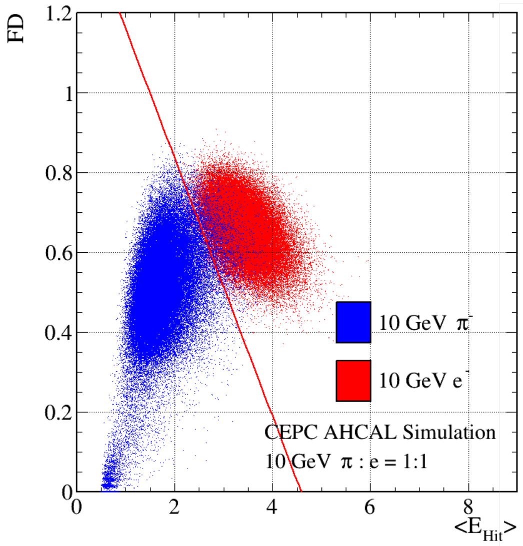
PID cut optimization and automation

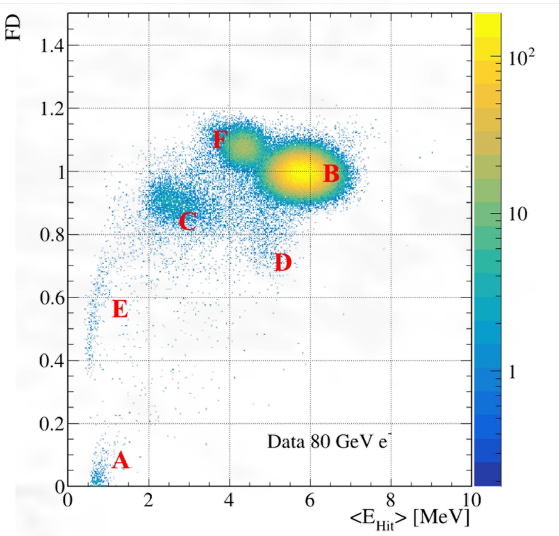
Xin Xia; Yuzhi Che
xiax@ihep.ac.cn

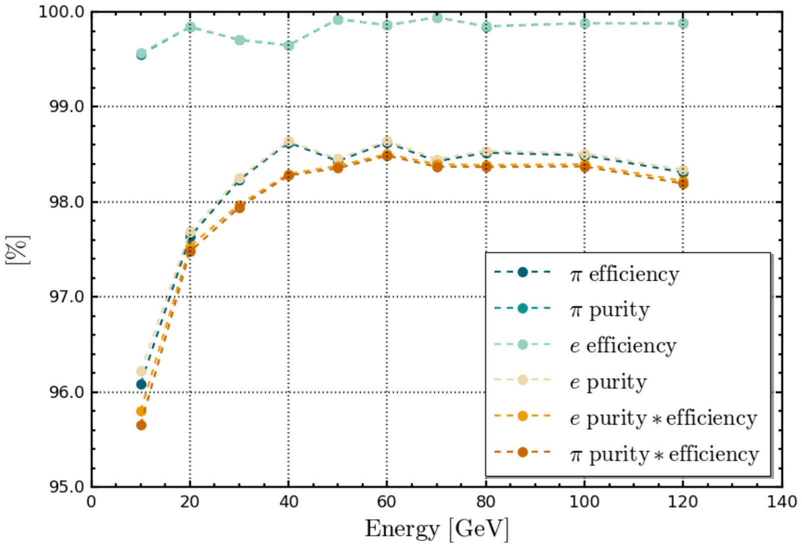
- Try to find an automatic PID method
- Variables optimization
- MC PID performance quantification

Raw PID method

- Method
  - Use two variables:
    - Fractal dimension
    - Average hit energy
  - Artificial cut
- Advantage：
  - Easy
  - clearly distinguish different components in the beam
- Disadvantage
  - Time-consuming
  - with a certain degree of arbitrariness
  - did not achieve the optimal level of PID performance for this method.



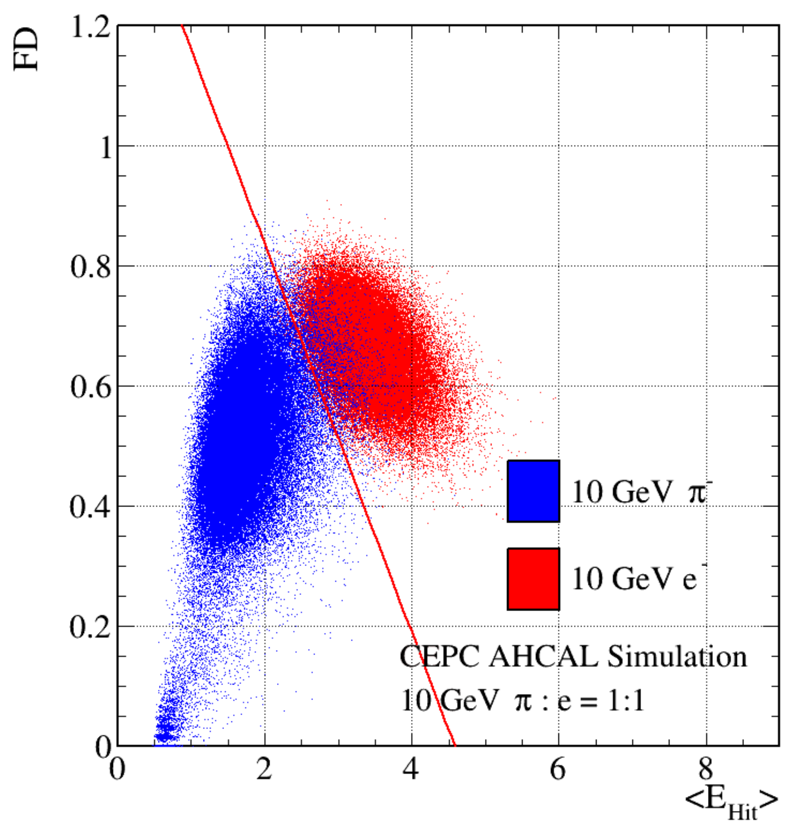




Misjudgment rate minimization

- Step 1: prepare MC sample
  - 1:1 (100k pion + 100k electron)
  - ratio of e and pi in real data (obtained by Miunit fit)
- Step 2: define misjudgment rate
  - Linear cut: FD = k * <E_{Hit}> + b
  - Mis-id-num = (pi_all – pi_select) + (e_all – e_select)

    = Func (k, b)
- Step 3: minimize misjudgment rate
  - Minuit (fail)
  - Brute force
  - Other method



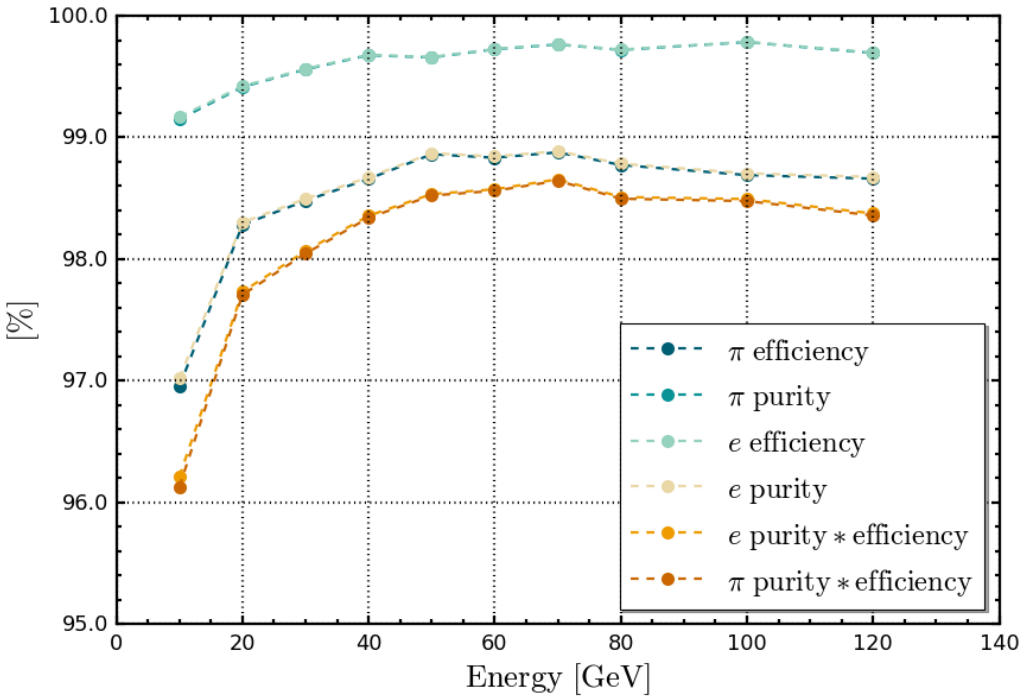
Brute force

- Parameter
  - For k: range (-0.6, -0.005), number = 100
  - For b: range (1, 3), number = 100
  - 100 ×100 array
- (k, b) → Mis-id-num
- Result = Min(Mis-id-num) → k, b
- Time: ~ 10s

MC sample 1 : 1 mixture + Brute force

-- Results

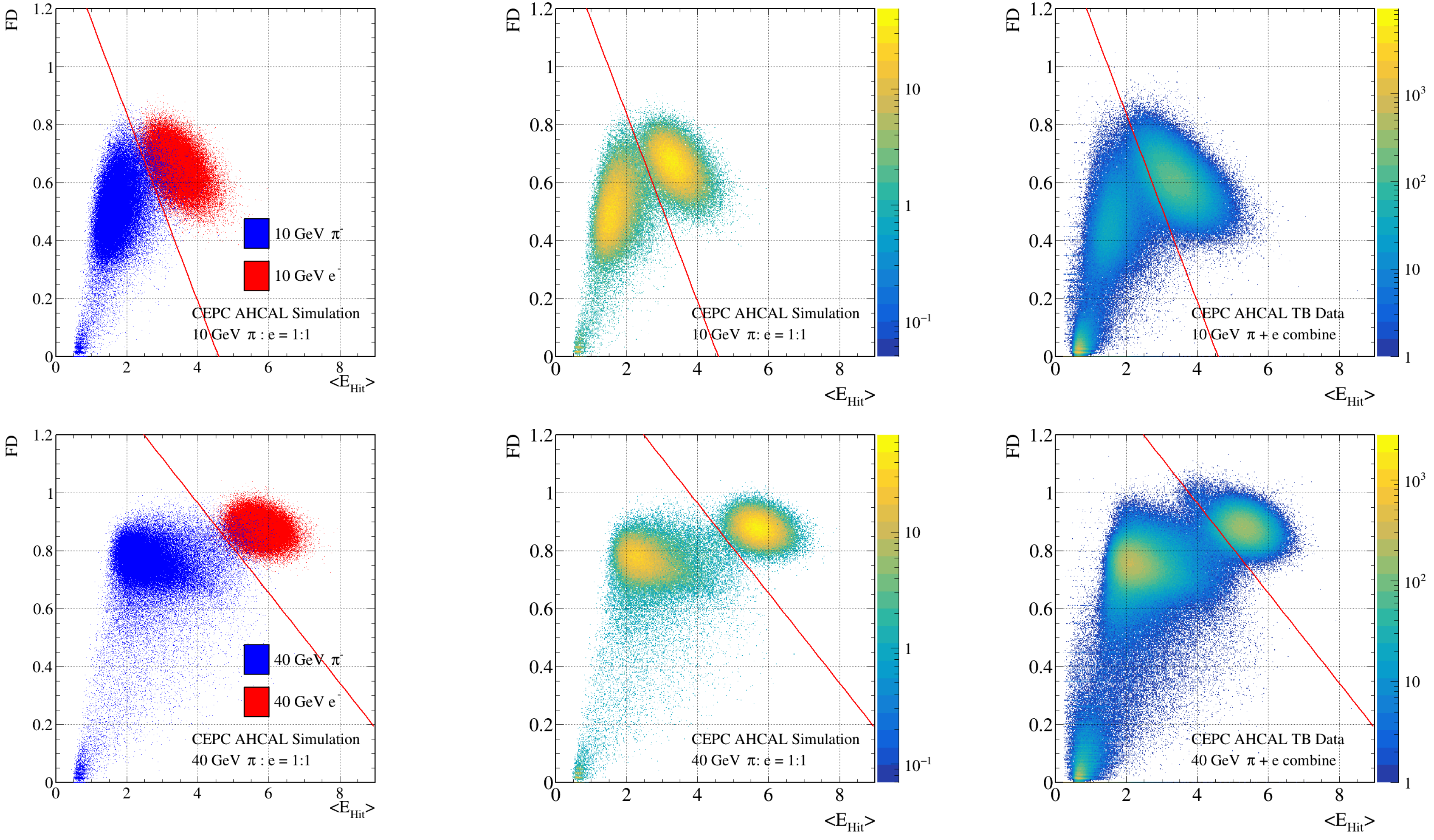
- For pion:
  - Efficiency > 97%; Purity > 99%.
  - Efficiency can be improved with larger energy and enters a plateau region when the energy reaches more than 40 GeV.
  - Purity is relatively stable.
- For electron:
  - Efficiency > 99%; Purity > 97%.
- Pion efficiency behaves similarly to electron purity, and electron efficiency behaves similarly to pion purity.
- PID performance improves with energy.



MC sample 1 : 1 mixture + Brute force

-- FD-<E_{Hit}> plots

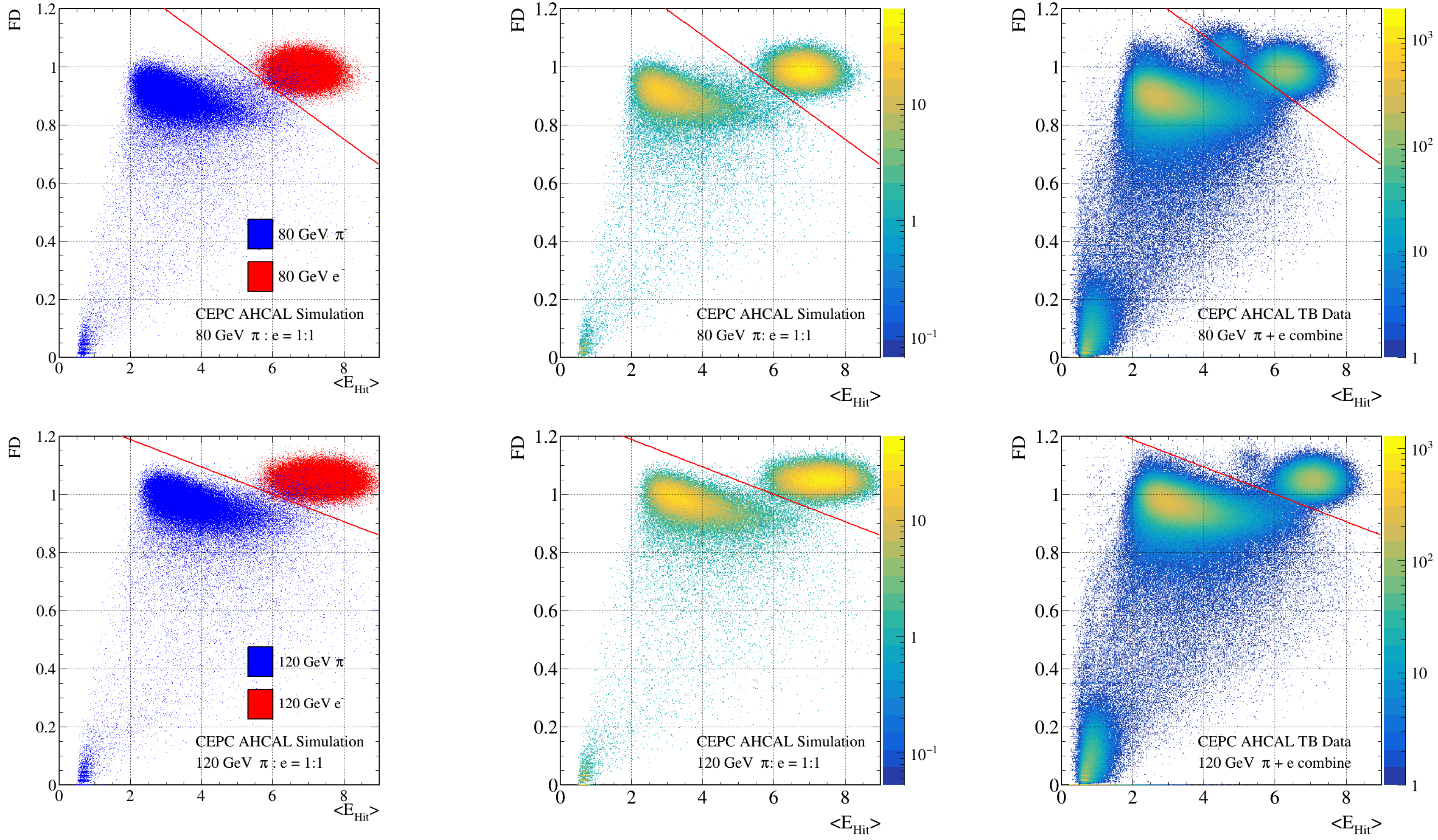
MC/Data discrepancy, so the cut for MC is not suit for Data. Especially at low energy points.



MC sample 1 : 1 mixture + Brute force

-- FD-<E_{Hit}> plots

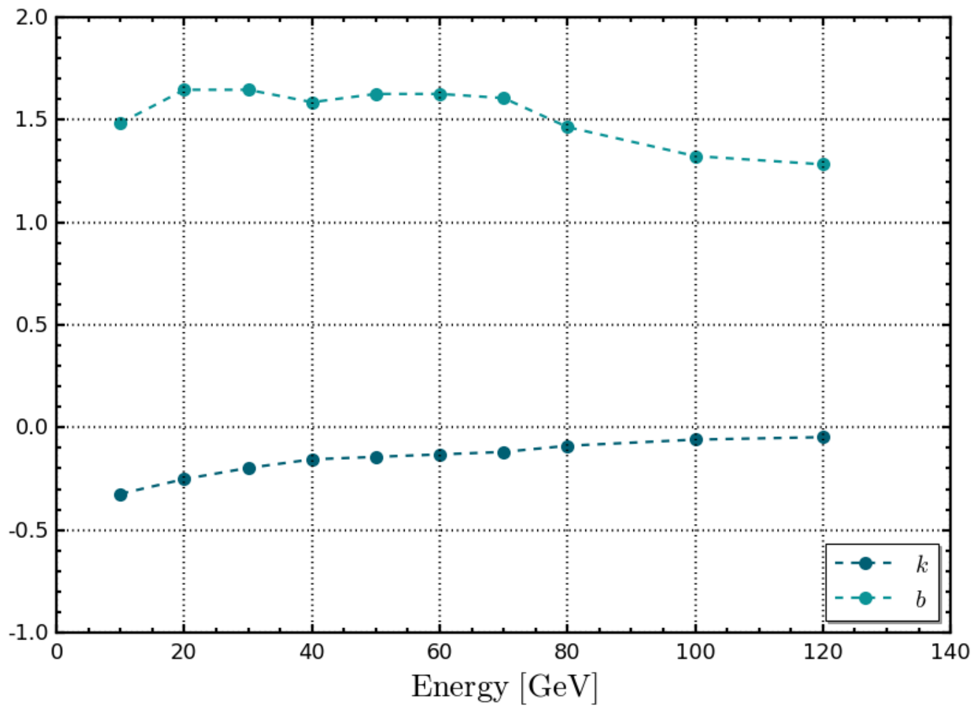
MC/Data discrepancy, so the cut for MC is not suit for Data. Especially at low energy points.

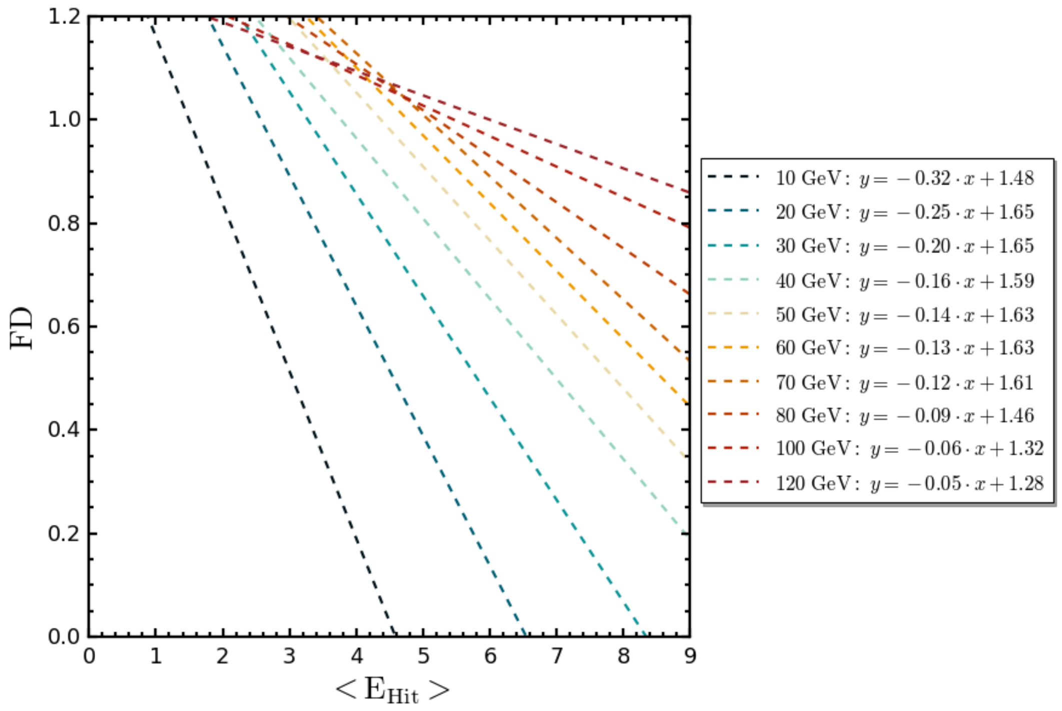


MC sample 1 : 1 mixture + Brute force

-- cut

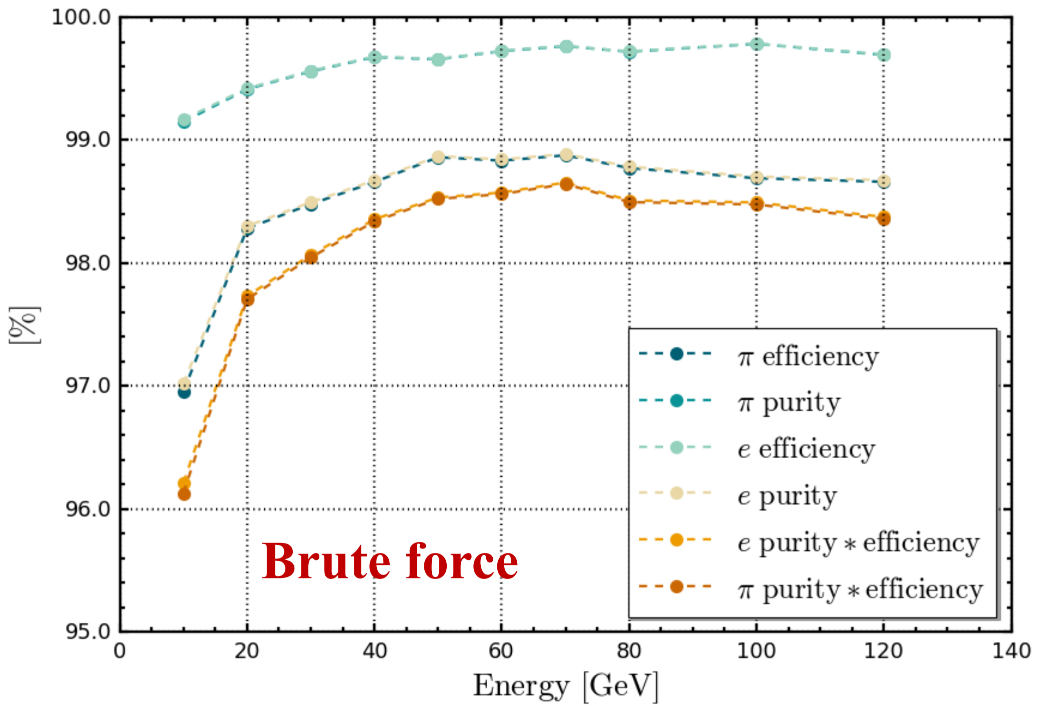
The cut that varies with energy.

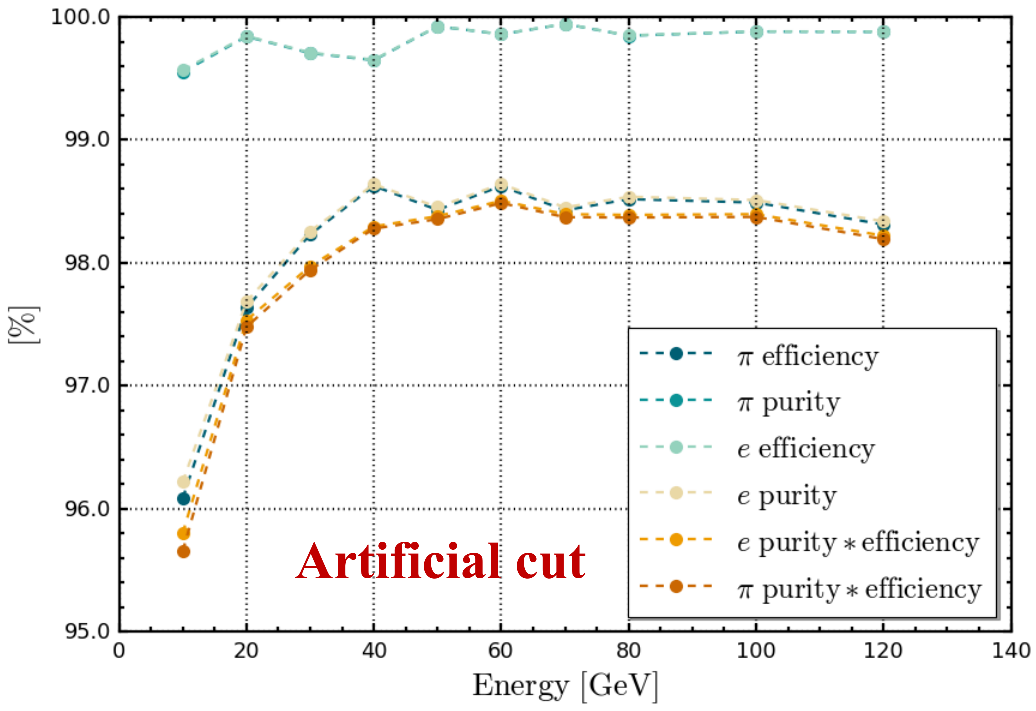




Comparison between Brute force and Artificial cut

The performance of Brute force method is ~ 1% better than Artificial cut.

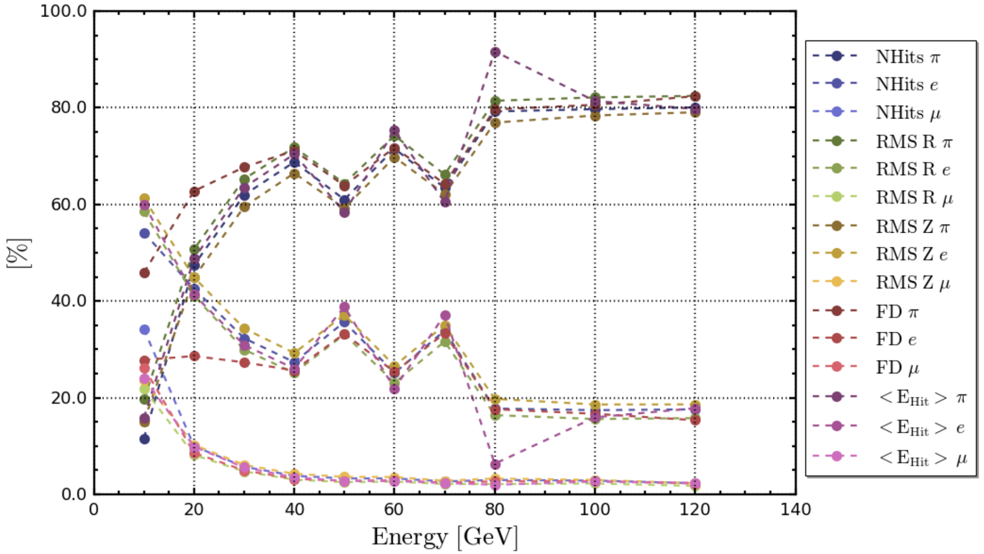


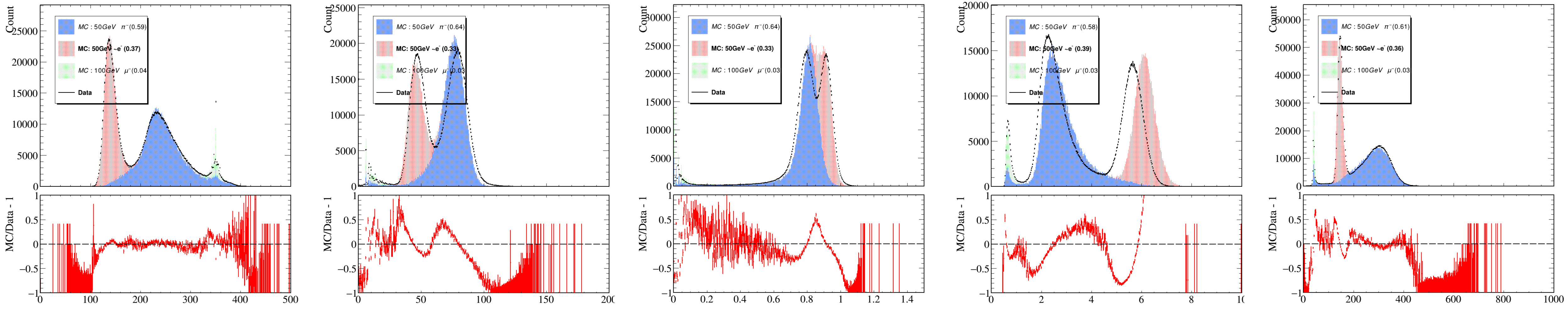


MC sample real ratio mixture + Brute force

-- Ratio estimate

- Data: combine all the runs at the same energy
- Method: single variable + Minuit
  - Here we compared five different variables: NHits, RMS R, RMS Z, <E_{Hit}>, FD
- Difference in ~ 5%.
- For RMS Z, Best consistency between MC and data. So we choose the fit result obtained by using RMS Z.

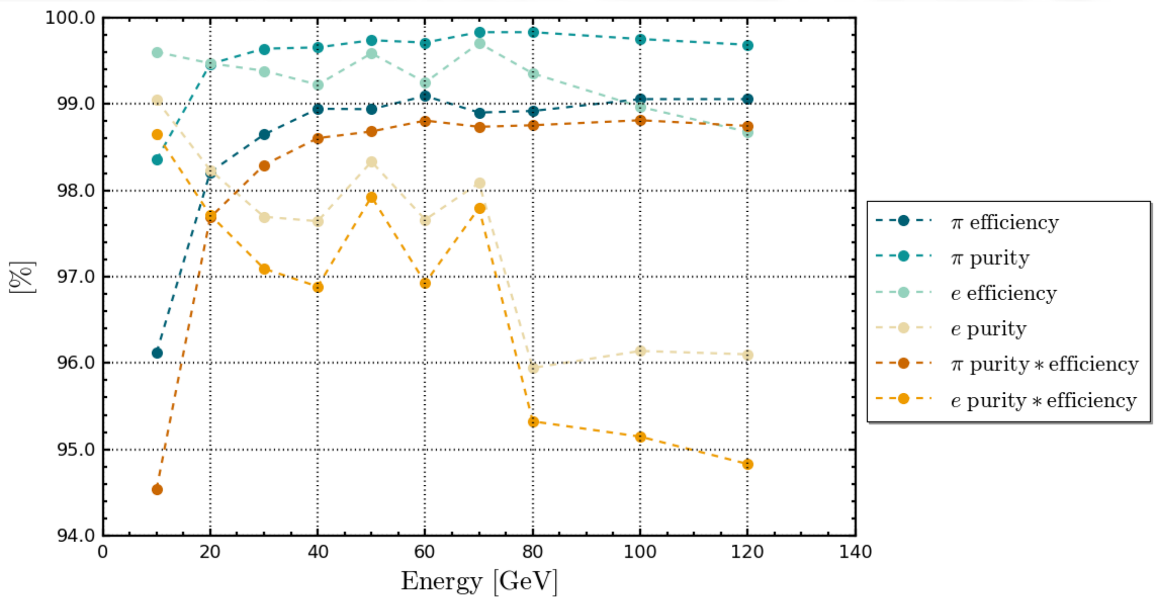




MC sample real ratio mixture + Brute force

-- Results

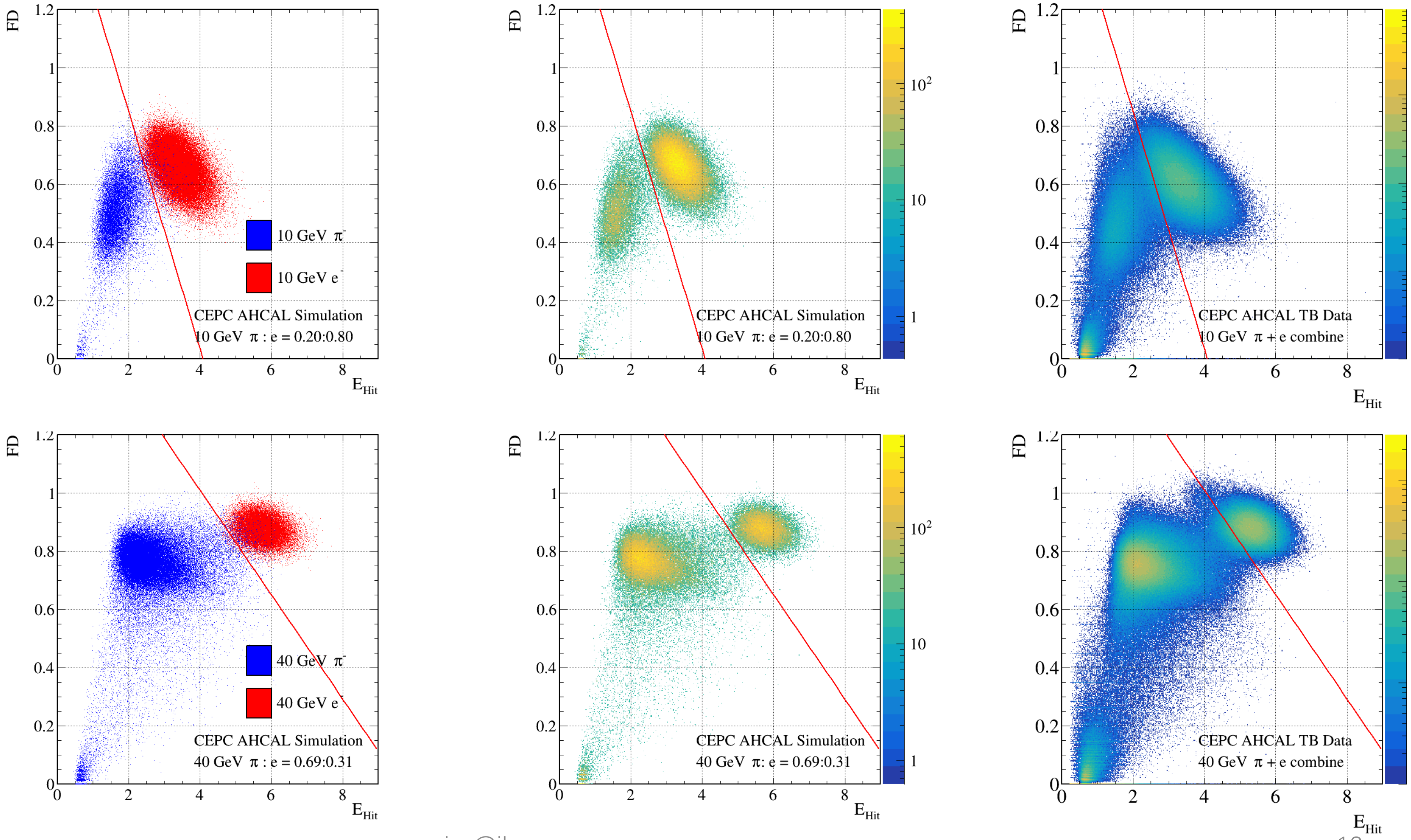
- For pion:
  - Efficiency > 96%; Purity > 98%.
  - Efficiency can be improved with larger energy and enters a plateau region when the energy reaches more than 40 GeV.
  - Purity is relatively stable.
- For electron:
  - Efficiency > 98%; Purity > 96%.
  - As the energy increases, the purity decreases.
  - Efficiency is relatively stable.



MC sample real ratio mixture + Brute force

-- FD-<E_{Hit}> plots

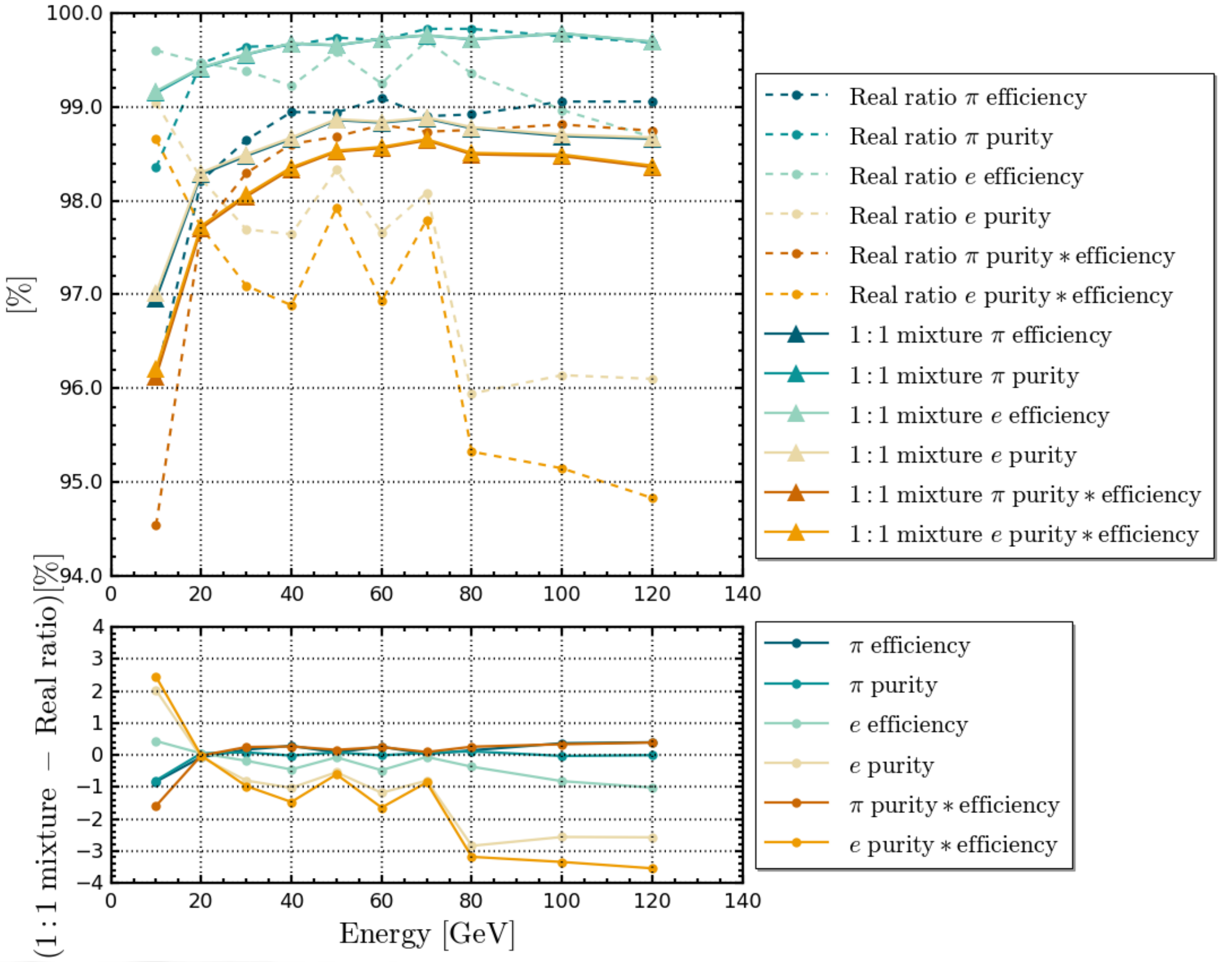
Performance same as 1 : 1 mixture.



Comparison between 1:1 mixture and real ratio mixture

-- Performance

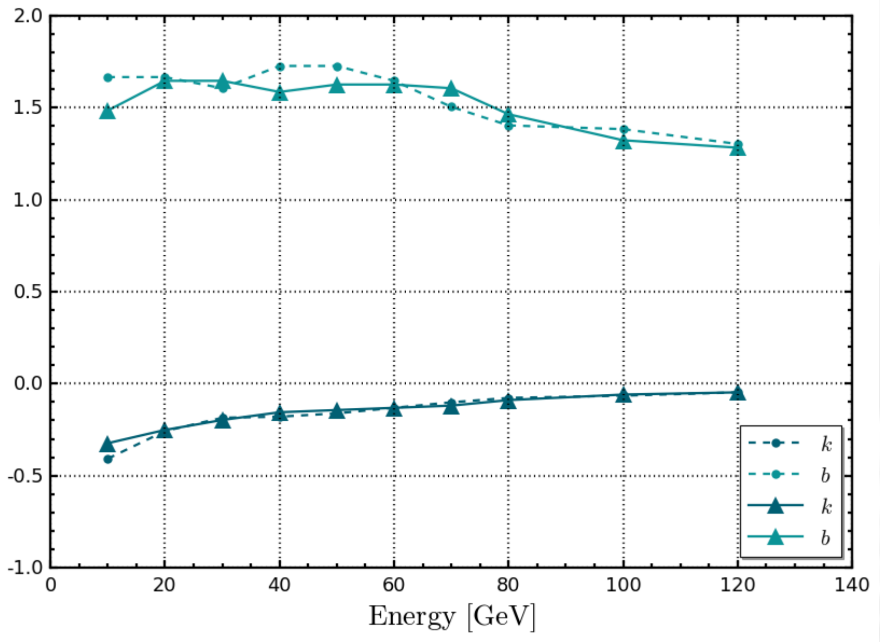
- Difference
  - Efficiency: in 1%
  - Purity: in 3%
  - Efficiency* Purity: in 4%

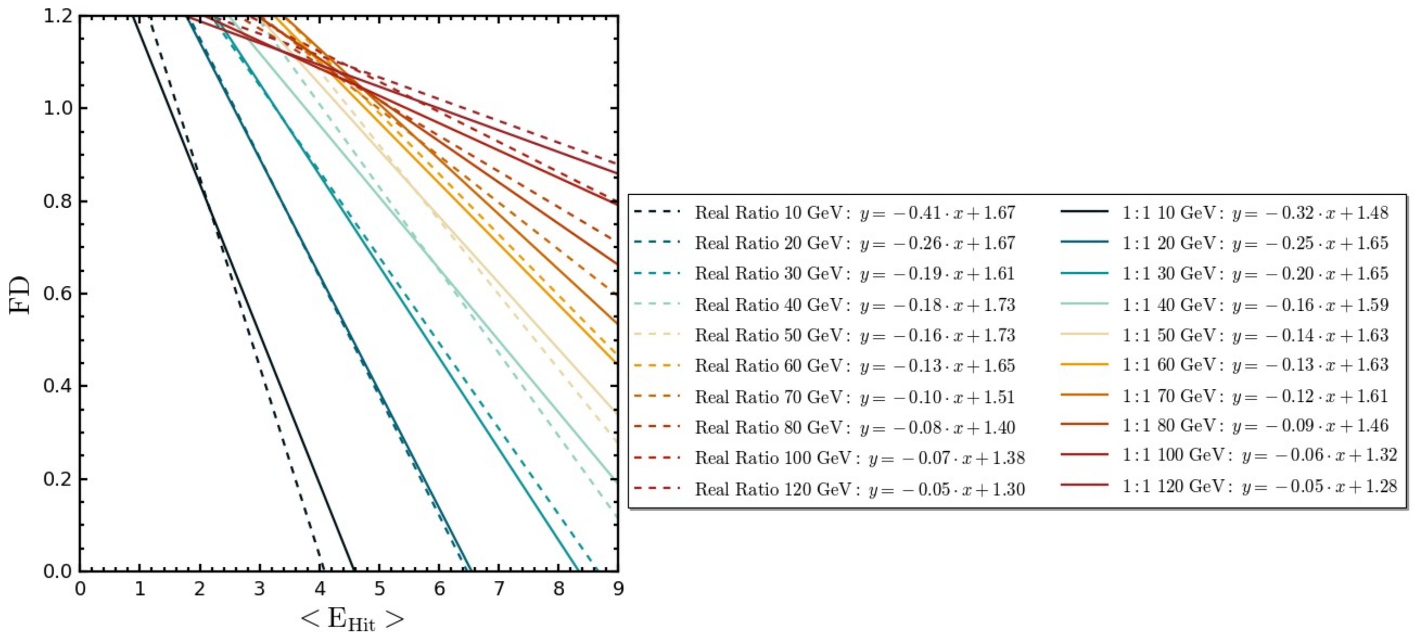


Comparison between 1:1 mixture and real ratio mixture

-- Cut

The cut that varies with energy.



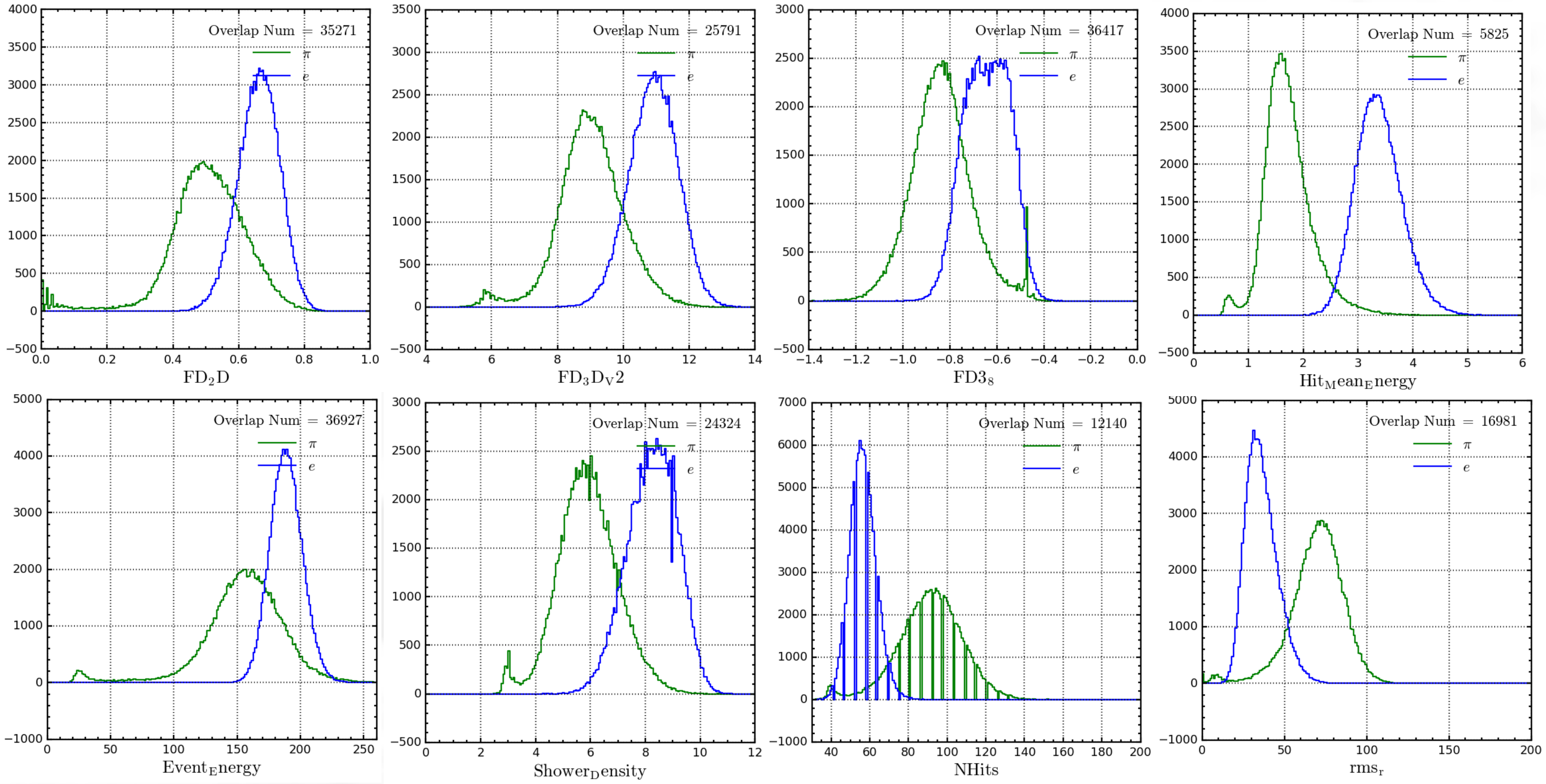


xiax@ihep.ac.cn

Variables optimization and selection

Separation power:

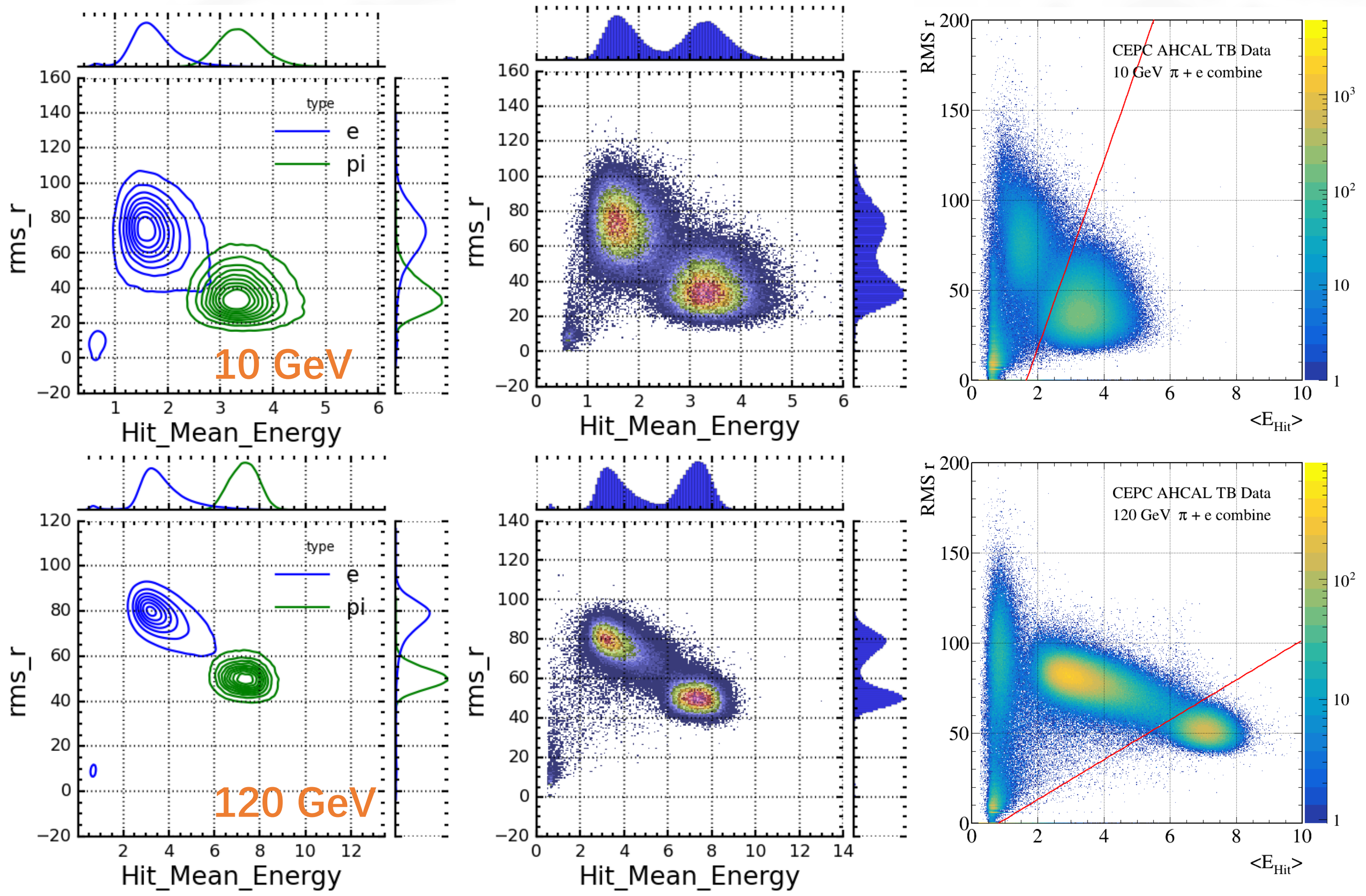
<E_{Hit}> > NHits > RMS r > shower density > FD > Event energy



<E_{Hit} > + RMS r + Brute Force

-- plots

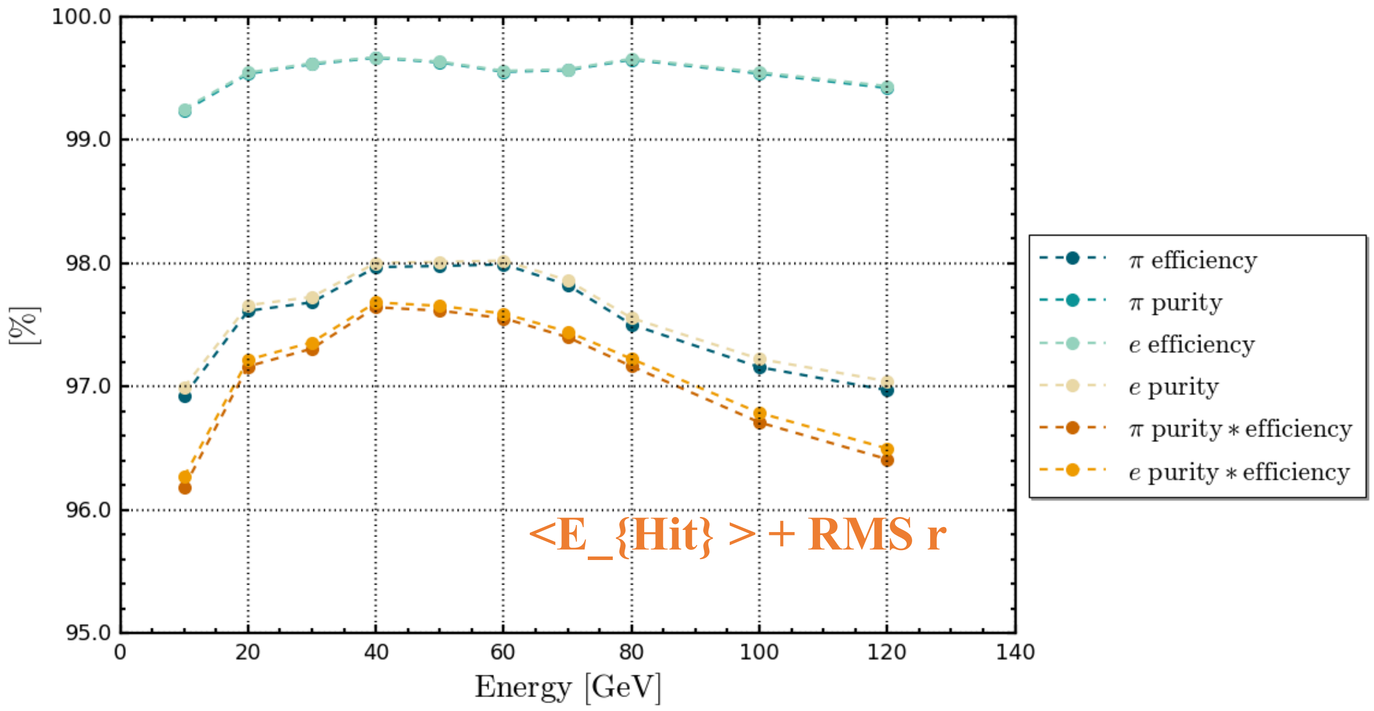
PID performance looks good.

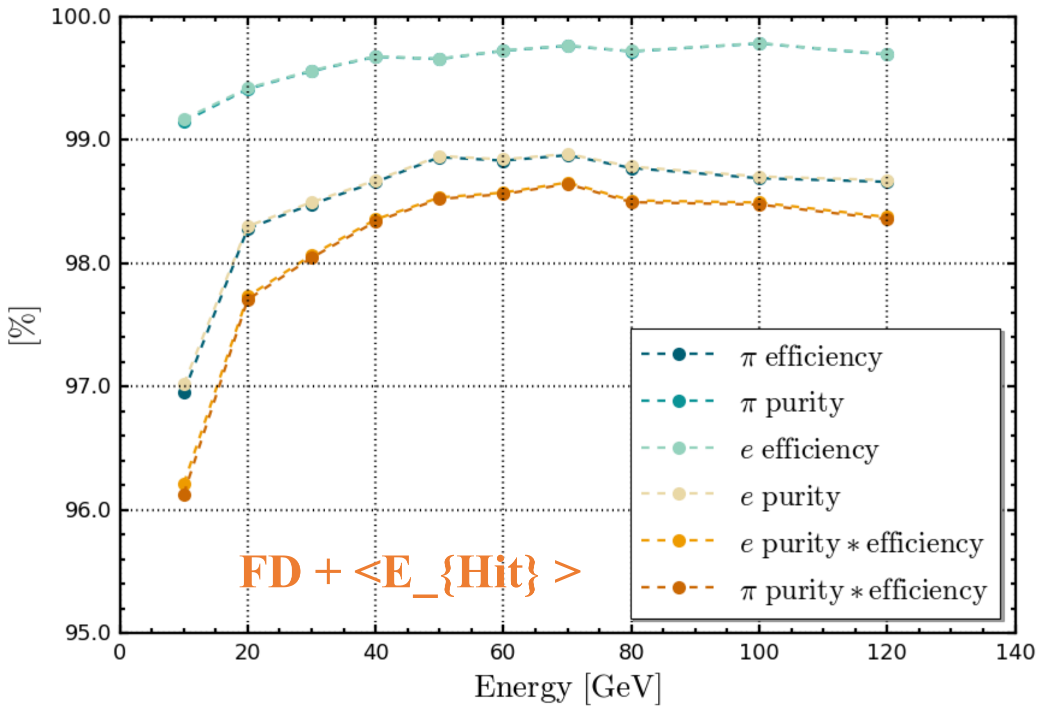


<E_{Hit} > + RMS r + Brute Force

-- performance

PID performance worse than FD + <E_{Hit} > .

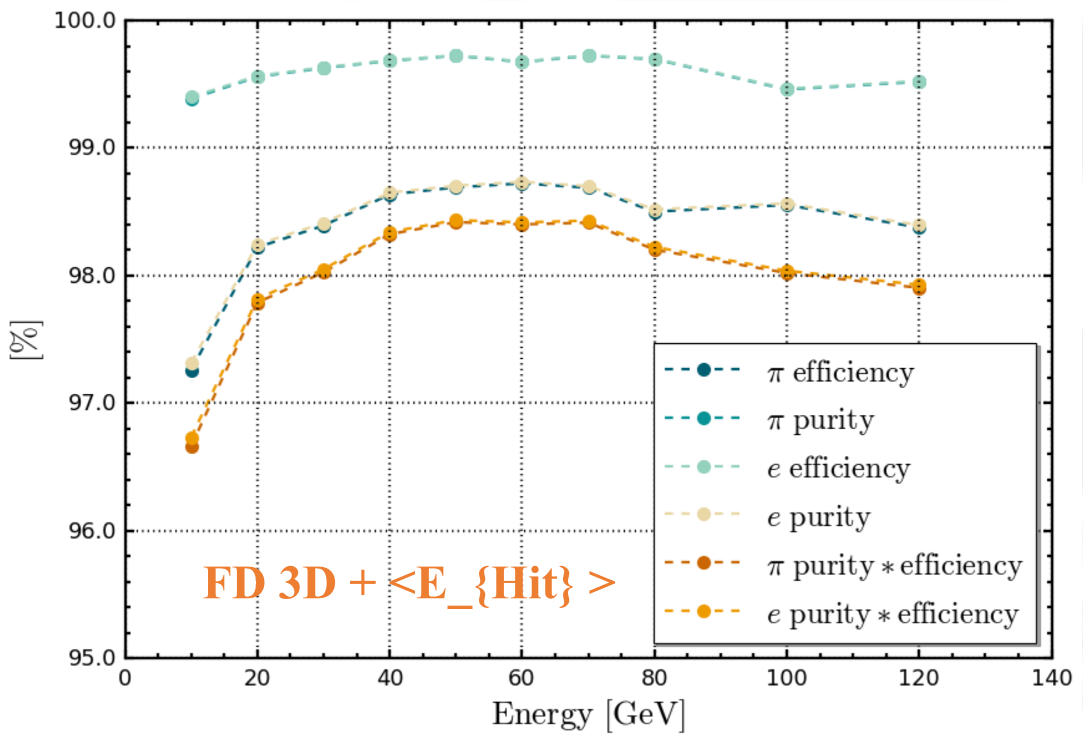


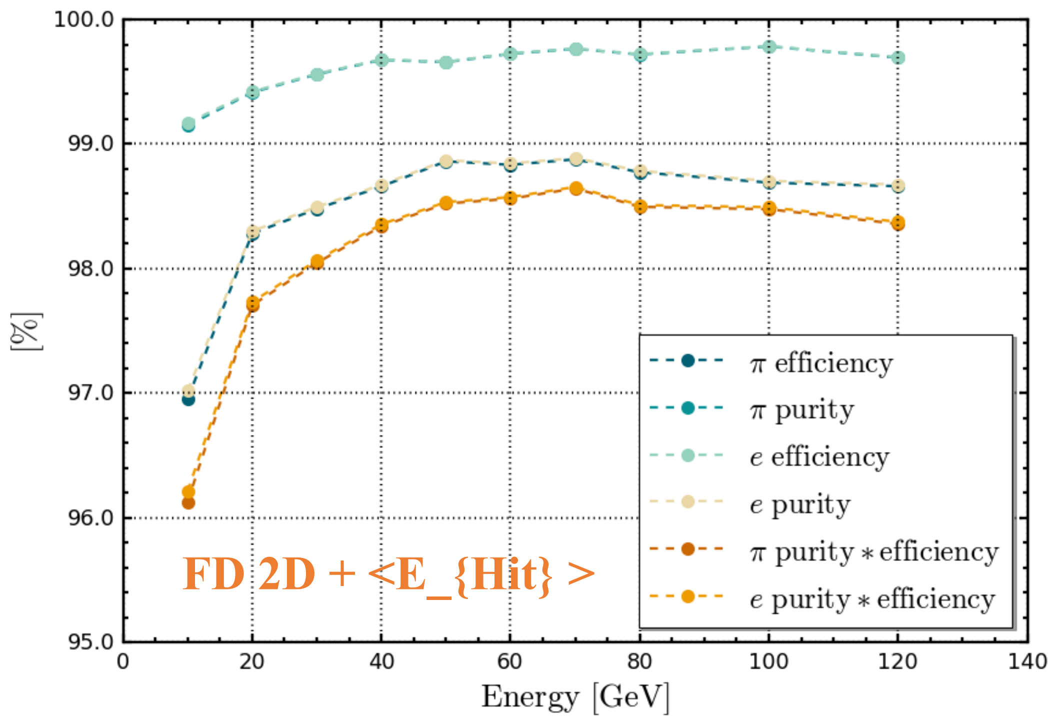


<E_{Hit} > + FD 3D + Brute Force

-- performance

Low energy point FD 3D better than FD 2D, high energy point converse.
The performance of FD 3D is stable than FD 2D.





Summary

- A method for automatic PID cut has been developed.
  - FD 2D + <E_{Hit}> + Brute froce
  - Efficiency and purity better than 97% (1:1 mixture) / 96% (real ratio mixture)
- Compare the separation power of different variables
- Try another PID variable
  - – RMS R, but the performance is worse than FD 2D + <E_{Hit}> .
  - FD 3D, performance is similar with FD 2D

Thanks for your listening!

Xin Xia
xiax@ihep.ac.cn