CEPC Jet@Clusters

Kaili Zhang

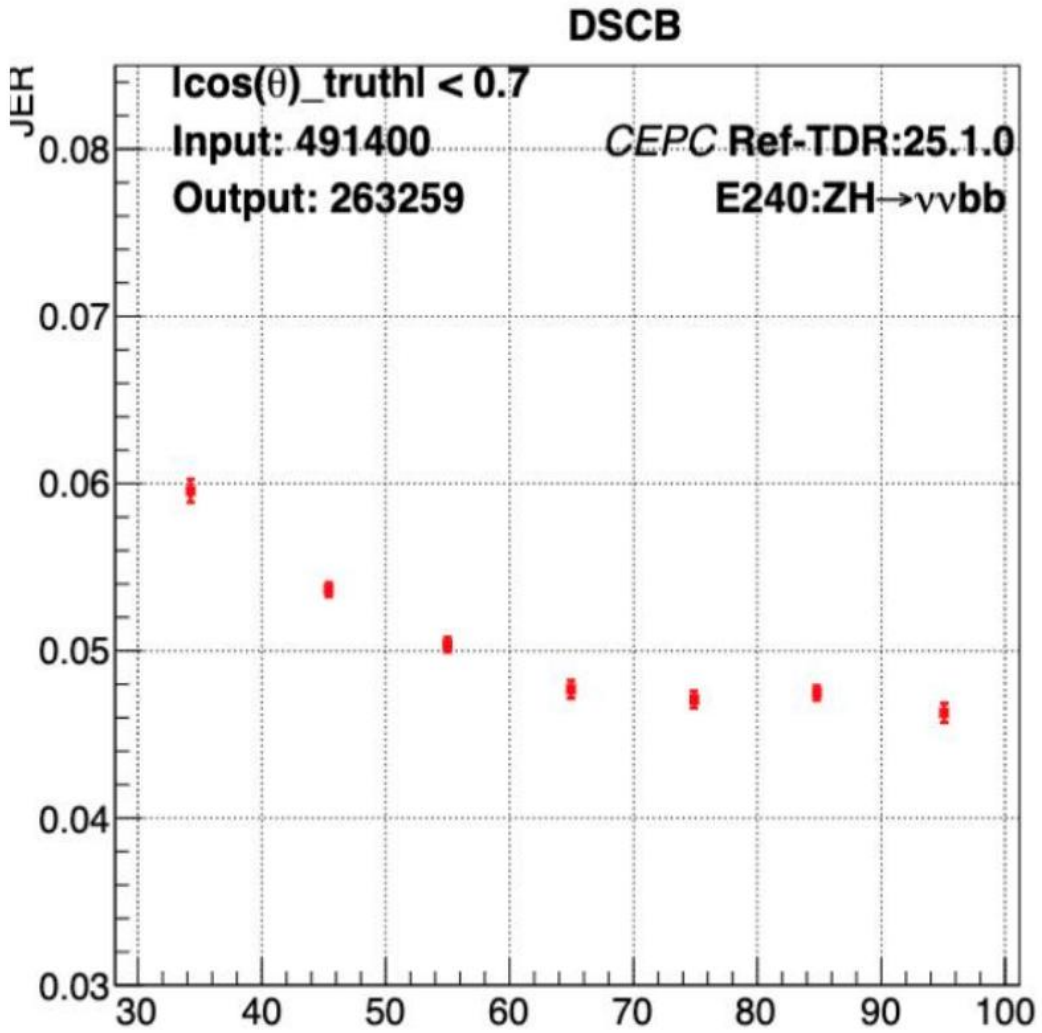
IHEP

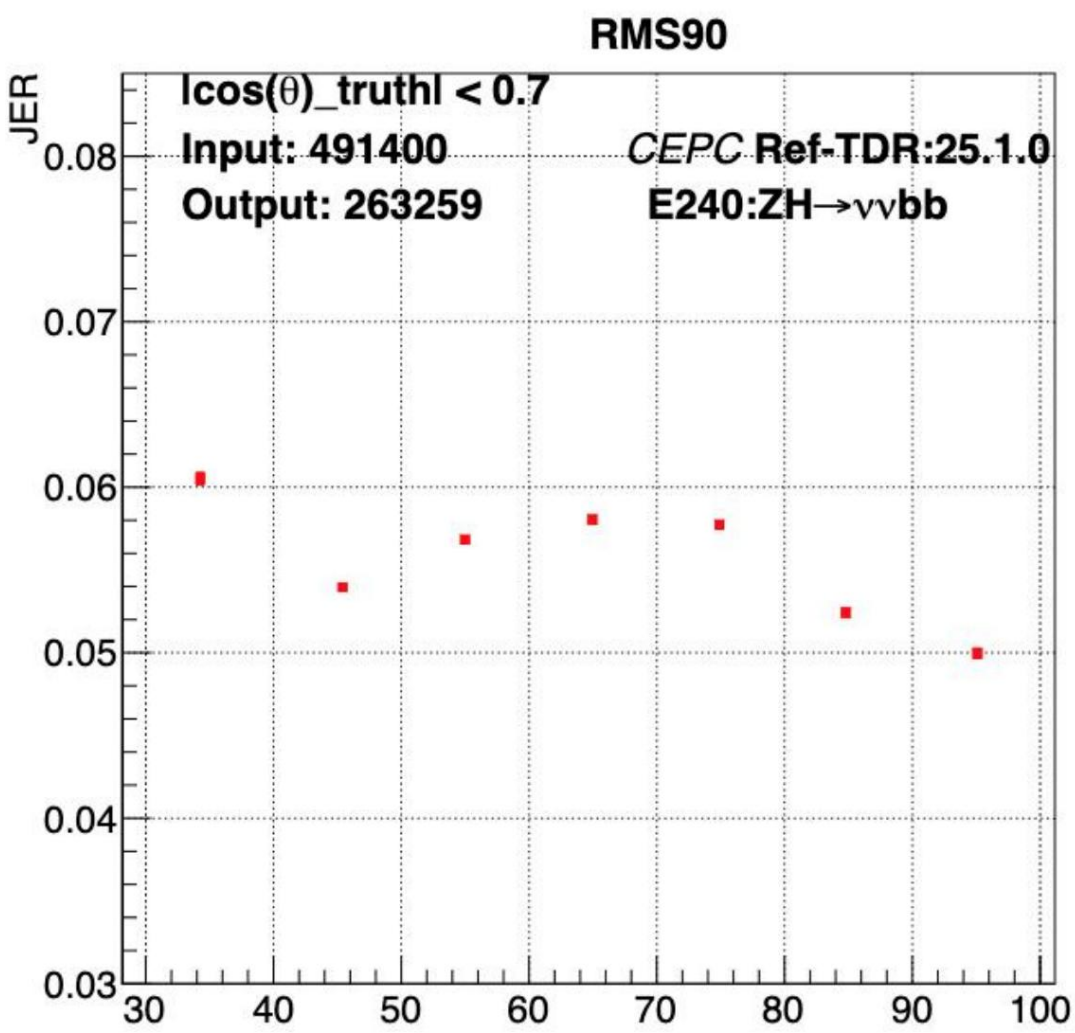
zhangkl@ihep.ac.cn

JER

@Yingqi

RMS90 using sliding window for 90% to fit.
Different behavior in ~60-90GeV.

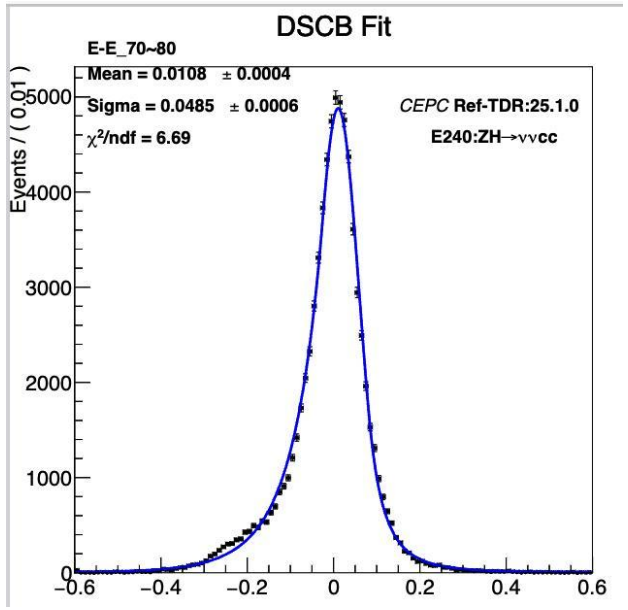




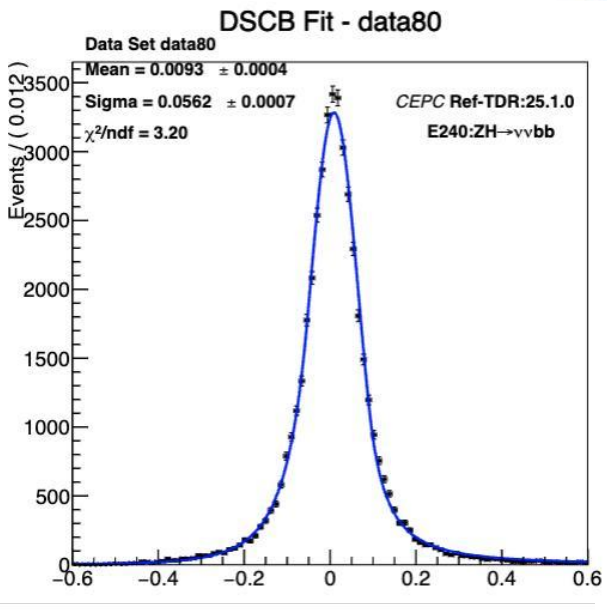
JER

E=70 GeV Asymmetry in left side

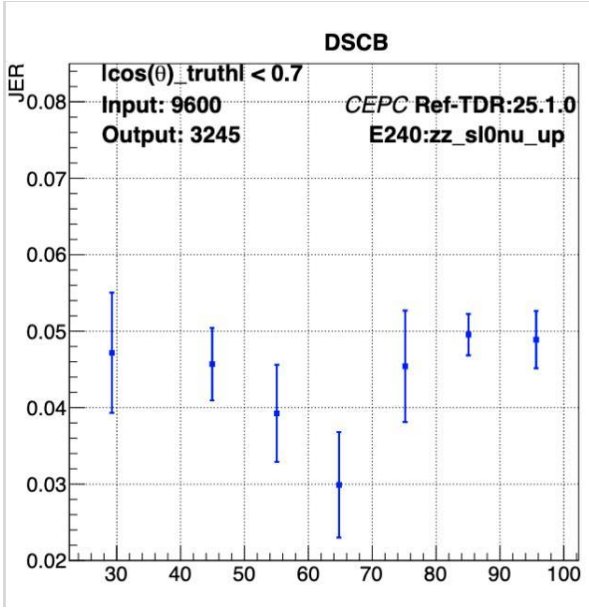
@Yingqi



ee->qq process. Choose to use sample without JER to avoid the impact of ee->Z->qq.



ee->ZZ->vvqq process.
Comparable with ZH->vvqq.
Need more statistics.

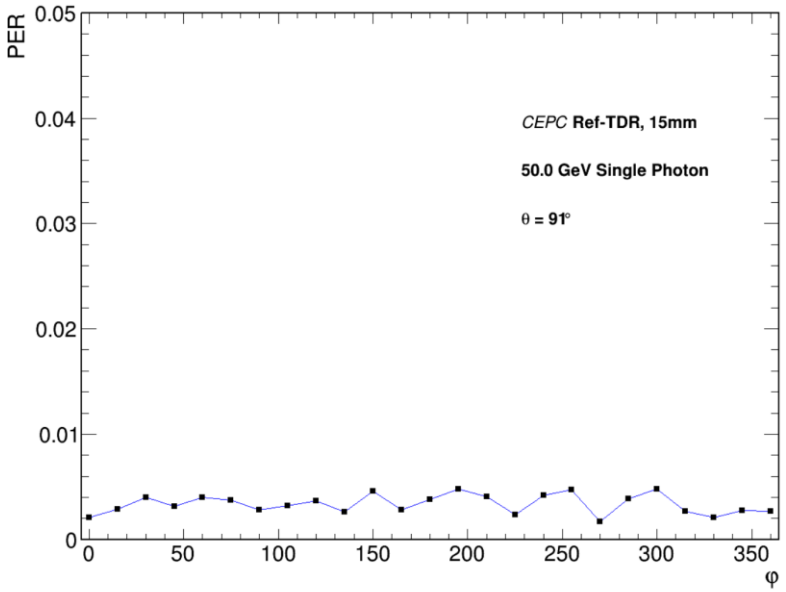


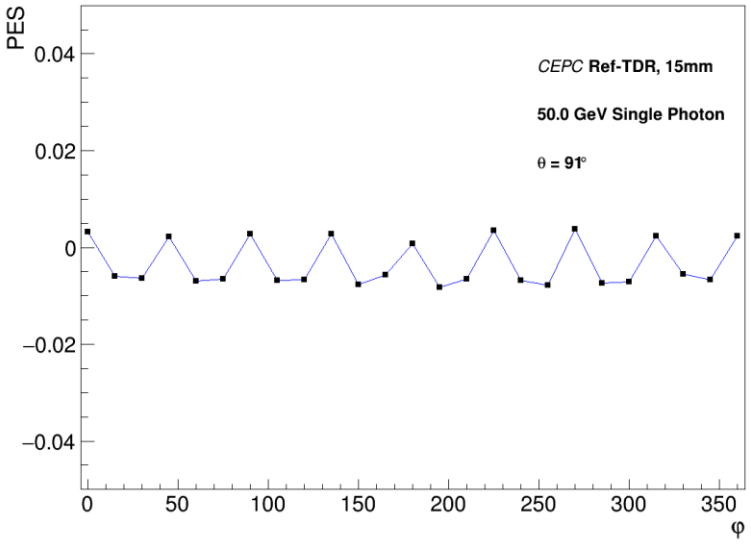
Photon scan

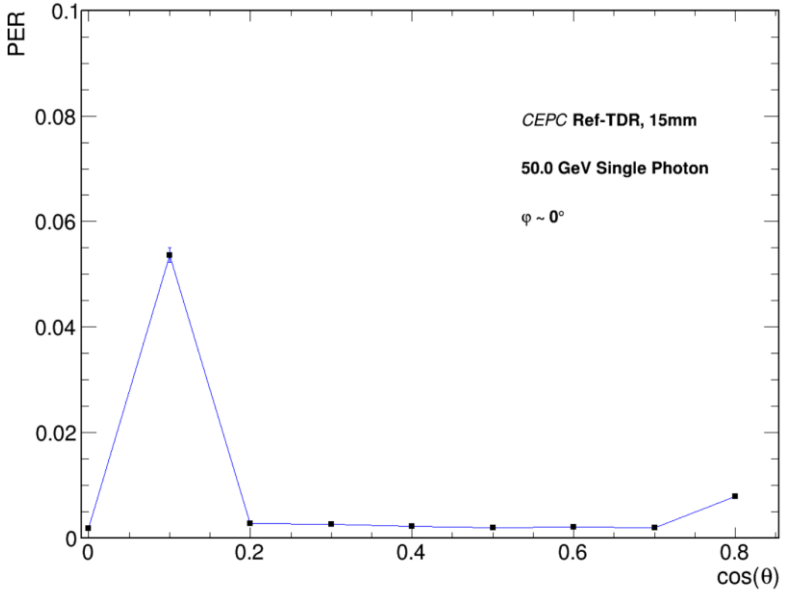
Photon scale has two pattern:

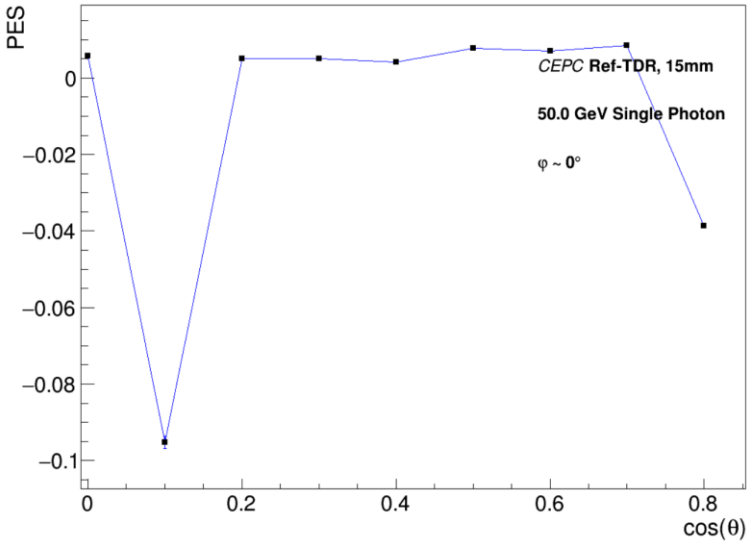
1. Generally flat on theta/phi.
2. Bad result on crack between modules.

@Reda



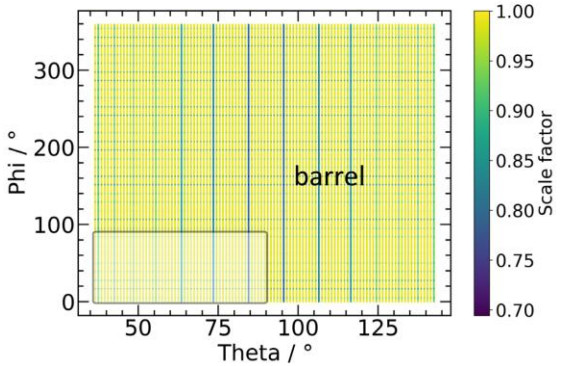


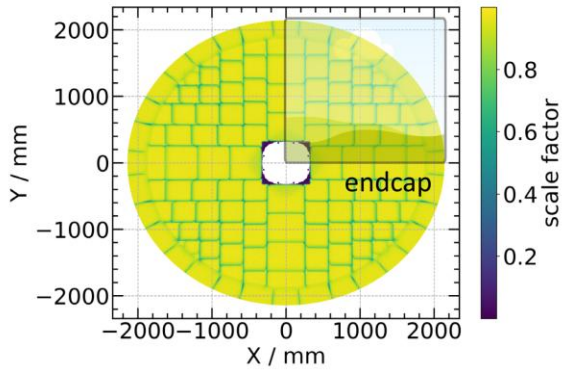




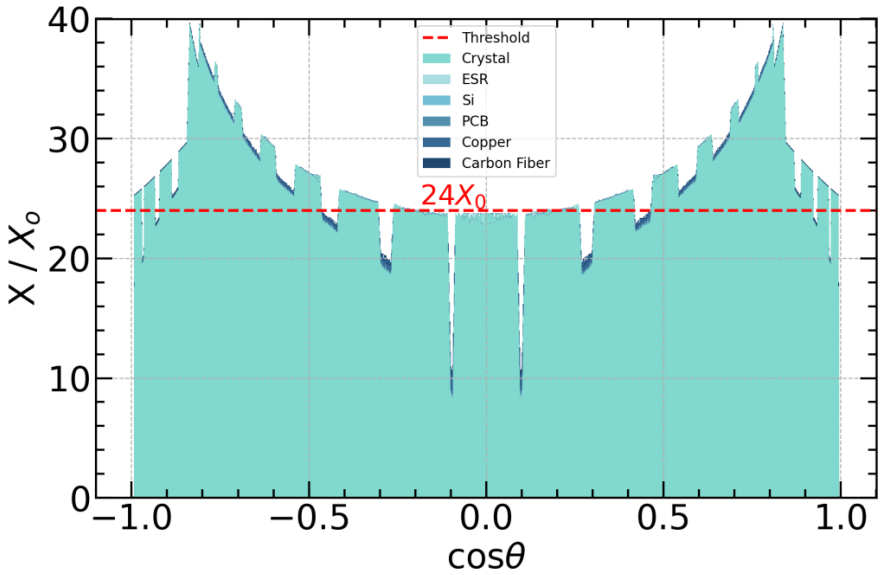
2D scan for integral

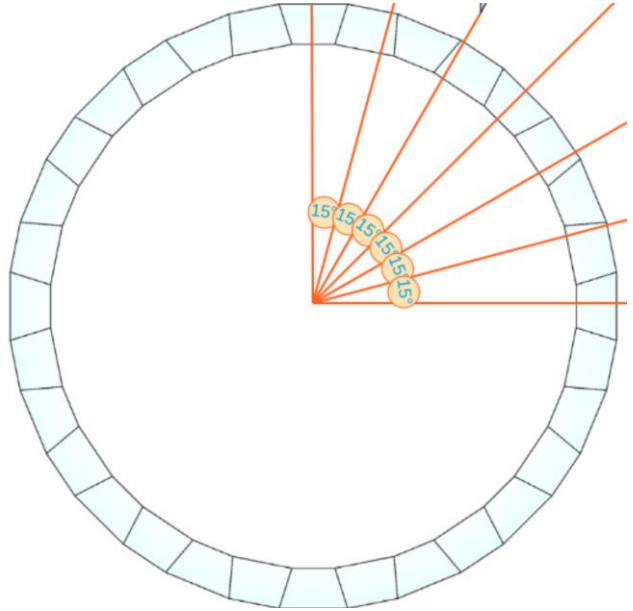
@Reda





- One 2D scan for single photon on theta/phi is undertaking in latest 15mm layout in barrel.
- To study the integral impact of crack region for resolution & scale.

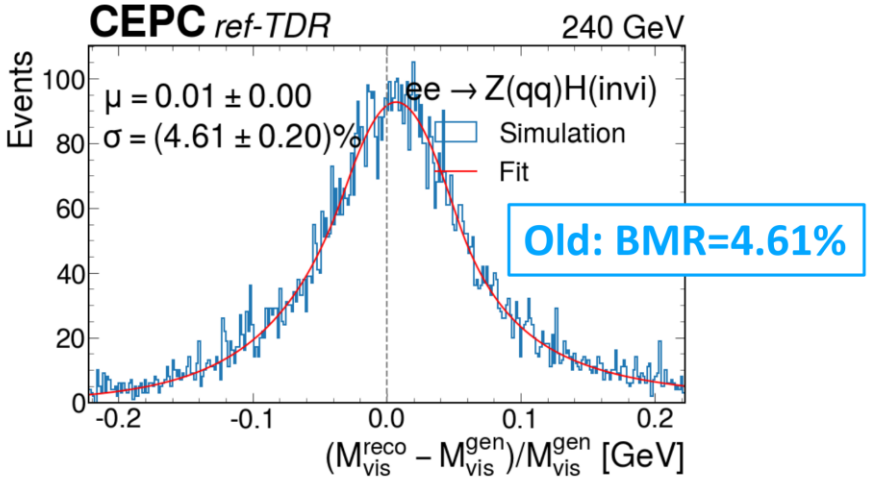


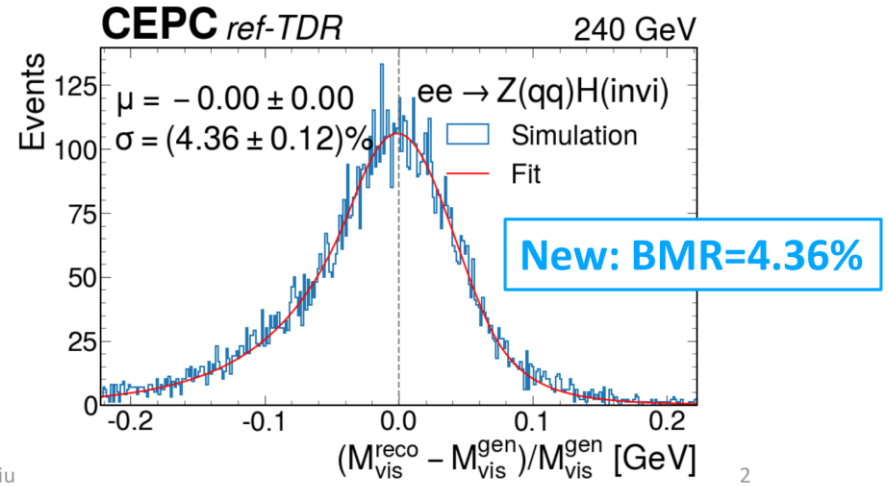


CEPCSW & BMR

- https://code.ihep.ac.cn/guofangyi/cepcsw-release/-/tree/CyberPFA-6.0.3
- Latest update on Reco PFA.
- Showing bettwe performance for BMR.
- Software group will release a new one

-> Before the end of this month.



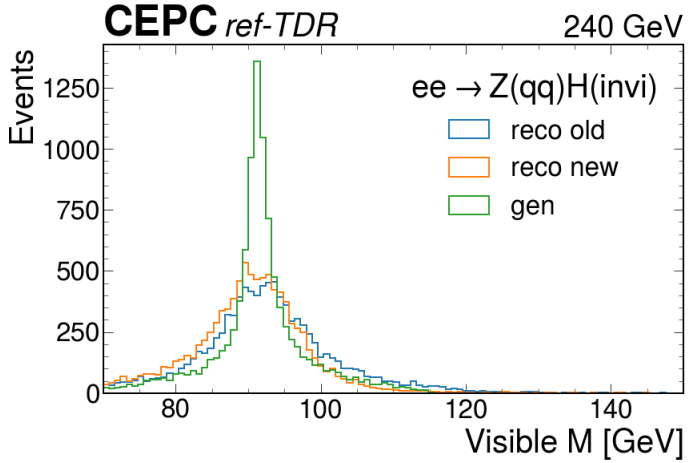


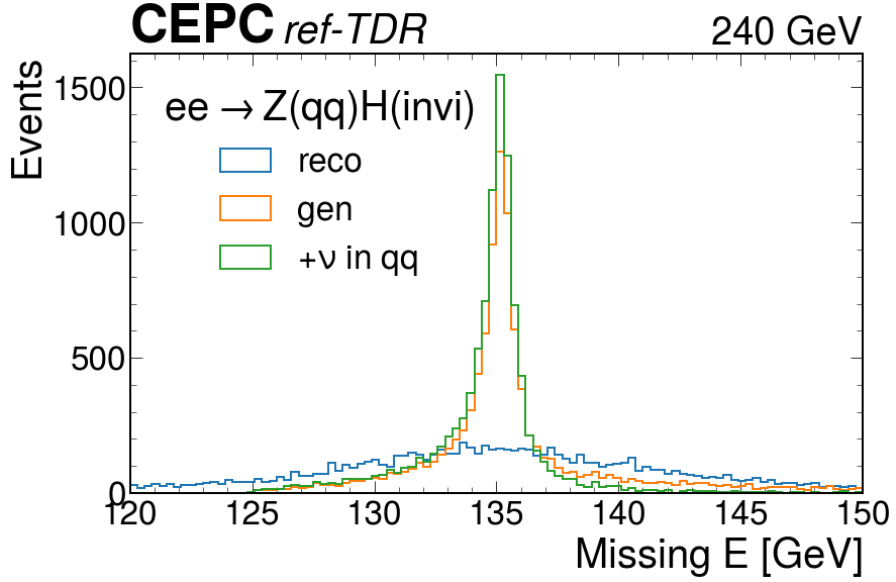
Invisible study

- With endcap calorimeter, now CEPC TDR with complete knowledge for missing Mass/Energy/MET system.

->other analysis relying on MET can join.

- Missing_E, Missing_mass can be used rather than MET.(which in LHC Pz is unknown)
- Under study which variable to use.

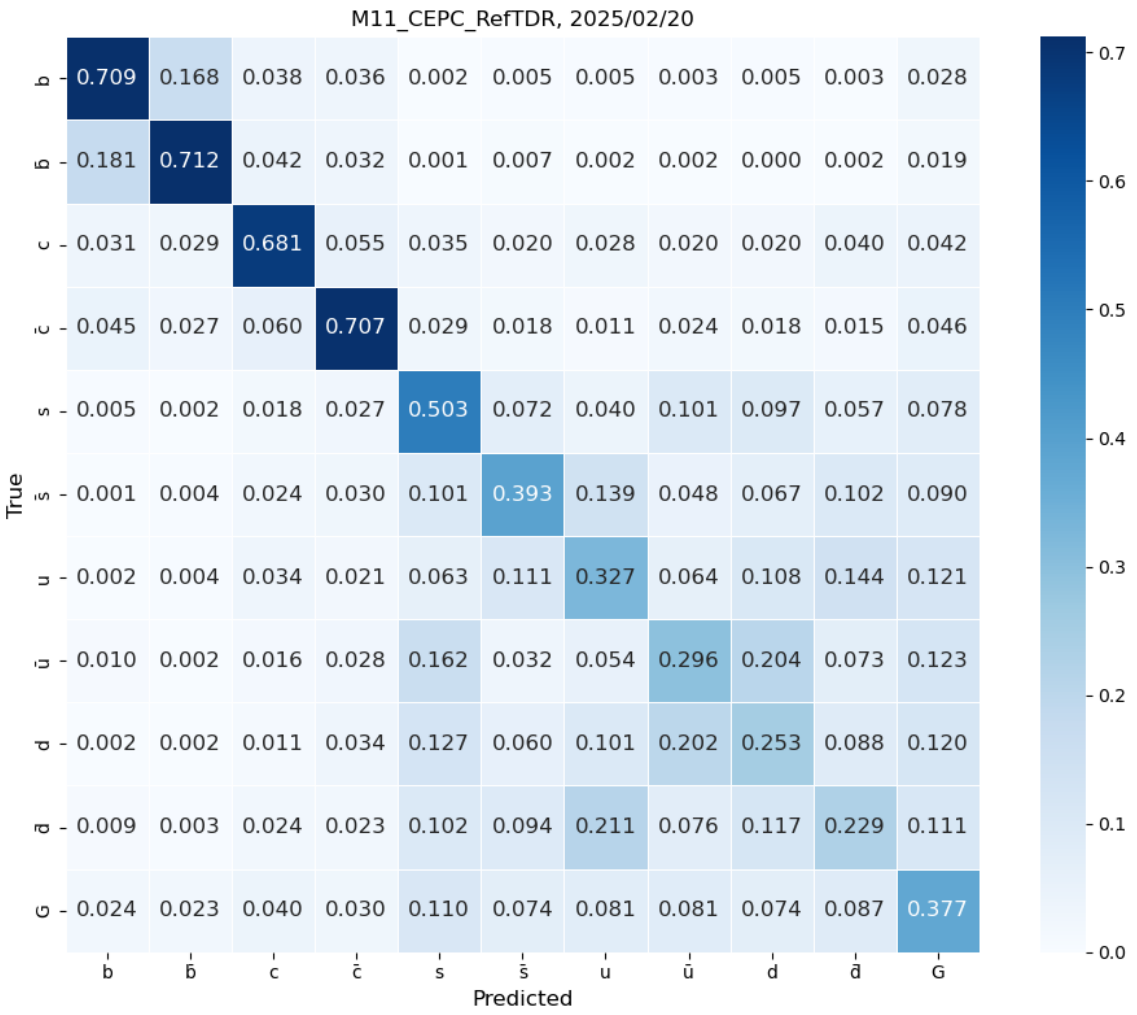




JOI

- Current result
- Truth-Reco matching eff
- Input variable
  - TDR-CDR Compare
- Scaling Law
  - Overtraining
- Application

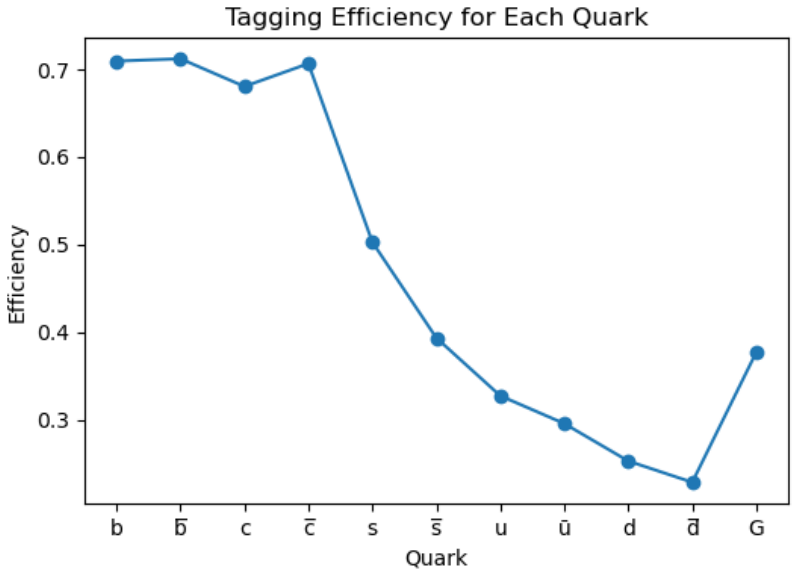
Current Best JOI

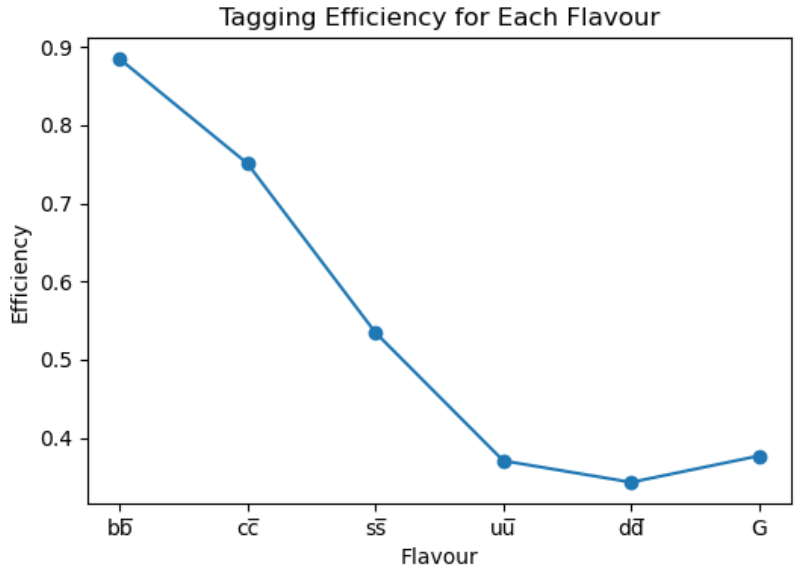


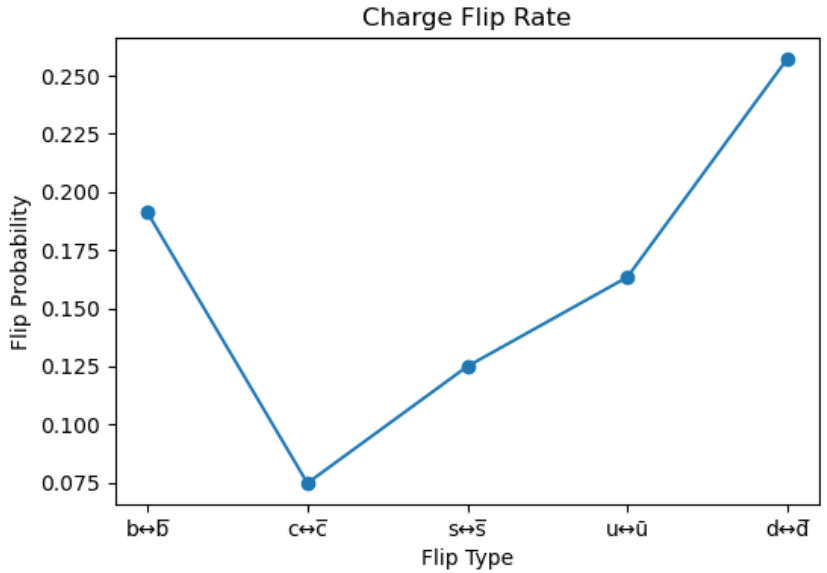
- This result is done in:
- Length: 40. Batchsize: 40
- Total sample size: 980W
- Train: Test: Validation=8:1:1

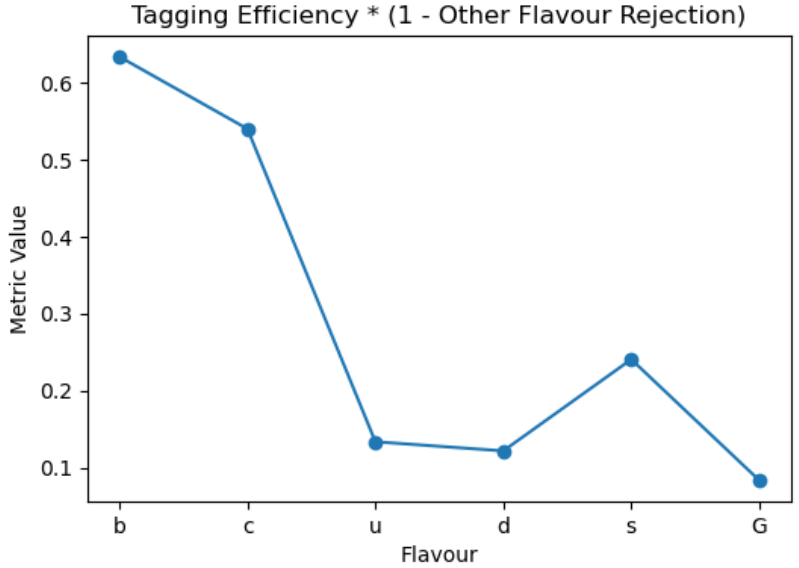
- With overtraining issue. (Epoch #2 gives best result. Following worse;)
- Under tuning. But with ~89% btagging eff.

Current JOI

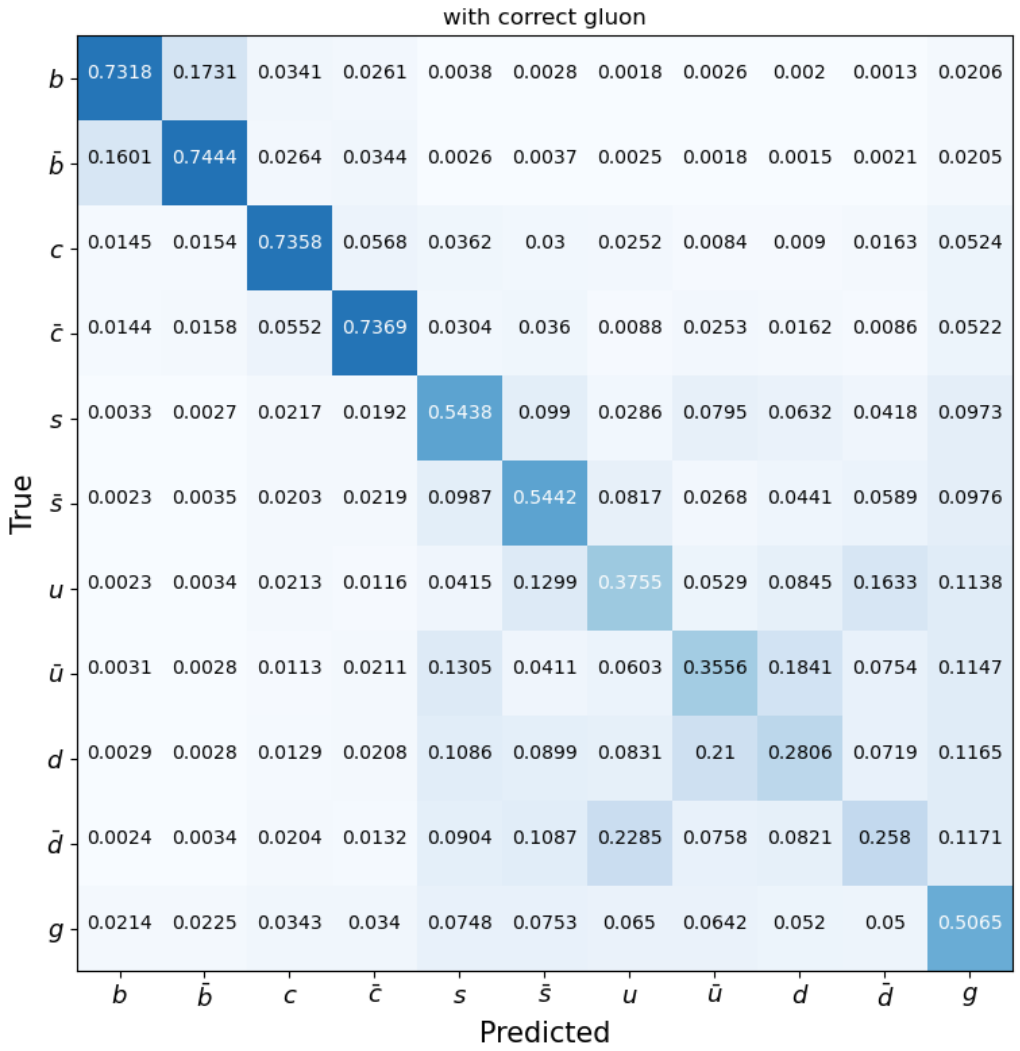

Tagging Efficiency for Each Quark
Efficiency
0.7
0.6
0.5
0.4
0.3
b
b̄
c
c̄
s
s̄
u
ū
d
d̄
G
Quark

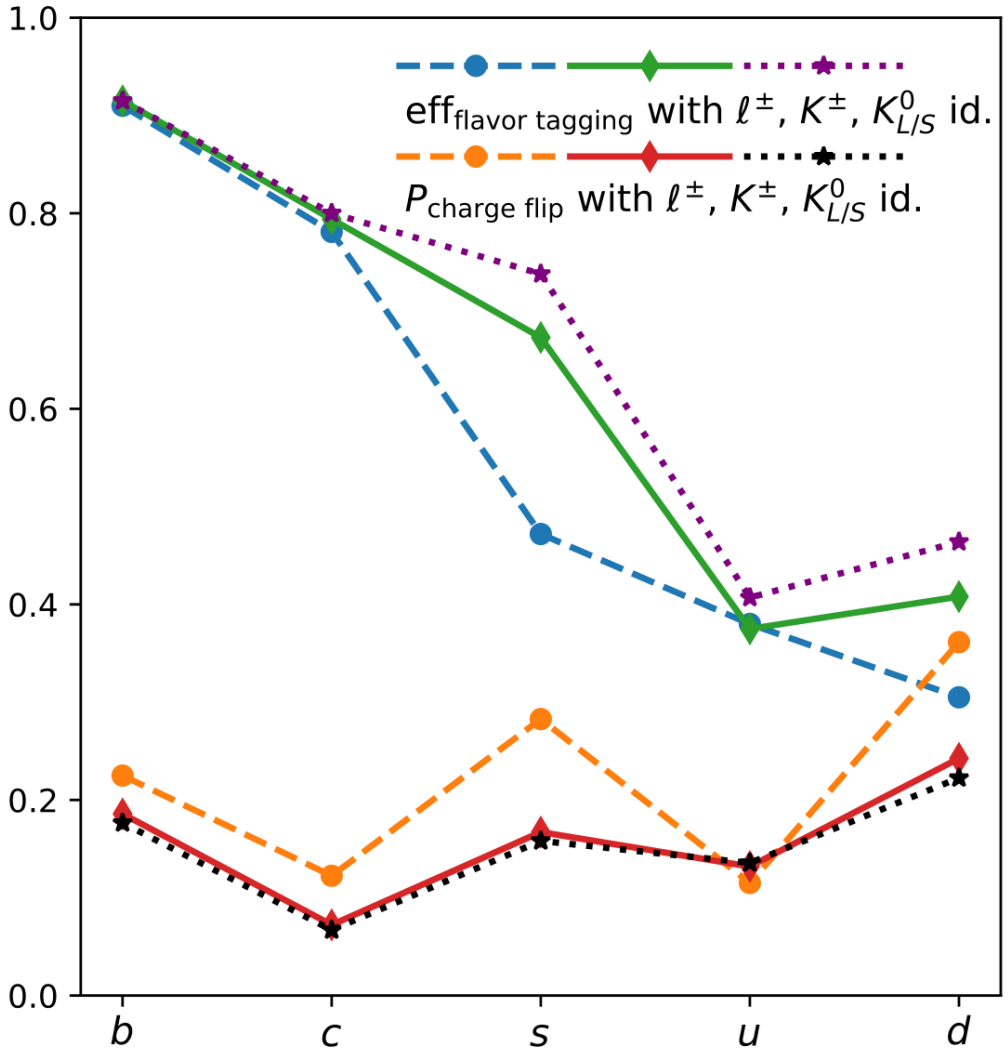

Tagging Efficiency for Each Flavour
Efficiency
0.9
0.8
0.7
0.6
0.5
0.4
bb̄
cc̄
ss̄
uū
dd̄
G
Flavour


Charge Flip Rate
Flip Probability
0.250
0.225
0.200
0.175
0.150
0.125
0.100
0.075
b↔b̄
c↔c̄
s↔s̄
u↔ū
d↔d̄
Flip Type


Tagging Efficiency * (1 - Other Flavour Rejection)
Metric Value
0.6
0.5
0.4
0.3
0.2
0.1
b
c
u
d
s
G
Flavour

CDR


with correct gluon
True
Predicted
b 0.7318 0.1731 0.0341 0.0261 0.0038 0.0028 0.0018 0.0026 0.002 0.0013 0.0206
b̄ 0.1601 0.7444 0.0264 0.0344 0.0026 0.0037 0.0025 0.0018 0.0015 0.0021 0.0205
c 0.0145 0.0154 0.7358 0.0568 0.0362 0.03 0.0252 0.0084 0.009 0.0163 0.0524
c̄ 0.0144 0.0158 0.0552 0.7369 0.0304 0.036 0.0088 0.0253 0.0162 0.0086 0.0522
s 0.0033 0.0027 0.0217 0.0192 0.5438 0.099 0.0286 0.0795 0.0632 0.0418 0.0973
s̄ 0.0023 0.0035 0.0203 0.0219 0.0987 0.5442 0.0817 0.0268 0.0441 0.0589 0.0976
u 0.0023 0.0034 0.0213 0.0116 0.0415 0.1299 0.3755 0.0529 0.0845 0.1633 0.1138
ū 0.0031 0.0028 0.0113 0.0211 0.1305 0.0411 0.0603 0.3556 0.1841 0.0754 0.1147
d 0.0029 0.0028 0.0129 0.0208 0.1086 0.0899 0.0831 0.21 0.2806 0.0719 0.1165
d̄ 0.0024 0.0034 0.0204 0.0132 0.0904 0.1087 0.2285 0.0758 0.0821 0.258 0.1171
g 0.0214 0.0225 0.0343 0.034 0.0748 0.0753 0.065 0.0642 0.052 0.05 0.5065
b b̄ c c̄ s s̄ u ū d d̄ g


1.0
0.8
0.6
0.4
0.2
0.0
$\text{eff}_{\text{flavor tagging}}$ with $\ell^{\pm}$, $K^{\pm}$, $K^0_{L/S}$ id.
$P_{\text{charge flip}}$ with $\ell^{\pm}$, $K^{\pm}$, $K^0_{L/S}$ id.
b c s u d

Truth Matching Eff

/cefs/higgs/zhangkl/CEPCSW/Analysis/JetOrigin/src

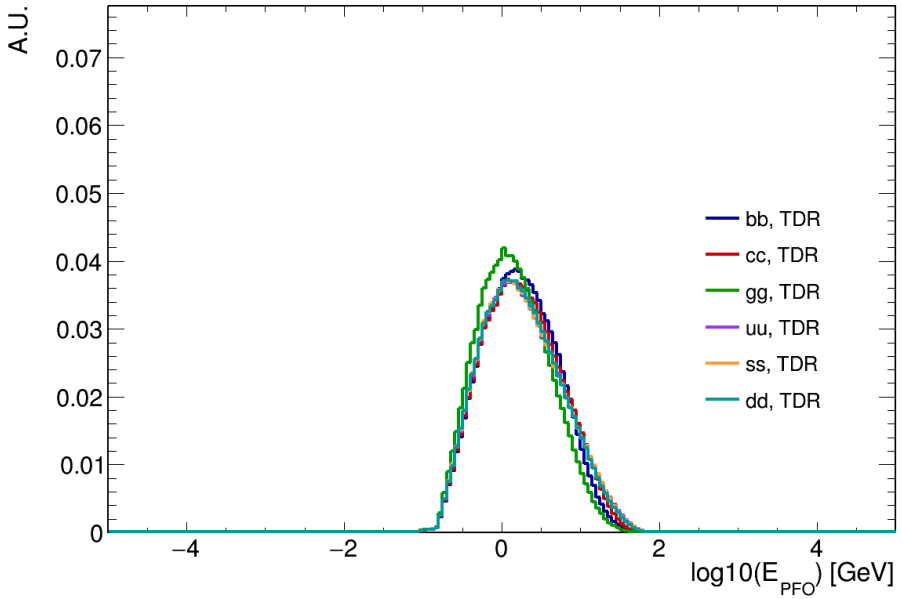
- Current CEPCSW do not have reliable truth link between reco PFO and mc truth.
- Use DeltaR and DeltaE to match.
  - Currently, for each charged track, eff >90%
- Fail case:
  - All stable truth particle already matched with additional track;
  - Track with small energy (like, <0.5GeV)
  - Other under tuning issue.

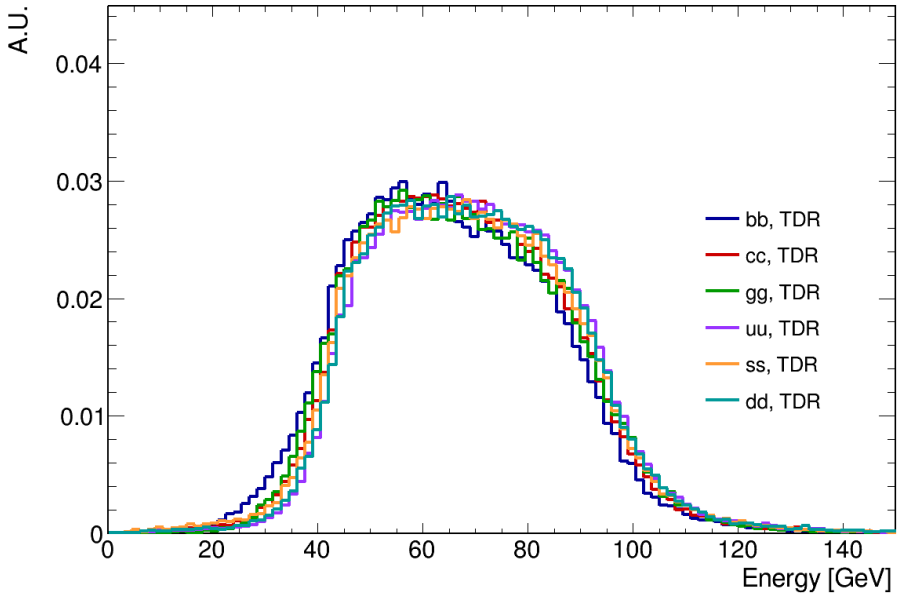
Debug matching info

```
JetOrigin           INFO  Matched with truth particle Energy: 0.321898: 0.310204 MCIndex: 107
JetOrigin           INFO  (reco-truth)/truth E: -0.0363294
JetOrigin           INFO  (reco-truth)/truth Pt: -0.0450998
JetOrigin           INFO  (reco-truth)/truth Px: -0.0438147
JetOrigin           INFO  (reco-truth)/truth Py: -0.106302
JetOrigin           INFO  (reco-truth)/truth Pz: -0.0448689
JetOrigin           INFO  (reco-truth)/truth Eta: -0.000167079 (reco-truth)/truth Phi: 0.00943658 (reco-truth)/truth DeltaR: 0.00943806
JetOrigin           INFO  PFO Px: 0.198272 PFO Py: -0.0273029 PFO Pz: -0.191523 PFO E: 0.310204
JetOrigin           INFO  E_Ecal: 0        E_Hcal: 0        ESize: 0        HSize: 0
JetOrigin           INFO  D0:    -0.0924448 Z0:    -0.0011889 D0err: 0.00154263 Z0err: 0.00247036
JetOrigin           INFO  Dphi:  -0.52123 Deta:  -0.0987253 DeltaR: 0.530497
JetOrigin           INFO  PID:   5        Charge: 1
JetOrigin           INFO  MResidual: 0.253956
JetOrigin           INFO  Track: 1        Truth: 1
JetOrigin           INFO  S E_Ecal: -inf     S E_Hcal: -inf     S ESize: 0        S HSize: 0
JetOrigin           INFO  C Dphi: -3.60986 C E:   -1.85753 C Pt:  -3.10872 C relE: -9.76728 C relPt: -10.053
JetOrigin           INFO  C Z0:  -1.03412 C D0:  -2.92486 C D0err: -1.40587 C Z0err: -1.30362
JetOrigin           INFO  id: 620d570 107
 PDG : 211
 generatorStatus : 1
 simulatorStatus : 0
 charge : 1
 time : 0
 mass : 0.13957
 vertex : 0 0 0
 endpoint : 69.2728 -21.2372 -69.6248
 momentum : 0.207358 -0.0305505 -0.20052
 momentumAtEndpoint : 0 -0 -0
 spin : 0 0 0
 colorFlow : 0 0
 parents : c68708c168
 daughters :
```

```
JetOrigin           INFO  Jet ID: 5        Energy: 61.7908 : 58.3185
JetOrigin           INFO  Eta:   -0.751861: 0.0909377 Phi:   0.384387: 0.0581178 Mass:  28.5455
JetOrigin           INFO  Px:    39.1934 : 57.7834  Py:    15.854  : 3.36203
JetOrigin           INFO  Pz:    -34.8682: 5.27083  Pt:    42.2785 : 57.8811
JetOrigin           INFO  Cos(θ): -0.636258: 0.0906879
JetOrigin           INFO  DeltaE: 0.0595414 DeltaPx: -0.321719 DeltaPy: 3.71562  DeltaPz: -7.61532
JetOrigin           INFO  DeltaEta:-0.842799 DeltaPhi:0.326269 DeltaR: 0.903748
JetOrigin           INFO Jet Nparticles: 21       Matched Truth: 18
JetOrigin           INFO Jet Ymerge1: 0.566634 Jet Ymerge2: 0.0735972 Jet Ymerge3: 0.0288442
JetOrigin           INFO
JetOrigin           INFO 1: -2212 : 0.487163 : -0.352629 : -1.27141 : 1.69071
JetOrigin           INFO 1: 321 : 1.53739 : 0.0882302 : -0.996941 : 1.8997
JetOrigin           INFO 1: 211 : -0.24231 : 0.302205 : -0.174232 : 0.447078
JetOrigin           INFO 1: -211 : 0.0401332 : -0.26939 : -0.328194 : 0.448745
JetOrigin           INFO 1: 211 : 3.27854 : -0.223893 : -3.31824 : 4.67217
JetOrigin           INFO 1: -211 : 0.323064 : -0.165926 : -0.638264 : 0.747504
JetOrigin           INFO 1: -211 : 0.628152 : -0.174853 : -0.800565 : 1.04189
JetOrigin           INFO 1: -211 : 0.873888 : 0.681383 : -0.217063 : 1.13779
JetOrigin           INFO 1: 211 : 0.292688 : 0.163197 : -0.507021 : 0.623578
JetOrigin           INFO 1: -211 : 0.732942 : 0.0837761 : -0.456162 : 0.878513
JetOrigin           INFO 1: 211 : -0.329223 : -0.127718 : -0.00528008 : 0.379746
JetOrigin           INFO 1: -321 : 0.174421 : 0.0122814 : 0.563088 : 0.768949
JetOrigin           INFO 1: -211 : 0.383103 : -0.305514 : -0.660659 : 0.8343
JetOrigin           INFO 1: 211 : 0.207358 : -0.0305505 : -0.20052 : 0.321898
JetOrigin           INFO 1: 211 : 0.730673 : -0.379369 : -1.07901 : 1.36438
JetOrigin           INFO 1: -211 : 0.658582 : 0.841361 : -0.331626 : 1.12742
JetOrigin           INFO 0: 211 : 0.674515 : 0.651413 : 0.0740459 : 0.950932
JetOrigin           INFO 1: -2212 : 1.30374 : 1.44285 : -0.48827 : 2.21366
JetOrigin           INFO 0: 211 : 0.310803 : 0.204493 : -0.0155254 : 0.397664
JetOrigin           INFO 1: 2212 : 6.70552 : 5.50281 : -4.1895 : 9.6787
JetOrigin           INFO 1: -211 : 0.75547 : 1.03915 : -0.636641 : 1.44061
JetOrigin           INFO 1: 211 : 0.349126 : -0.0206927 : -0.15275 : 0.406361
JetOrigin           INFO 0: -211 : 0.133132 : -0.0104909 : 0.486927 : 0.523843
JetOrigin           INFO 0: -211 : 0.0187888 : 0.168162 : 0.774489 : 0.80495
JetOrigin           INFO 1: 211 : -0.285032 : 0.0804629 : 0.304115 : 0.446859
JetOrigin           INFO 1: -321 : 0.145107 : -3.69627 : 9.07354 : 9.81103
JetOrigin           INFO 0: 211 : 0.0201625 : 0.00335537 : 0.285279 : 0.318248
JetOrigin           INFO 1: 211 : -2.55145 : 2.10102 : -2.2907 : 4.02381
JetOrigin           INFO 1: -211 : -0.319055 : -0.169639 : 1.18457 : 1.2463
JetOrigin           INFO 0: 211 : -0.0303393 : 0.0552844 : 0.110099 : 0.188623
JetOrigin           INFO 1: -11 : 0.0257235 : 0.4499 : 0.182313 : 0.486117
JetOrigin           INFO 1: -321 : 0.543428 : -0.204375 : 1.92704 : 2.07225
JetOrigin           INFO 1: 321 : -11.7738 : 4.9272 : -9.10983 : 15.6886
JetOrigin           INFO 1: -211 : -0.94387 : 0.426599 : -1.00987 : 1.45334
```

Input variable distribution

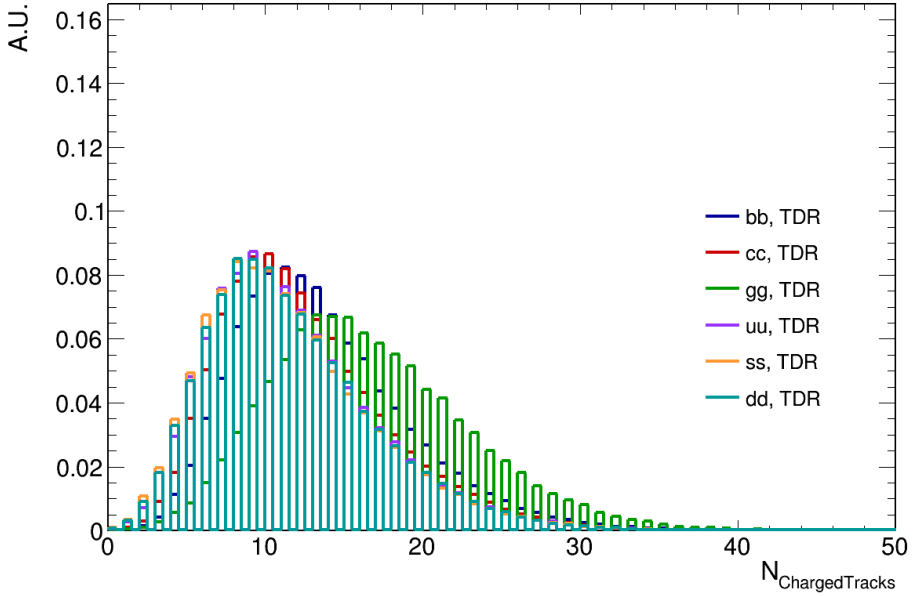




TDR: /cefs/higgs/zhangkl/JOI/datasets0224
CDR: /cefs/higgs/zhangkl/AI/datasets/zhuyf_higgs125
Full comparison will be ready in Wednesday.

N_charged track for TDR.
So now choose length 40.

```
bbTDR: 13.3142 ±5.46265
ccTDR: 12.1757 ±5.33673
ggTDR: 16.3898 ±6.29841
uuTDR: 11.4786 ±5.38694
ssTDR: 11.241 ±5.41545
ddTDR: 11.3787 ±5.3928
```



Scaling Law

- Trying use new small size sample for training/testing. 420000 : 50000

```
[2025-02-24 02:35:50,058] INFO: Processed 420000 entries in total (avg. speed 634.2 entries/s)
[2025-02-24 02:35:50,058] INFO: Train AvgLoss: 0.57868, AvgAcc: 0.79346
[2025-02-24 02:35:50,059] INFO: Train class distribution:
    [(0, 38122), (1, 38131), (2, 38500), (3, 38496), (4, 38507), (5, 38489), (6, 38312), (7, 38304), (8, 37711), (9, 37819), (10, 37609)]
```

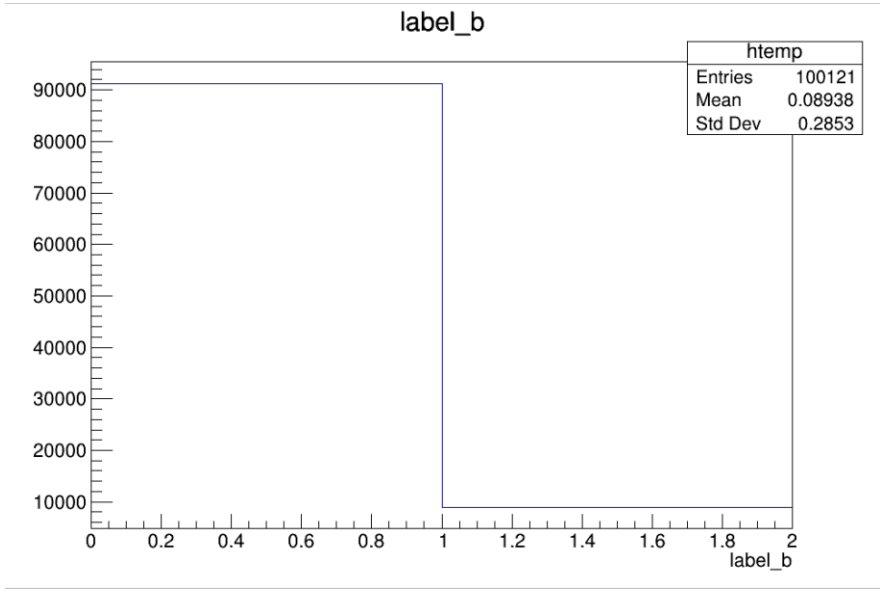
```
[2025-02-24 02:47:02,547] INFO: Processed 50000 entries in total (avg. speed 2342.1 entries/s)
[2025-02-24 02:47:02,547] INFO: Evaluation class distribution:
    [(0, 4452), (1, 4458), (2, 4447), (3, 4472), (4, 4357), (5, 4360), (6, 4451), (7, 4470), (8, 4398), (9, 4410), (10, 5725)]
```

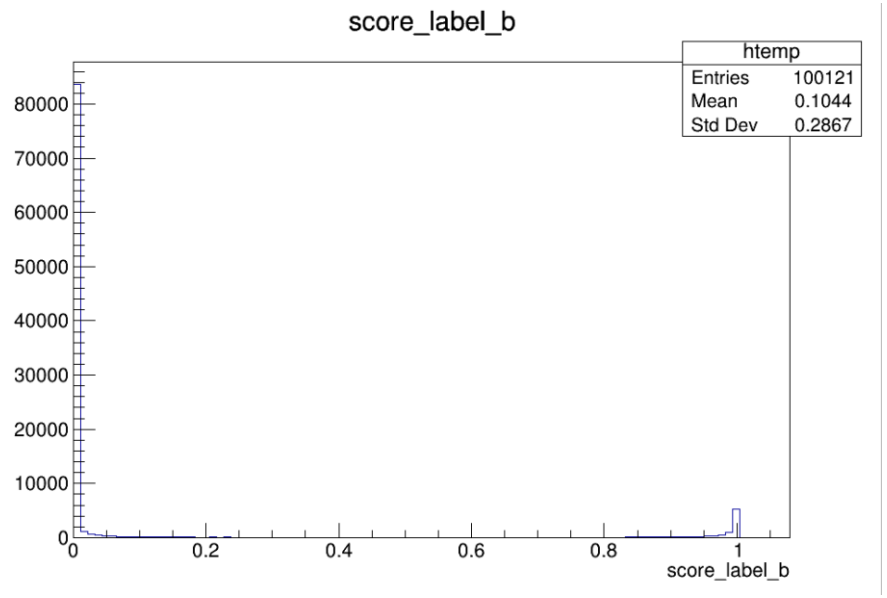
- 1/20 size of full set. Giving b-eff ~0.6 (best 0.7)
  - Even though small amount size giving enough result.
  - Overtraining issue is studying with small size sample for rapid response. (~10mins for 1 epoch).

Application

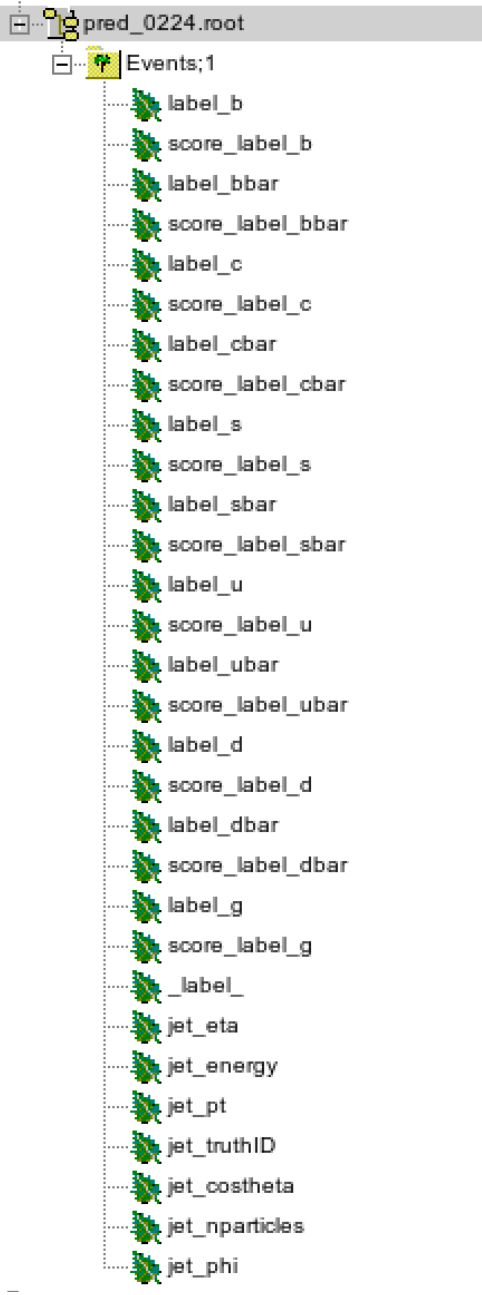
/cefs/higgs/zhangkl/AI/JOI/higgs/pred_0224.root

- For test sample 100k, category and score given.





- With onnx model, for each candidate, score given.
- B tagging score can be : score_b/(score_b+score_c+……11)



backup

Machine Learnings on Jets

- P-CNN
  - https://scipost.org/10.21468/SciPostPhys.7.1.014
- Particle Flow Network
  - https://arxiv.org/abs/1810.05165
  - CEPC@Xiaotian : https://arxiv.org/abs/2410.04465v2
- LundNet
  - https://doi.org/10.1007/jhep03(2021)052
- ParticleNet
  - Arxiv:1902.08570
  - https://github.com/hqucms/ParticleNet

ParticleTransformer

- https://arxiv.org/abs/2202.03772
- https://github.com/jet-universe/particle_transformer
- Platforms: https://github.com/hqucms/weaver-core
- Application on CEPC: 2309.13231 , PRL 132, 221802 (2024)
- Tutorial on CEPC: https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification
- Inputs from CEPCsoft: /cefs/higgs/zhangkl/AI/datasets
- Inputs from LHC, JetClass: /cefs/higgs/zhangkl/AI/jetclass
- Require higgsgpu group. Request on https://ccsinfo.ihep.ac.cn/
- Follow the tutorial, build the env if you are interested.

ParticleTransformer @ CEPC

https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification

- Variable list in M11origin.cc
  - Under development to CEPCSW
  - Unit as one jet: 4 momentum, M11 id information……
- Train in Weaver: JetClass_full.yaml
- Submit jobs on IHEP: train_JetClass.sh
- Output: Pred.root: Label and score for each jets.
- Application: onnx format

Inputs for JOI

/cefs/higgs/zhangkl/CEPCSW/Analysis/JetOrigin/src

- Jet->Event;
- PFO->Component;
- Length: 200
- Label: M11

- Current training use truth PID information, in application reco PID will be used.

Type	Var	Comment
PFO point distance	$\Delta\phi(pfo, Jet)$	Delta Phi, pfo to jet
	$\Delta\eta(pfo, Jet)$	Delta Eta, pfo to jet
PFO Vector variable	(px, py, pz, E)	4 momentum of PFO
PFO feature variable	P_t^{PFO}, $\log\frac{P_t^{PFO}}{P_t^{jet}}$	Pfo pt and relative pt
	E_t^{PFO}, $\log\frac{E_t^{PFO}}{E_t^{jet}}$	Pfo E and relative E
	$\Delta R(pfo, Jet)$	Delta R, pfo to jet
	N_charge, N_chargeflip	Charge of PFO
	D0, Z0, D0err, Z0err	(if with track) impact parameters
	N_Ecluster, N_Hcluster	
	E_ecal, E_hcal	
	PID	Truth PID type

Variable convention

- Feature variable, Transformer prefer normal distribution with mean ~0, range (-1,1) with cut edge maximum(-5,5).
- (4-momentum vector variable not included)
- Normalization functions like Tanh() used.
-