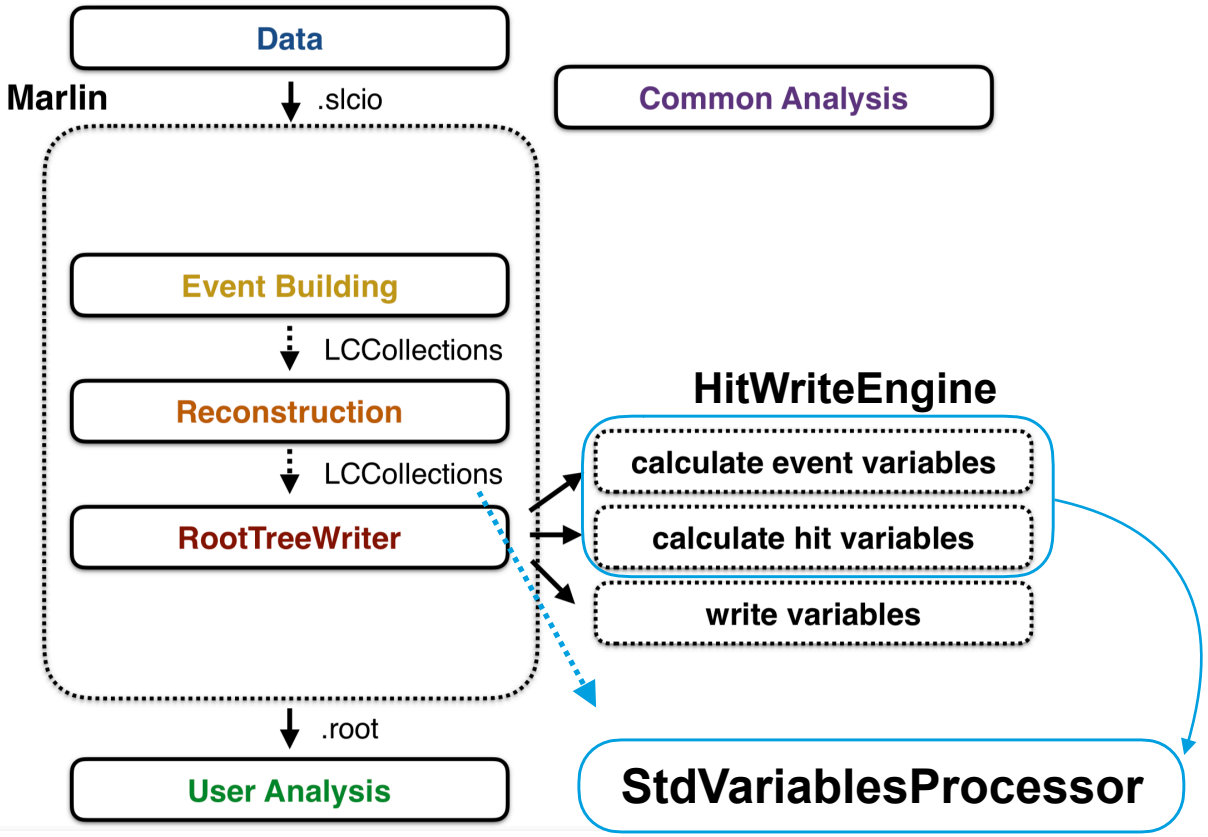
Software architecture

Variables calculation



Software architecture

Variables calculation



(…)

Reconstruction

HCALCalorimeter_Hits (TC_Hits)

StdVariablesProcessor

Hit variables collection

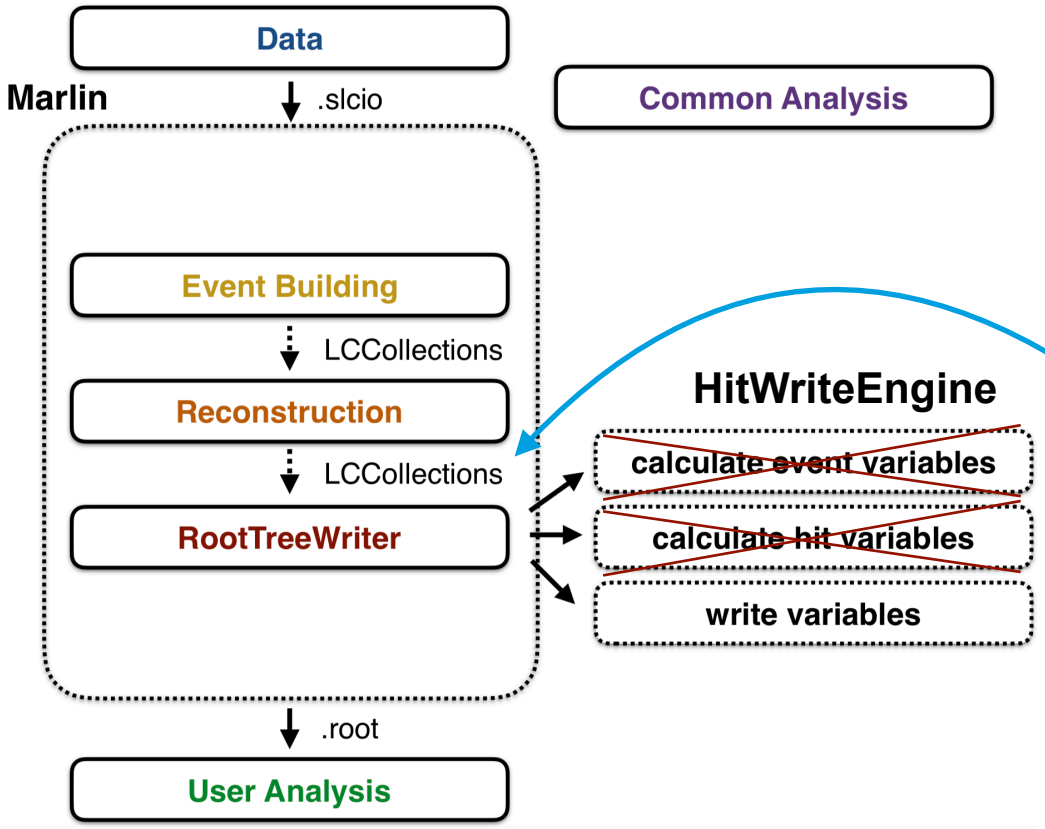
Layer variables collection
1 entry <=>1 Layer

Event general variables collection
1 entry <=>1 Event

(…)

Software architecture

Variables calculation



(…)

Reconstruction

HCALCalorimeter_Hits (TC_Hits)

StdVariablesProcessor

Hit variables collection

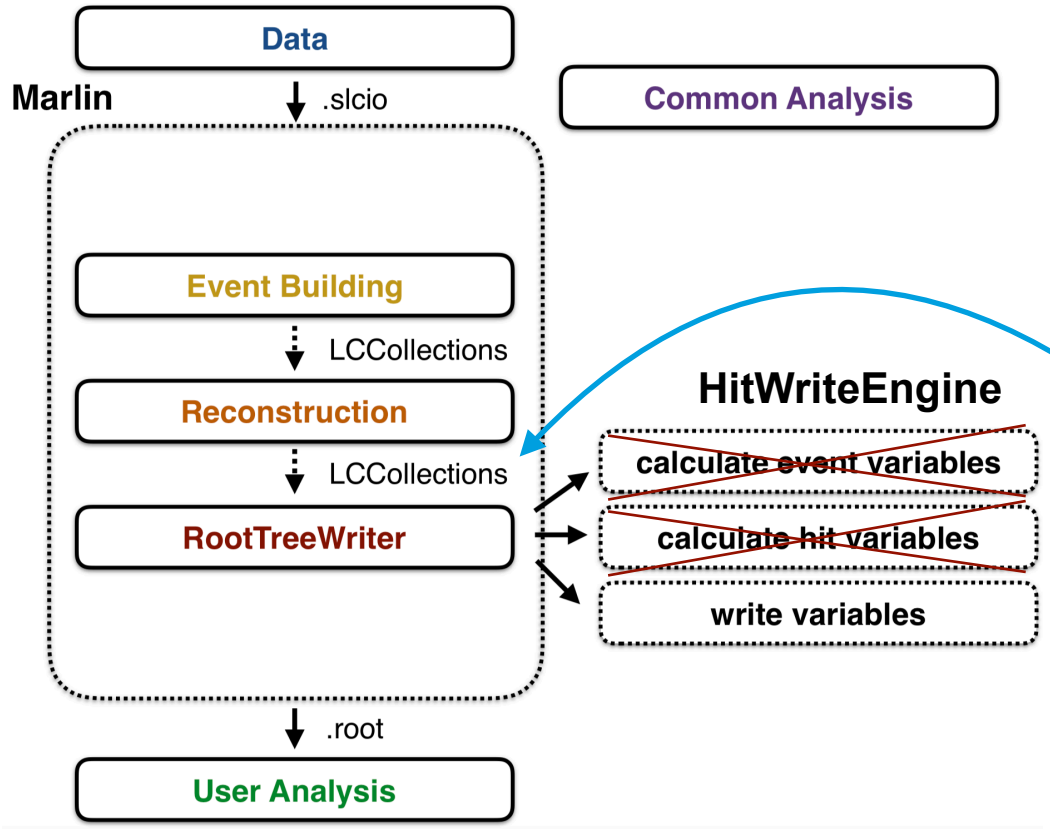
Layer variables collection
1 entry <=>1 Layer

Event general variables collection
1 entry <=>1 Event

(…)

Software architecture

Variables calculation



(…) **DONE**

Reconstruction

HCALCalorimeter_Hits (TC_Hits)

StdVariablesProcessor

Hit variables collection

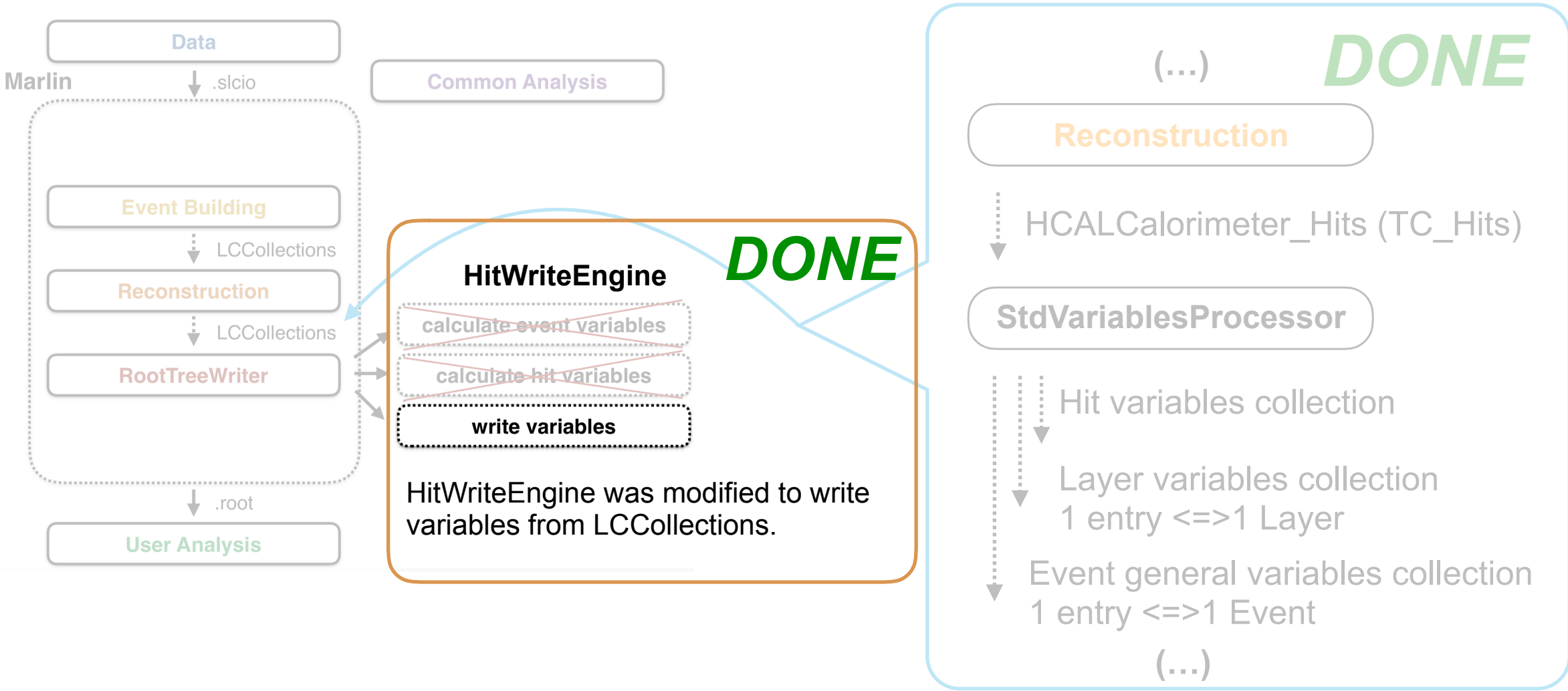
Layer variables collection
1 entry <=>1 Layer

Event general variables collection
1 entry <=>1 Event

(…)

Software architecture

Variables calculation



Software architecture

Variables calculation