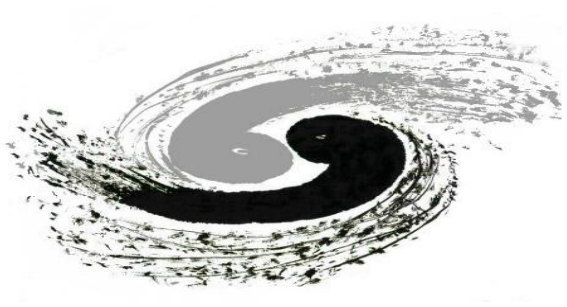


CEPC

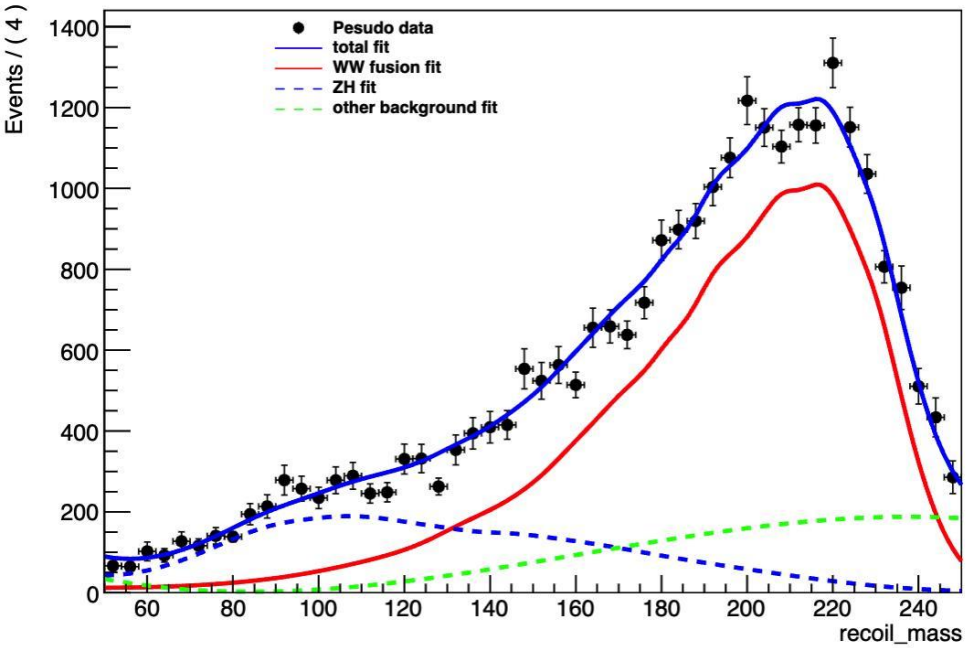
Jets, samples and Wednesday working meeting

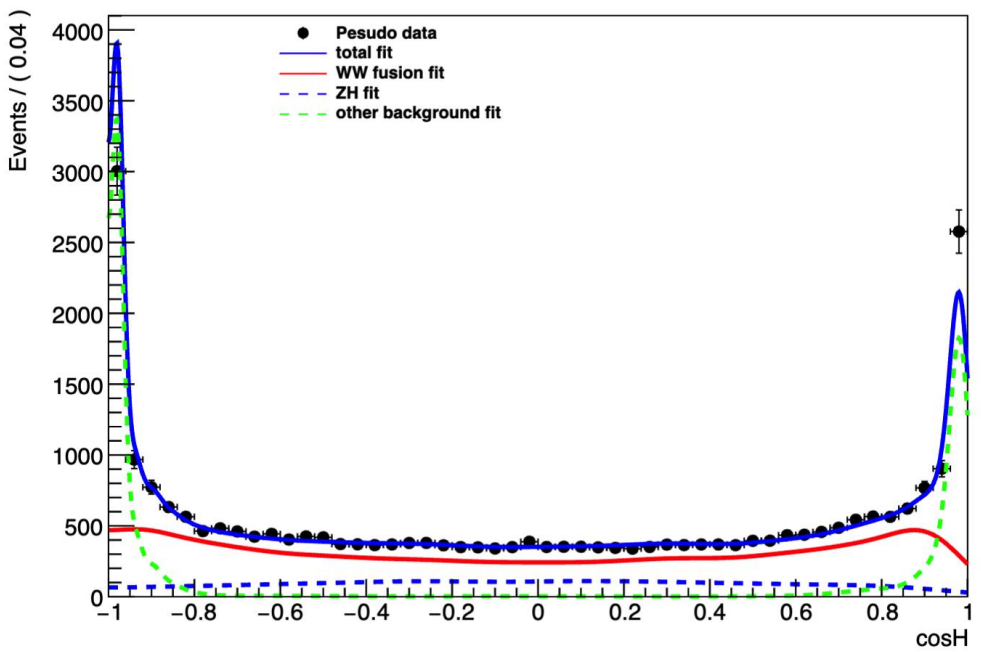
Kaili Zhang

zhangkl@ihep.ac.cn

WWfusion

@Yingqi





Disscusion with committee on Thursday;
Gen filter+ Sample Production issue;

Rb Measurement

@Hancen

Discussion analysis strategy on Wednesday;

Issues

Merging issues

If keep all the 11 categroies, the calculation is well defined for no merging issue. However, if some categroies such as u, d, s are combined ($q = u, d, s$), the merging issues will appear.

One thing should be argued is that, if the two confusion matrix (single tagging and double tagging) are equal.

Preliminary solution

The two confusion martixes used in tagging quation are not equal,

Normalization issues

The model now can not guarantee that for the same flavor, the number of reco flavor jet equals to the number of true flavor jet. From the opinion of Kaili, some bias of mean value will appear.

Preliminary solution

The difference between the number of two kinds of same flavor jets, will reflect in the total score of pred label is not equal to 1, which have no influence on the calculation.

System error estimation

Preliminary solution

This kind of error will reflect in the confusion matrix, which means, there is no so-called system error from detector.

Correlation issues

In the previous paper, the correlation was omitted due to its tiny value. However, our stat. error now is about 1e-6, the correlation, no matter between different flavor or quark-antiquark, could be the main contribution of error.

Preliminary solution

Aim to update the minimal chi2 method to avoid the uncertainty.

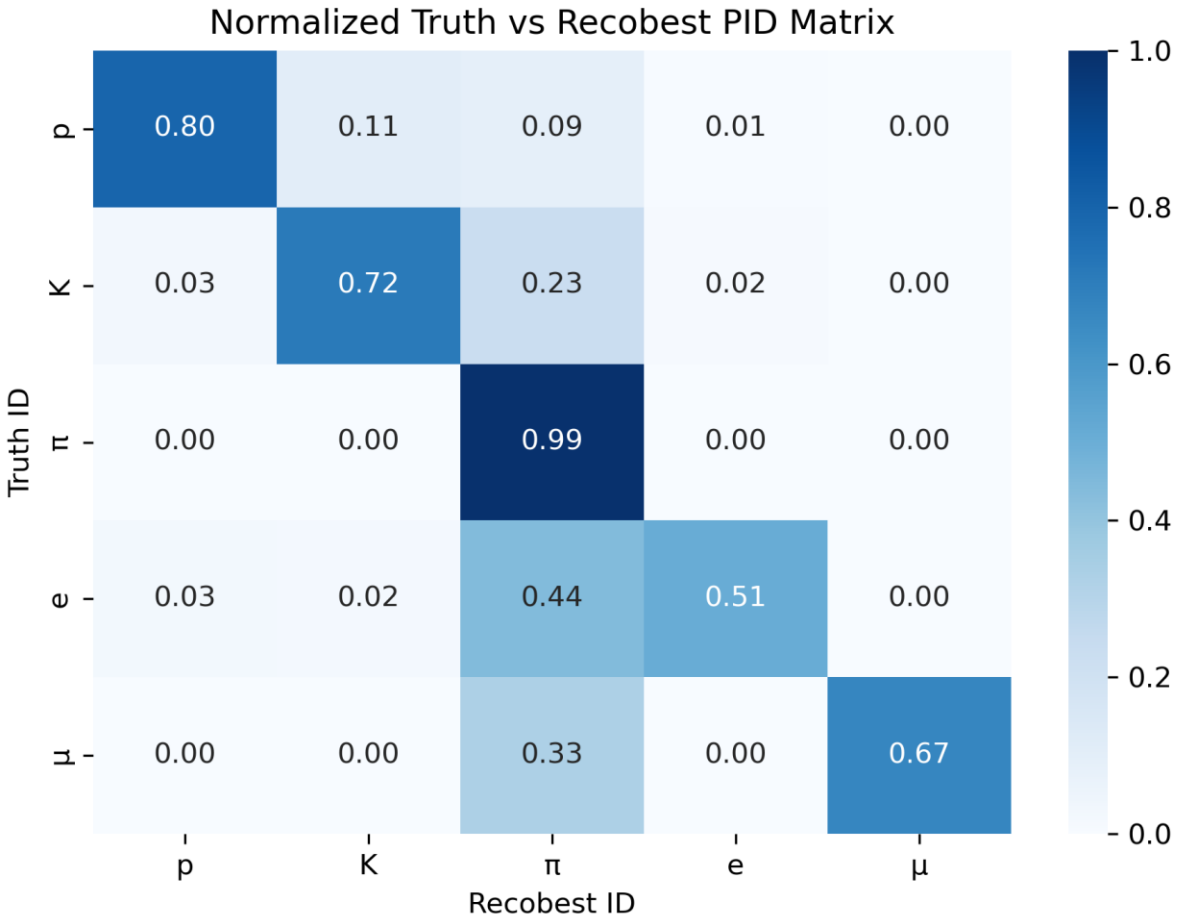
Current Result

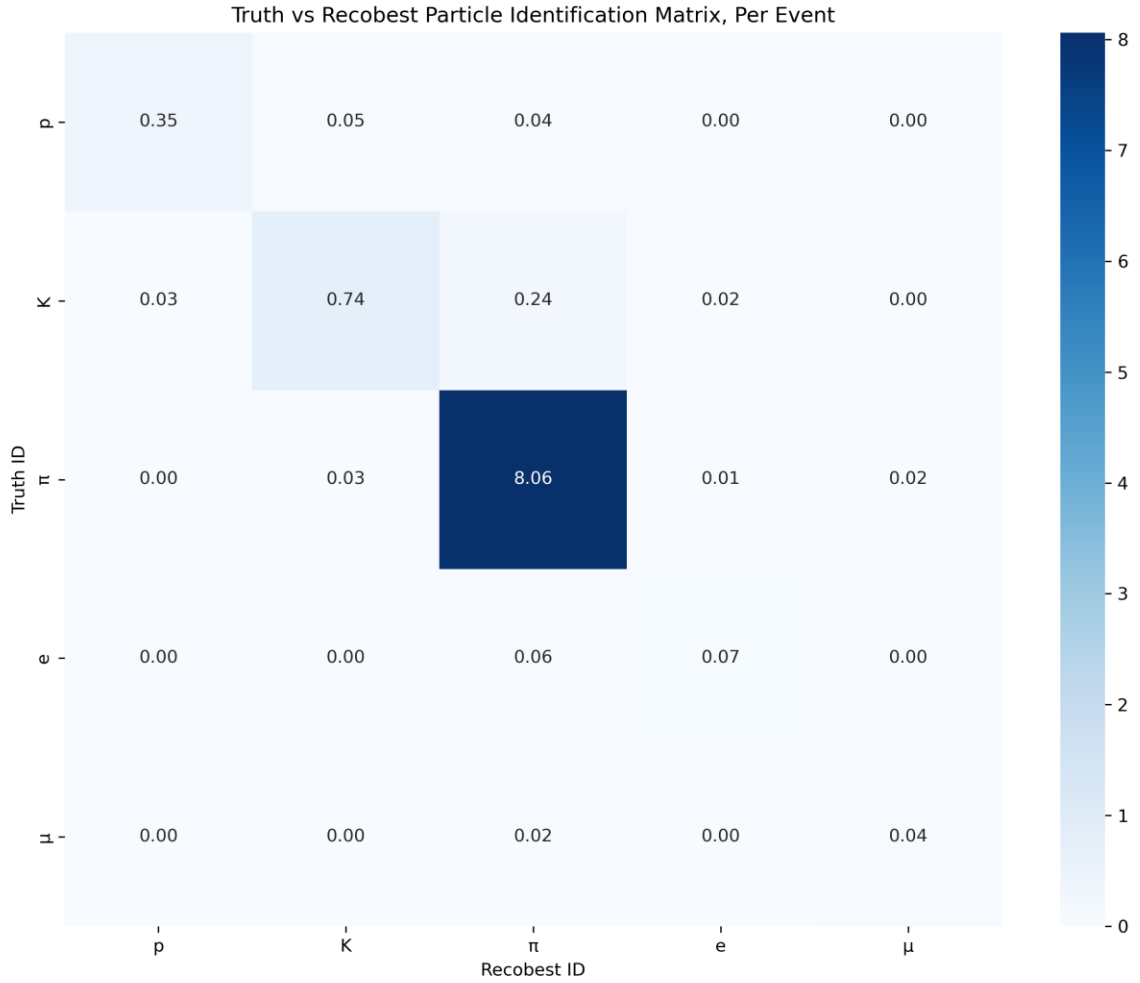
	$\sigma_{R_b}(10^{-6})$	$\sigma_{R_c}(10^{-6})$	$\sigma_{R_q}(10^{-6})$	**flavor tagging method**
LEP+SLC	659	3015	-	-
FCC-ee	2.1(0.3)	-	-	-
Template fit	1.2	2.3	2.1	LCFIPlus
Double + Single Three Categories	1.3	1.4	-	Particle Net 10^{11} events
Double + Single Two Categories	0.52	-	-	ParT (PFOAna) 10^{10} events
Double + Single Three Categories	1.32	1.51	-	ParT (PFOAna) 10^{10} events

XGBoost: In Zqq.

/hpcfs/cepc/higgsgpu/zhangkl/datasets_Zqq/999*.root

Without Energy cut,Eff*Purity: 0.718, 0.650, 0.950, 0.369, 0.413

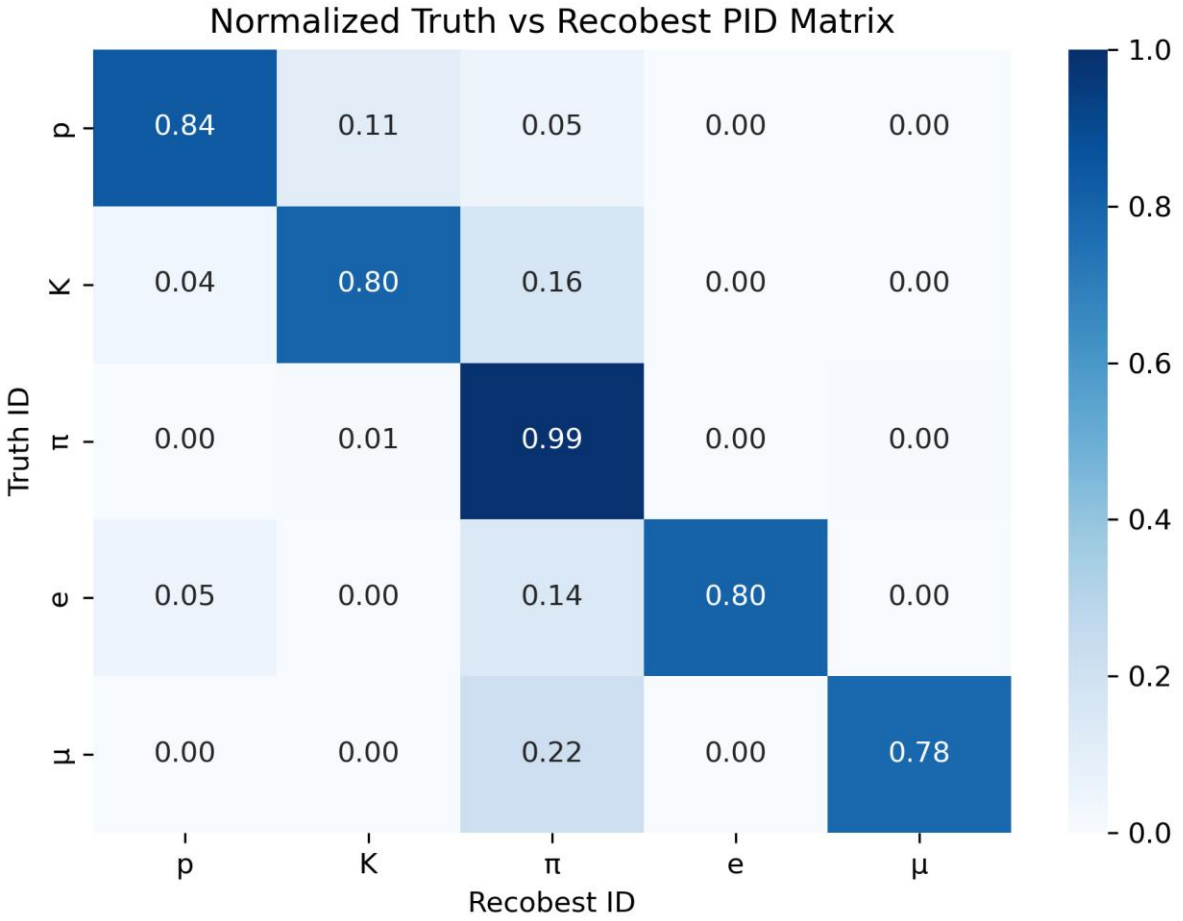


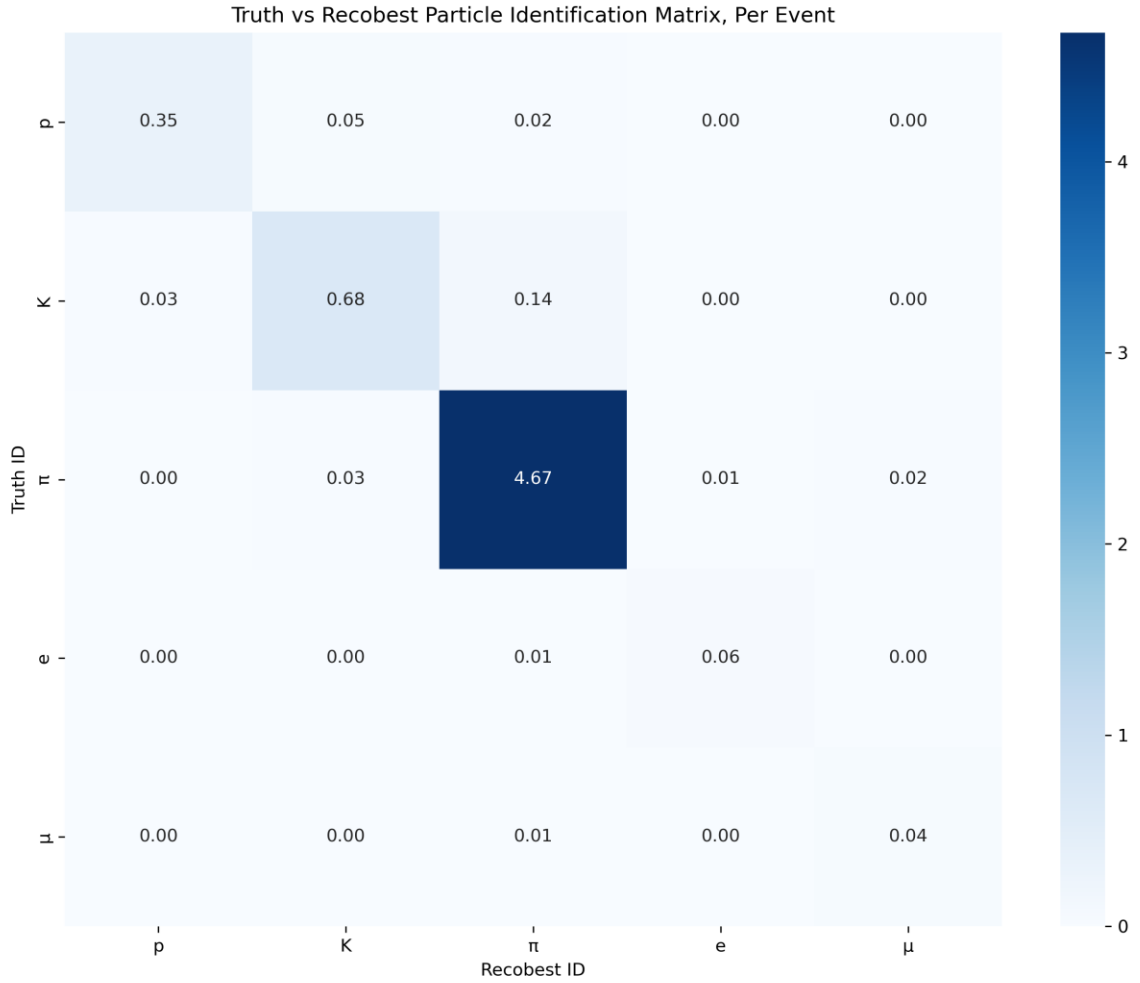


Zqq with E>1 GeV:

Reco_best	Efficiency	Purity	Eff*Purity
p	0.837	0.901	0.754
K	0.798	0.905	0.722
π	0.988	0.963	0.951
e	0.802	0.885	0.709
μ	0.784	0.636	0.499

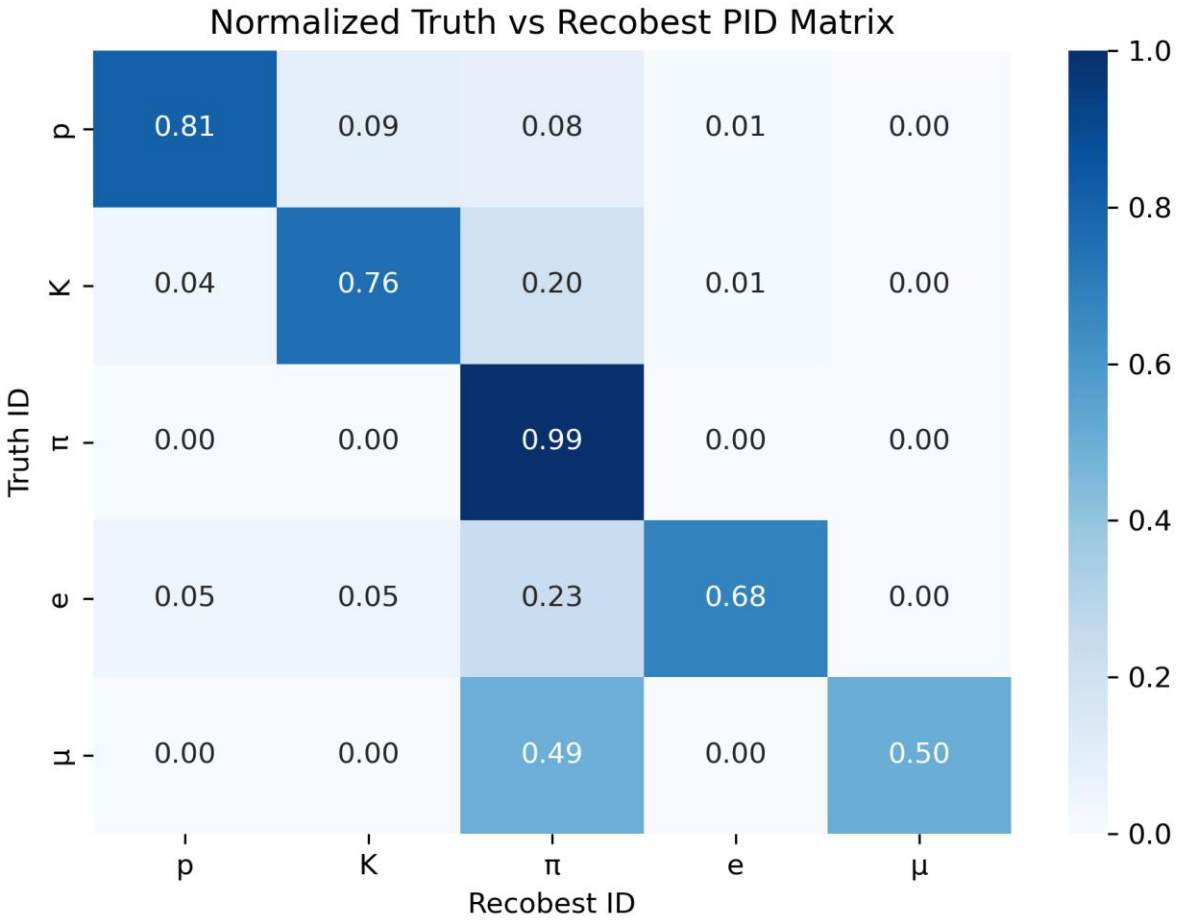
Universal ratio for bcuds jets.

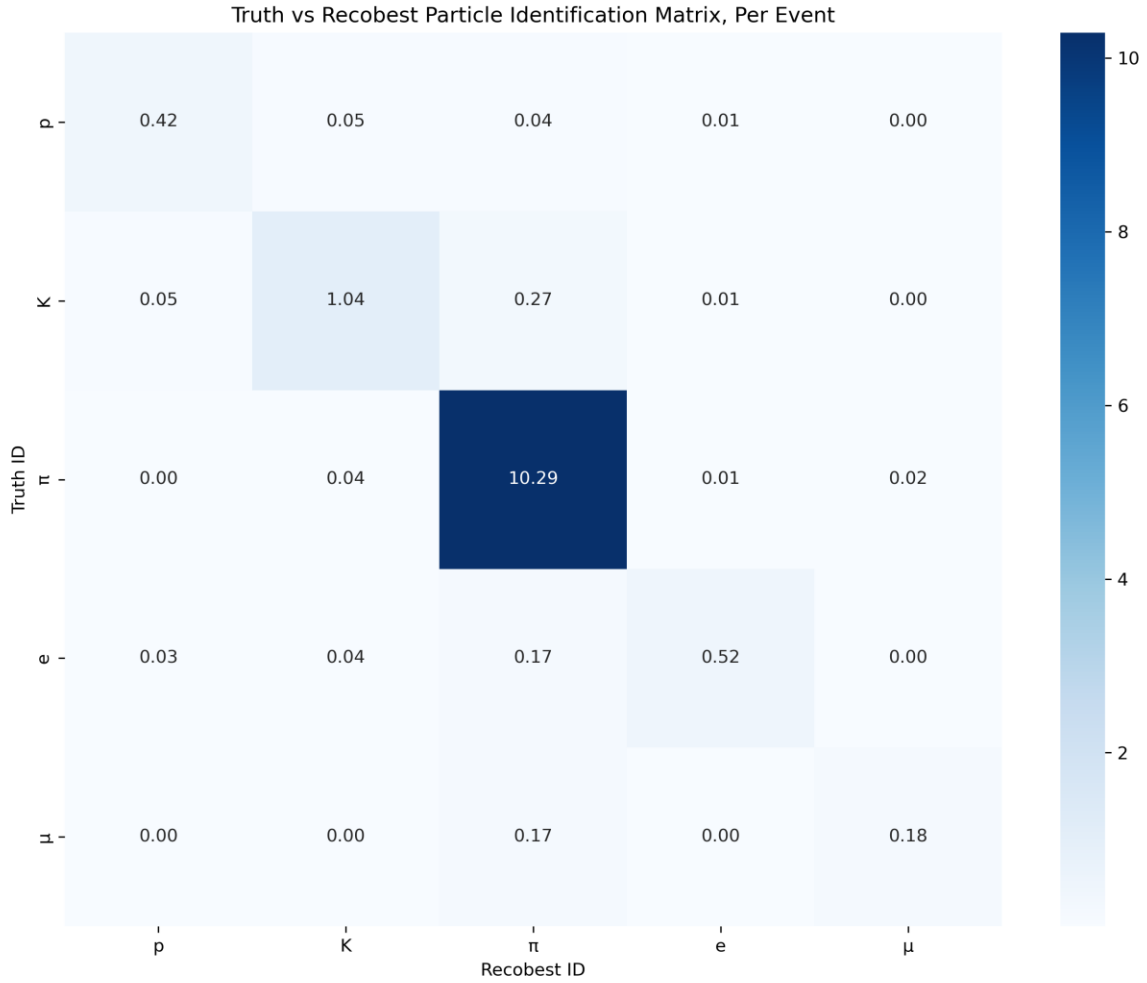




Hbb

b jets with more leptons.

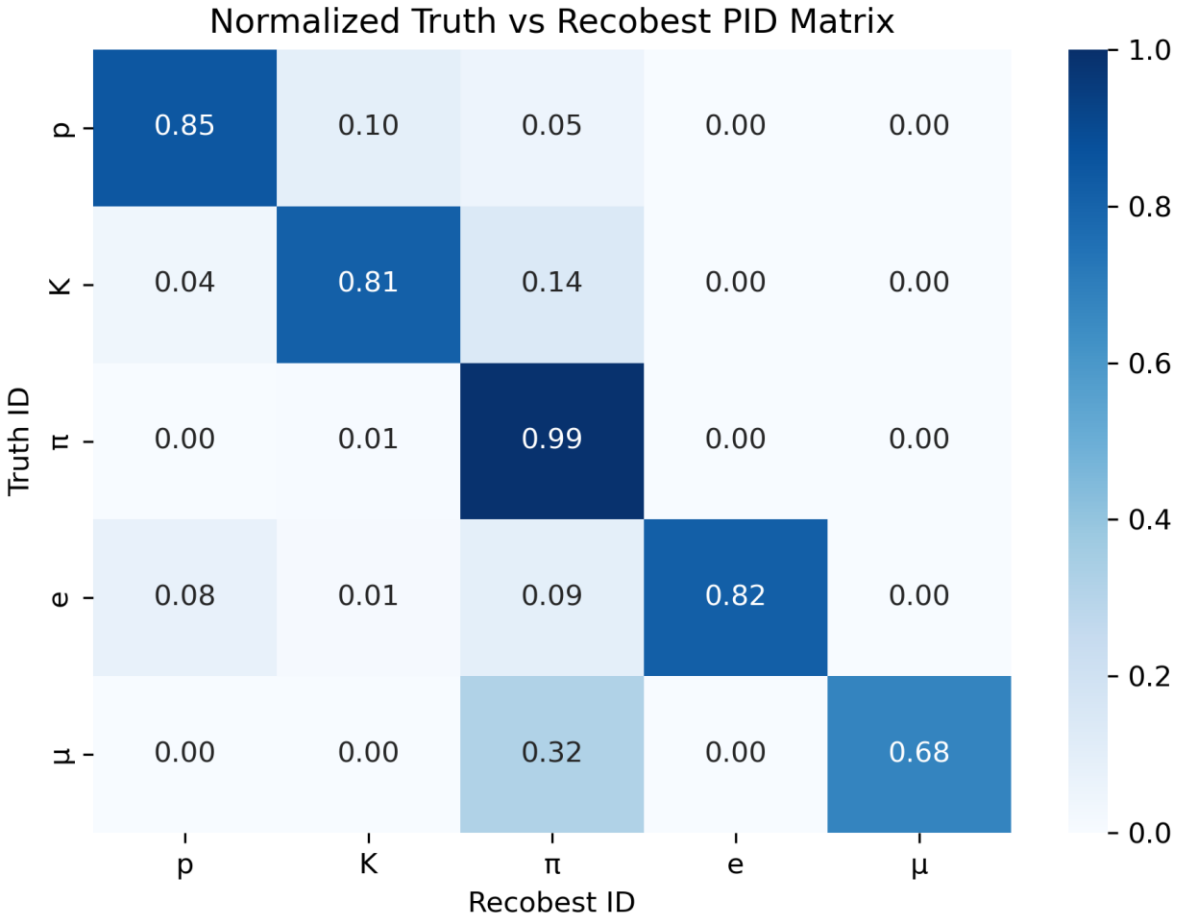


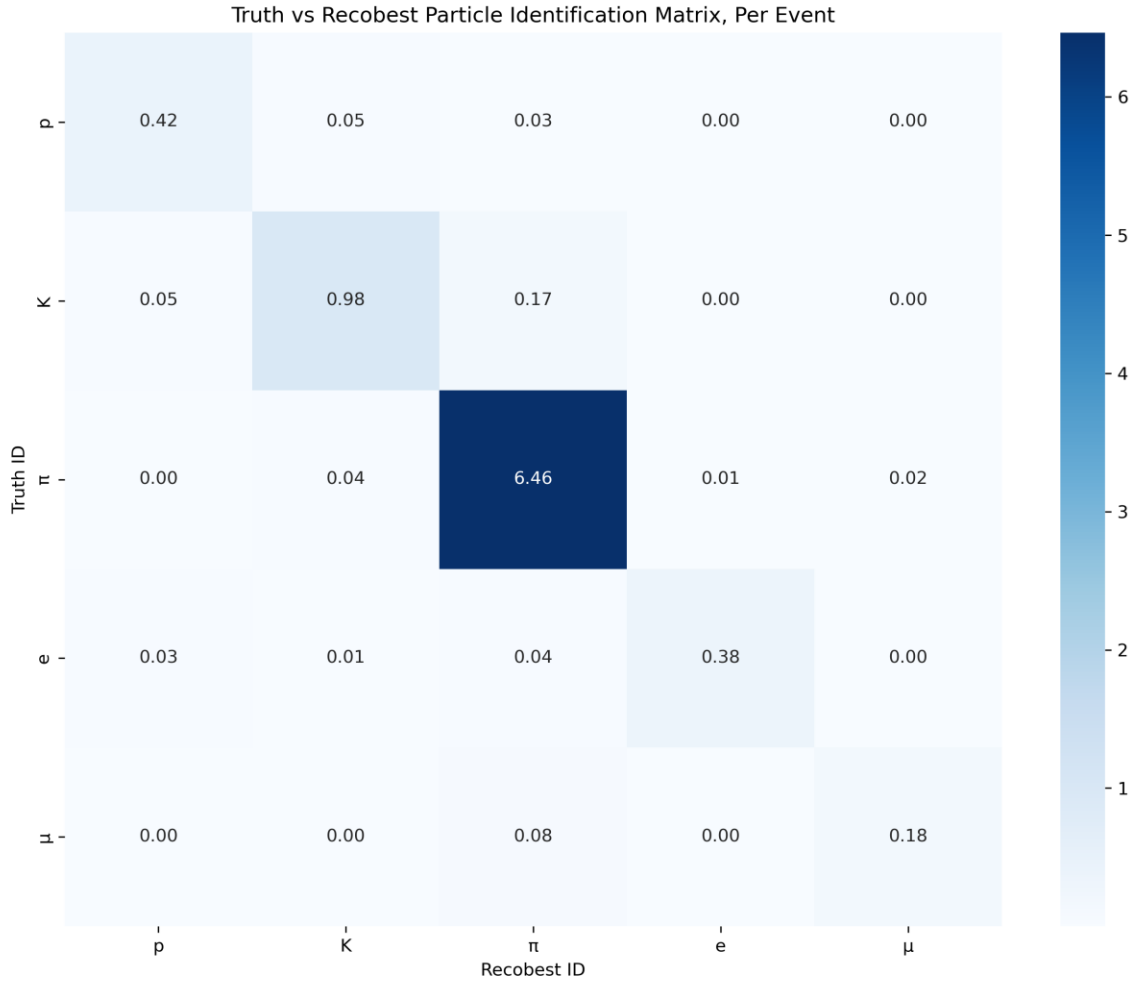


Hbb, with E>1 GeV:

Reco_best	Efficiency	Purity	Eff*Purity
p	0.851	0.827	0.704
K	0.813	0.914	0.744
π	0.99	0.952	0.943
e	0.818	0.975	0.798
μ	0.677	0.906	0.613

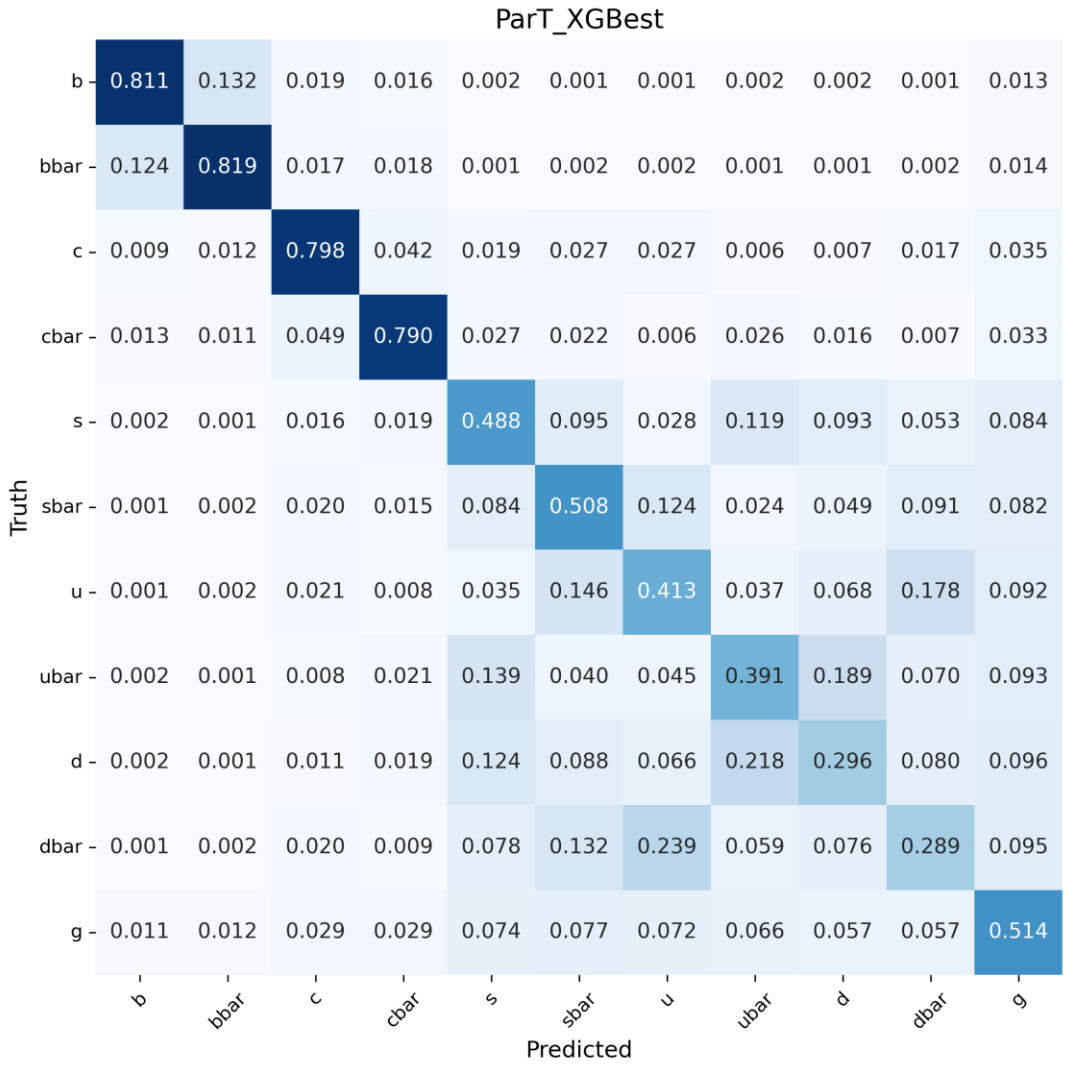
Lepton enhanced performance.

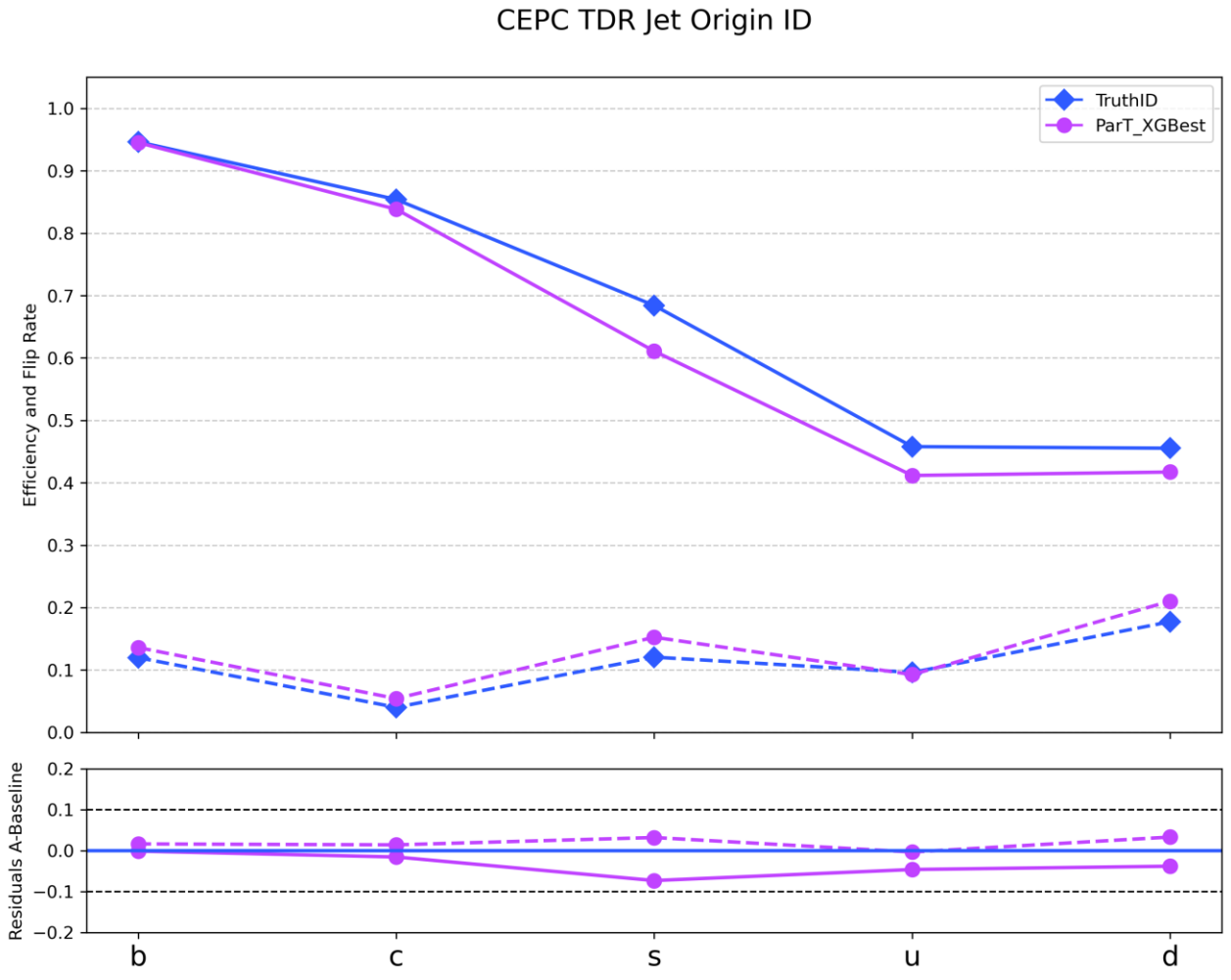




JOI for XGBoost Best Pid

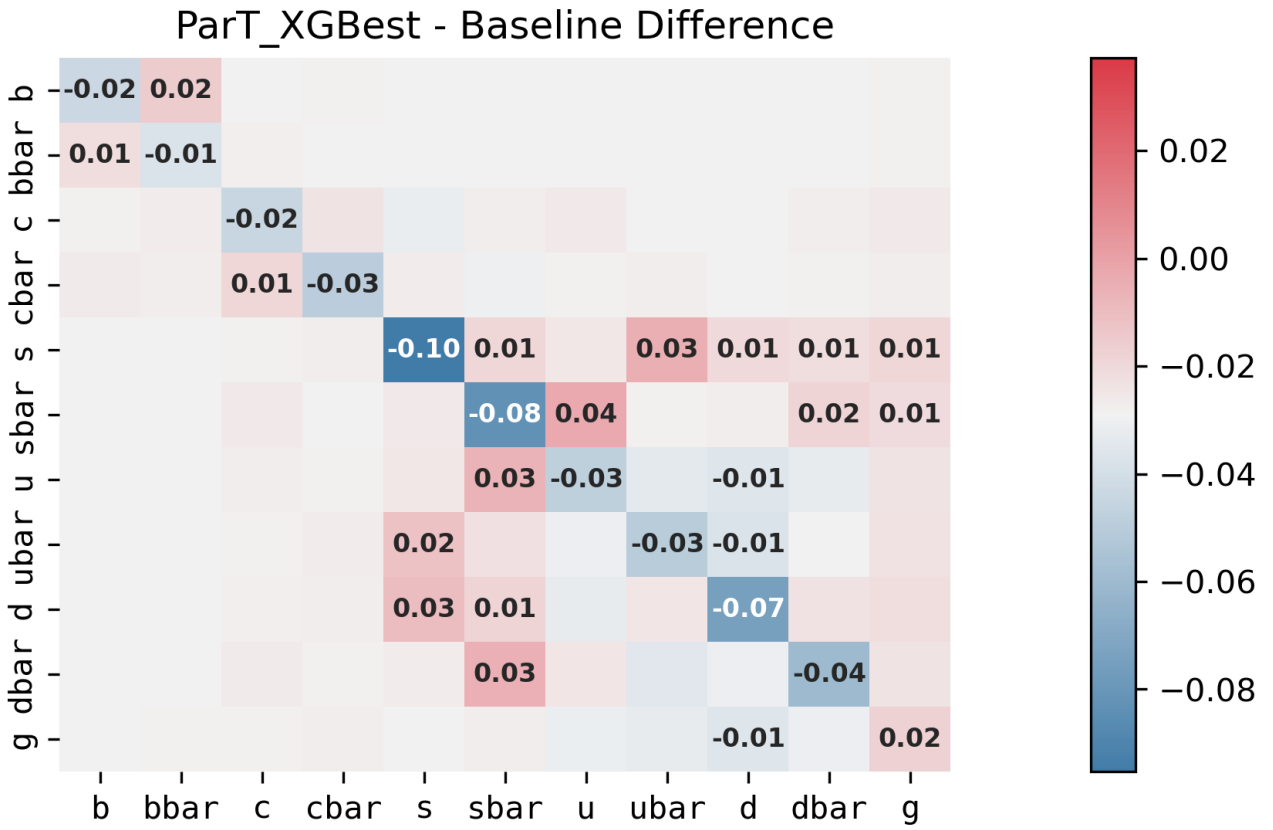
Truth ID:	0.5936
Reco 70:	0.5510
XGBest:	0.5562

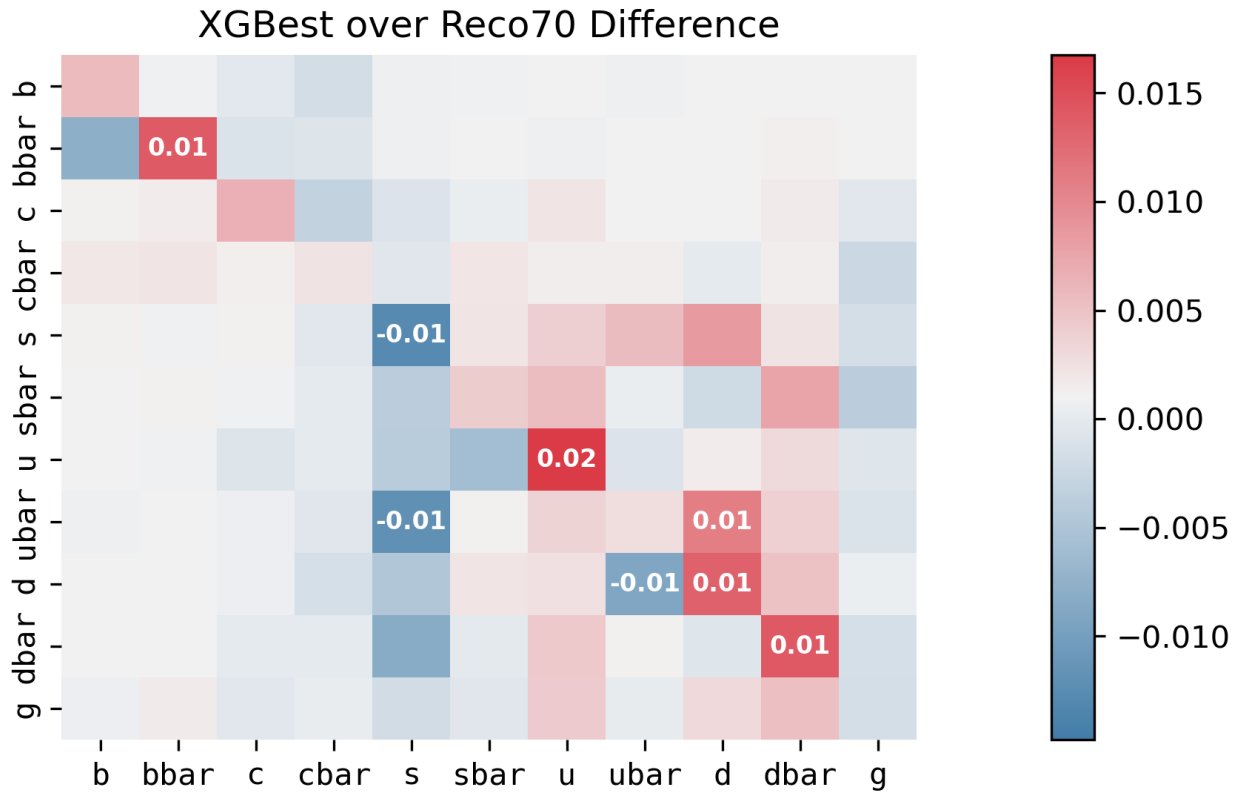




Matrix difference with Truth, Reco70

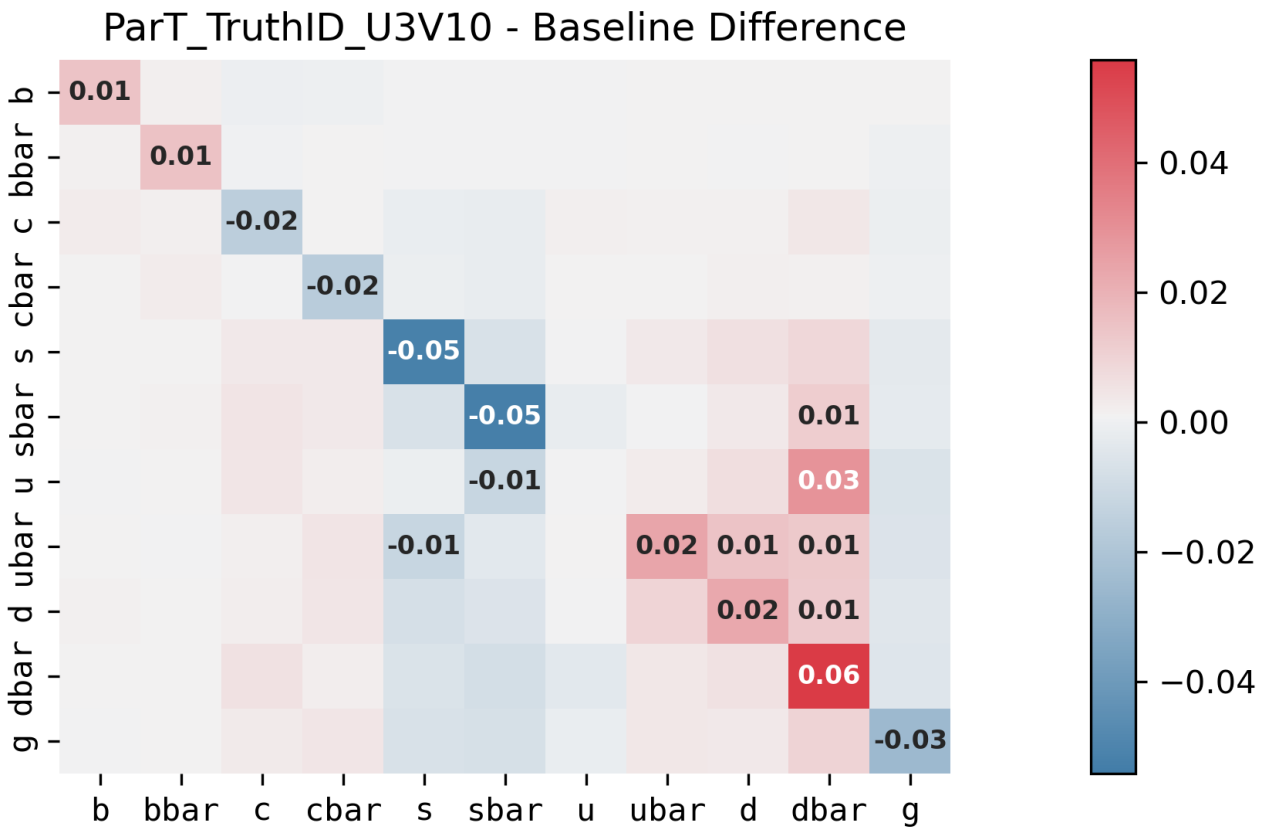
Compared to truth, current reco id lose ~9% efficacy in s tagging. While b,c quark performance kept (from VTX).
Compared to Reco70, XGBest gets better but the improvement is marginal.





VTX Impact for JOI

Different check in VTX U5V5 design and U3V10(alternative).
Metric: 0.5936 -> 0.5901



- VTX U5V5 also has limited space to improve.
- In Ref-TDR current setting, both reco pid and VTX is close to their best result.
- JOI performance close to the limit.

Using current PID in training:

- Chi2 Reco70: 0.5510
- XGBest: 0.55264 (Epoch 24/30, not finished yet), final should be <0.555.
- Expect XGBest has the current best reco ID performance, but improvement marginal.
- For VTX, also U5um, V5um has limited space to improve
- In current conditions, JOI performance is close to the best result.