CEPC Jet@Clusters

Kaili Zhang

IHEP

zhangkl@ihep.ac.cn

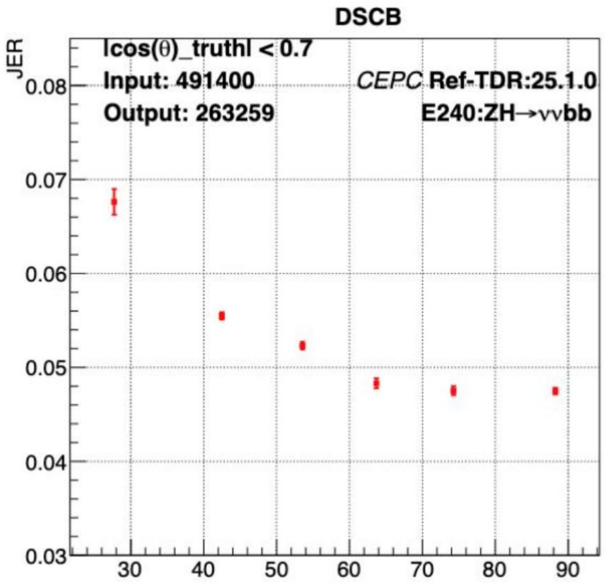
JER/JES

@Yingqi

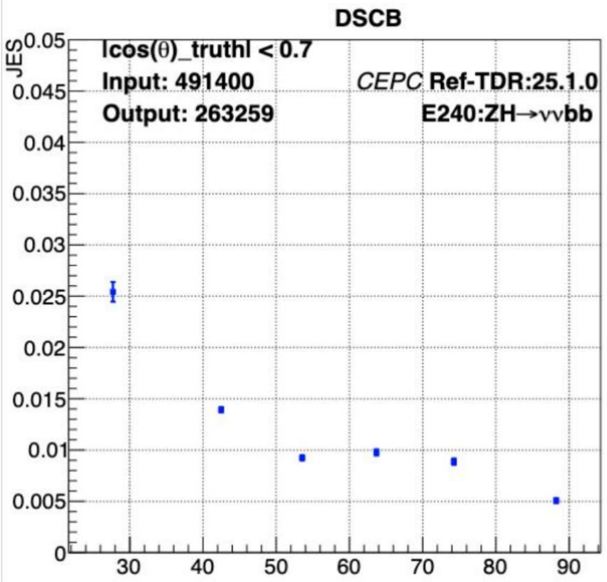
DSCB and RMS90 method tested in latest release, given different results.
Under further tuning.

DSCB

JER

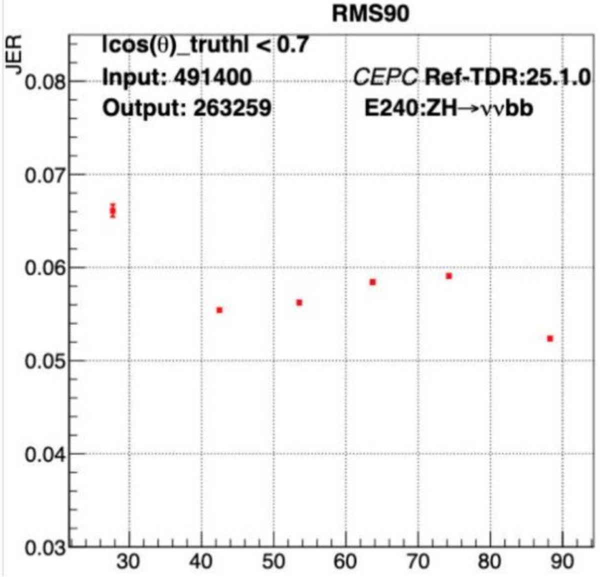


JES

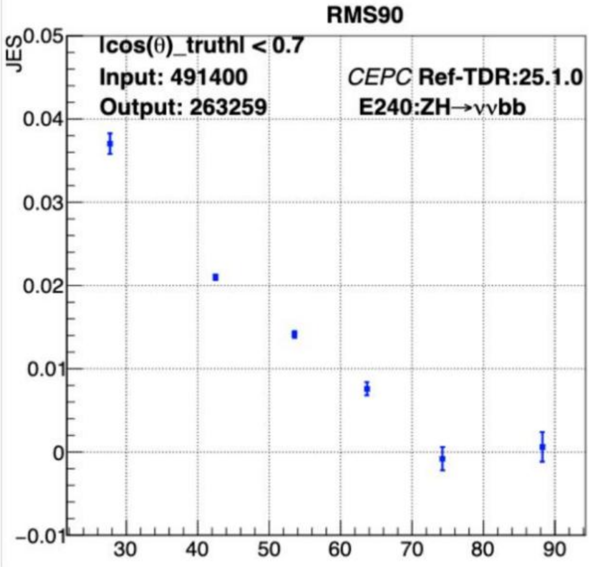


RMS90

JER



JES

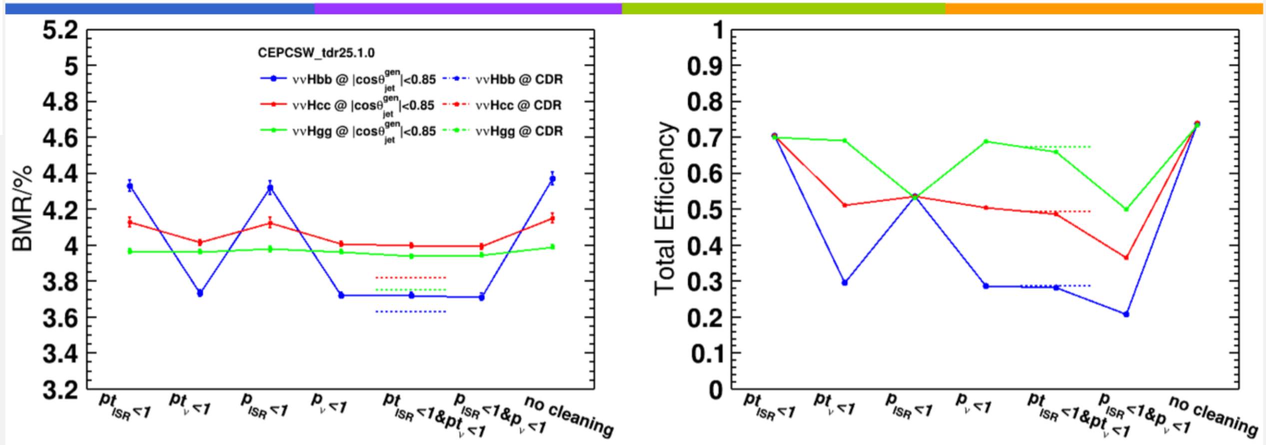


BMR Summary

@Xiaotian

Case	process	$ZH \to \nu\nu gg$	$ZH \to \nu\nu bb$	$ZH \to \nu\nu cc$	$ZH \to \nu\nu uu$	$ZH \to \nu\nu dd$	$ZH \to \nu\nu ss$
Physical level	BMR/%	4.00 ± 0.01	4.36 ± 0.03	4.16 ± 0.03	3.79 ± 0.01	3.97 ± 0.01	4.44 ± 0.01
	Efficiency/%	73.3	73.7	74.0	74.2	74.1	74.1
Detector level	BMR/%	3.95 ± 0.01	3.74 ± 0.02	4.01 ± 0.01	3.77 ± 0.01	3.95 ± 0.01	4.40 ± 0.01
	Efficiency/%	65.7	28.1	48.6	70.3	70.1	70.2

- Event cleaning: $\Sigma|Pt_{\mathrm{ISR}}| < 1\mathrm{GeV}/c \& \Sigma|Pt_{\nu}| < 1\mathrm{GeV}/c$
- Before event cleaning, BMR ranges from 3.79% to 4.44%
- After event cleaning, BMR ranges from 3.74% to 4.40%



Current result – efficiency consistent with CDR, BMR 0.11%/0.19%/0.20% higher

Table 1. Event cumulative efficiency for Higgs boson exclusive decay at the CEPC with $\sqrt{s} = 240$ GeV.

	gg(%)	bb(%)	cc(%)	WW^*(%)	ZZ^* (%)
Pt_ISR < 1 GeV	95.15	95.37	95.30	95.16	95.24
Pt_neutrino < 1 GeV	89.33	39.04	66.36	37.46	41.39
|Cos(Theta_Jet)| < 0.85	67.30	28.65	49.31	–	–

Table 3. Higgs boson mass resolution (sigma/Mean) for different decay modes with jets as final state particles, after event cleaning.

$H \to bb$	$H \to cc$	$H \to gg$	$H \to WW^*$	$H \to ZZ^*$
3.63%	3.82%	3.75%	3.81%	3.74%

CDR reference

PFN quick look

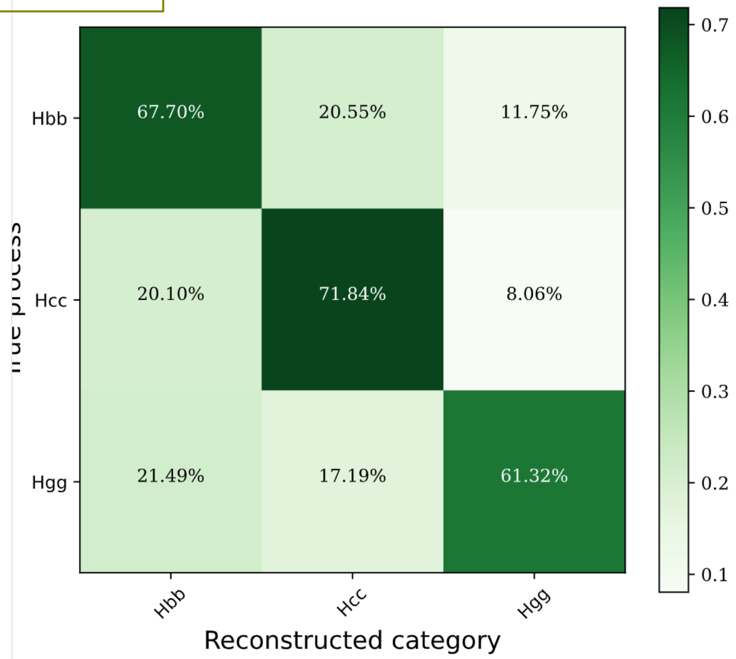
@Xiaotian Ma, Zuofei Wu

Samples：
- 150k for each category, (training: validation: test sets) = (8:1:1)
- Signal: $H \rightarrow b\bar{b}$、$H \rightarrow c\bar{c}$、$H \rightarrow gg$ (CEPCSW_tdr25.1.0)
- No event selection applied.

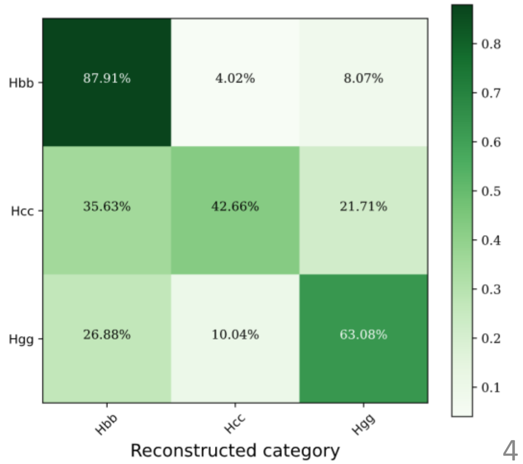
Training variables：
- Energy, momentum, $\cos\theta$, ϕ, D_0, Z_0,(w/o PID)

Training parameters：
- Φ_sizes: (64, 64, 50), F_sizes: (64, 64, 40)
- Fully connected layer: ReLU activation function and adam optimizer
- Output layer: SoftMax activation function
- loss function: cross-entropy
- Epoch: 50, Learning rate: 0.001, Batchsize: 1000

With PID info



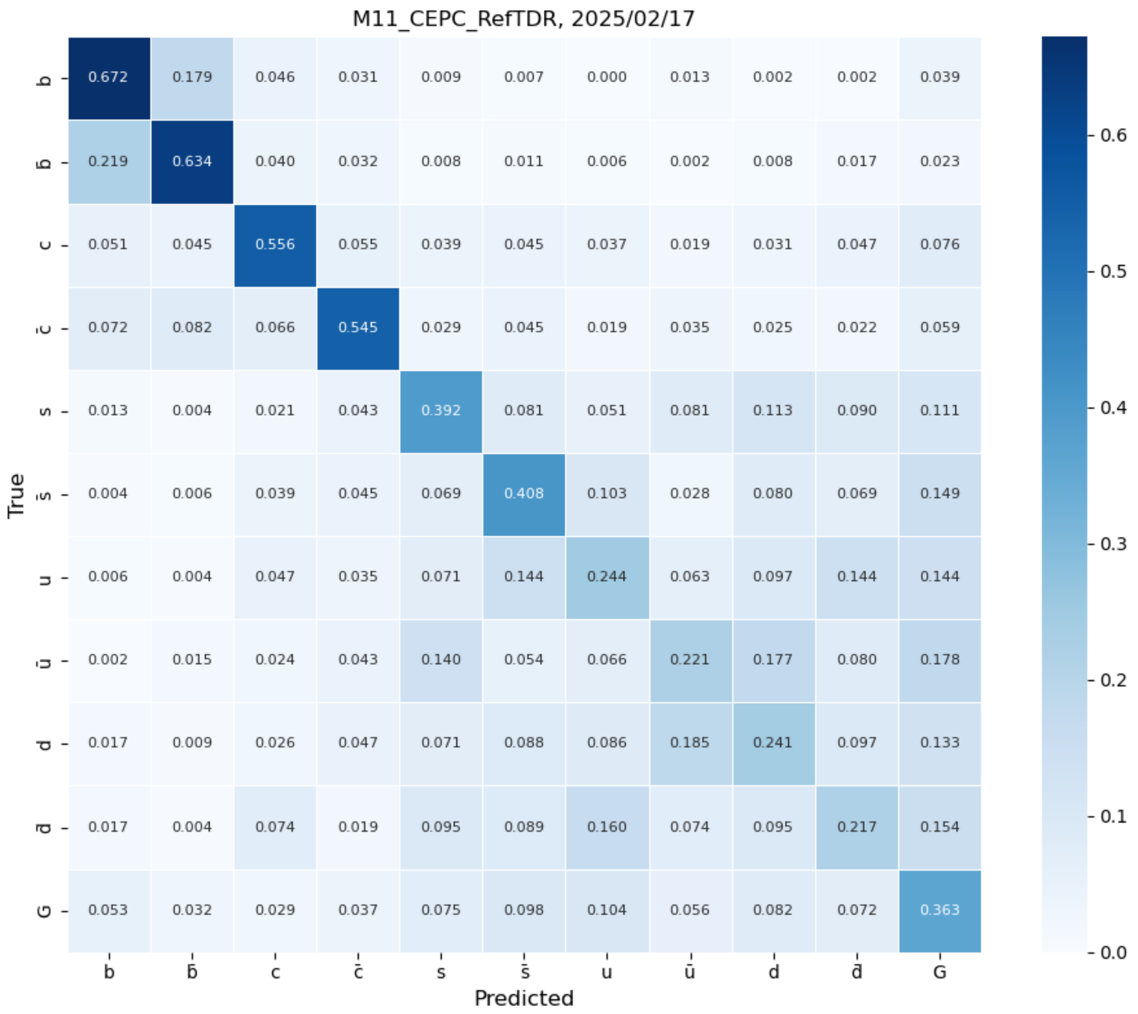
Without PID info

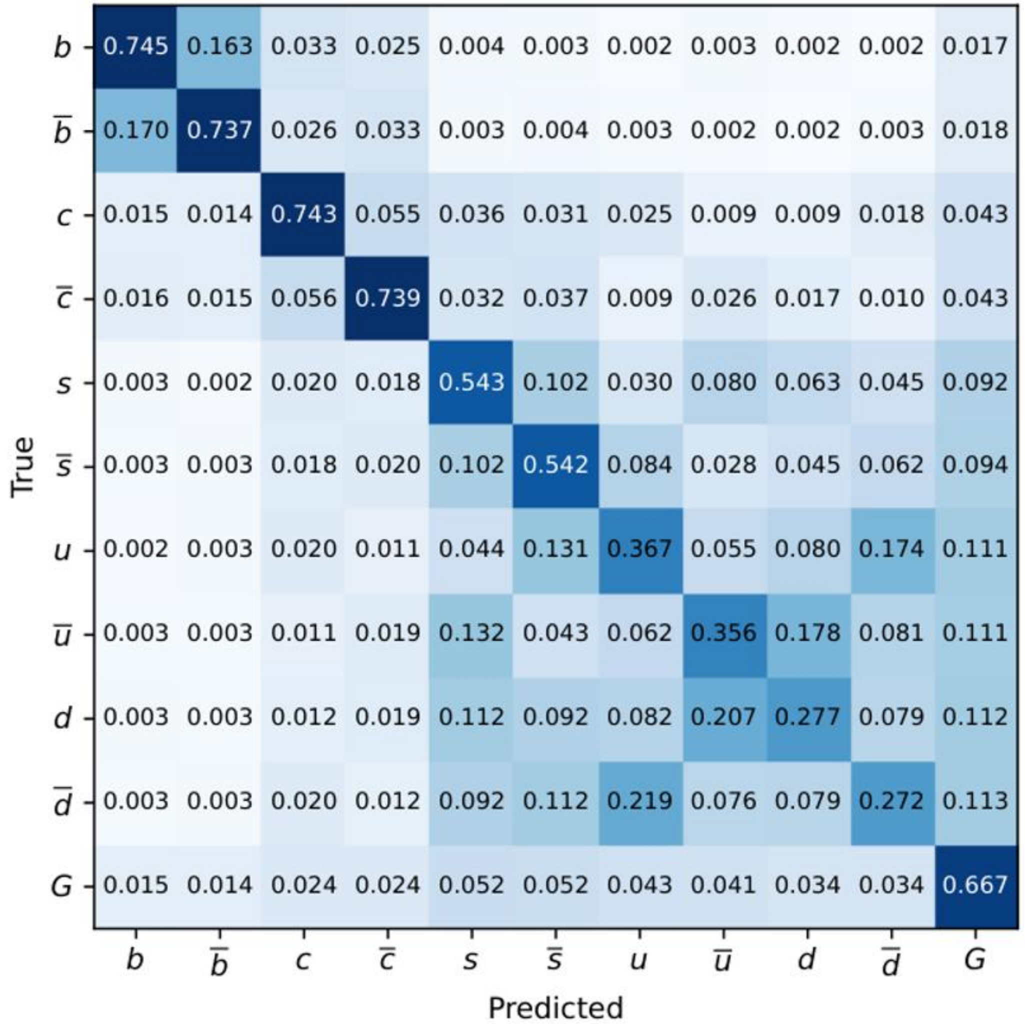


Current JOI

Discrepancy should be in training parameters. Under tuning.

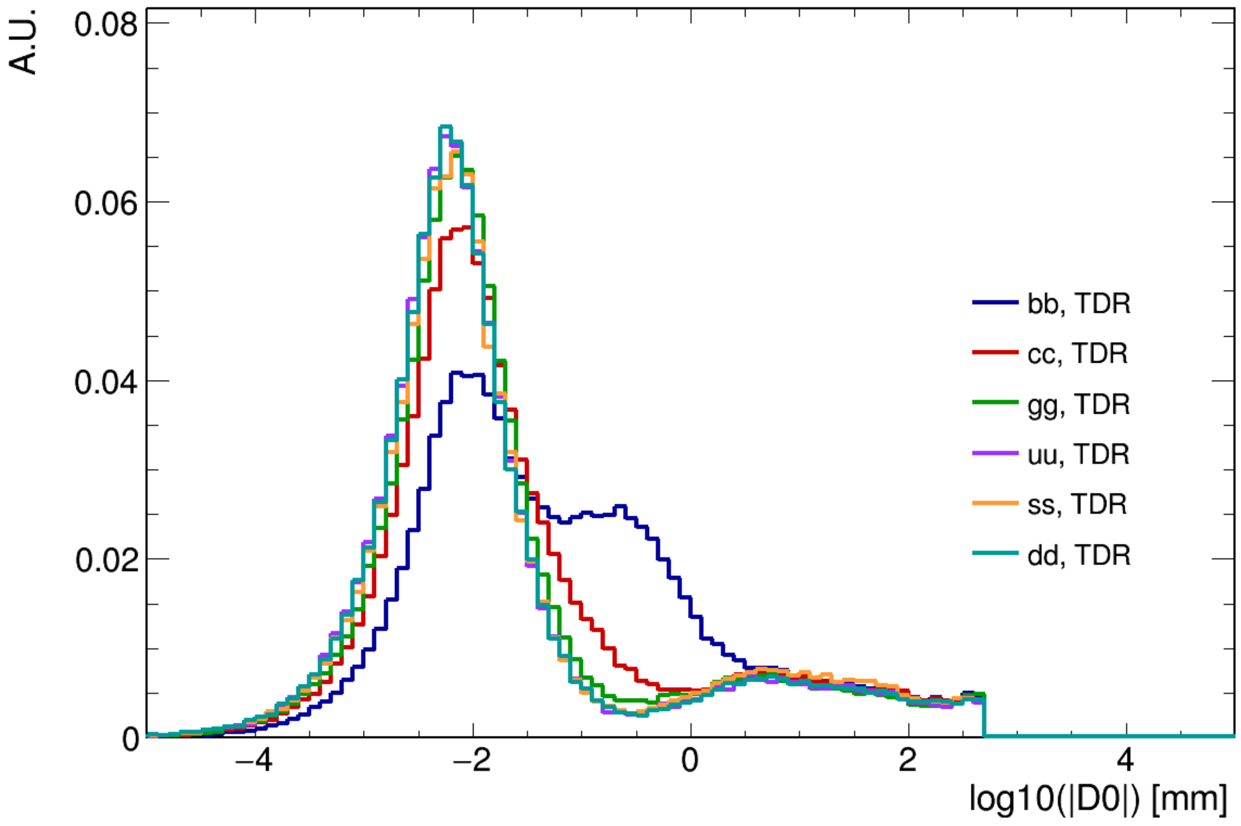
2309.13231

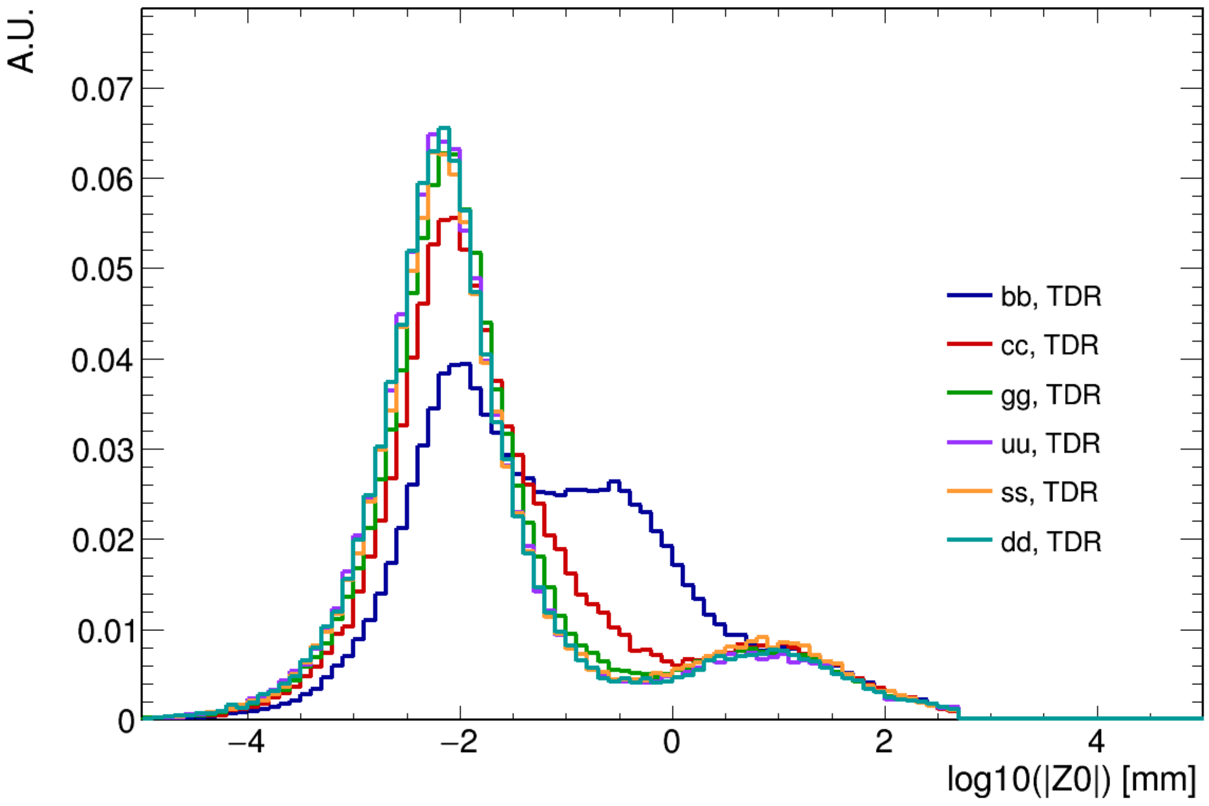




JOI inputs

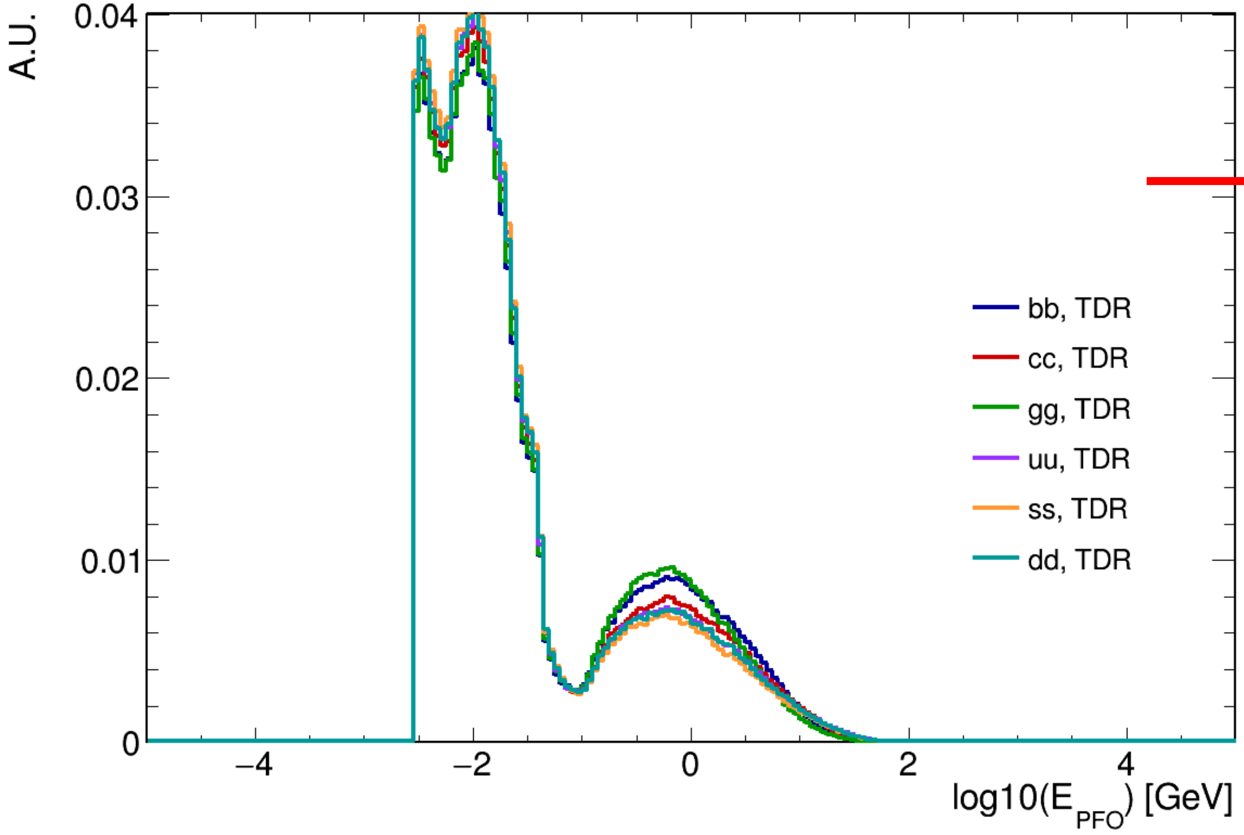
In D0 and Z0 plot, 3 pattern can be seen:
IP(Primary Vertex), Secondary/Thirdary Vertex(From b decay. Length~100um.)
and Long-Live Decay Vertex(From Kshort, Lambda……, length~cm.)





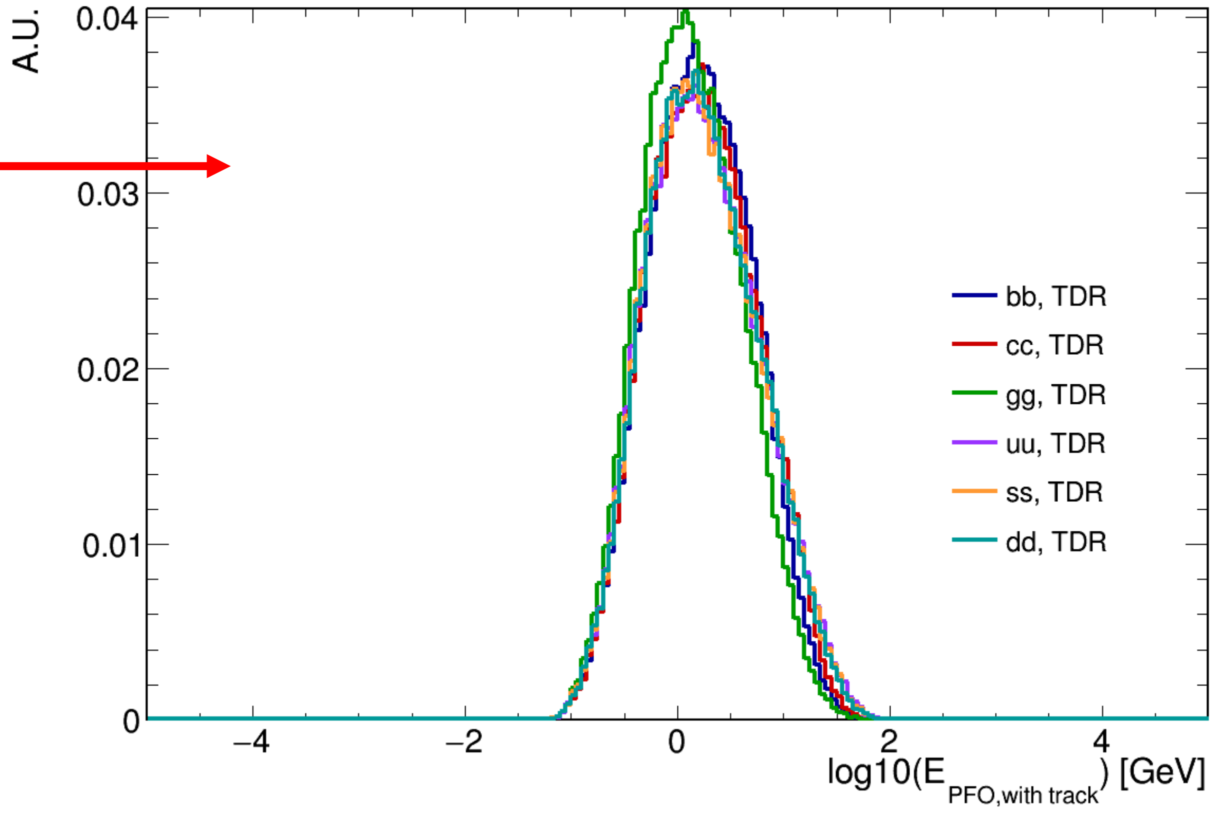
PFO Energy

```
bbTDR: -1.48234 ±0.949577
ccTDR: -1.53376 ±0.926584
ggTDR: -1.50001 ±0.923936
uuTDR: -1.55966 ±0.91385
ssTDR: -1.58785 ±0.891898
ddTDR: -1.56168 ±0.910634
```



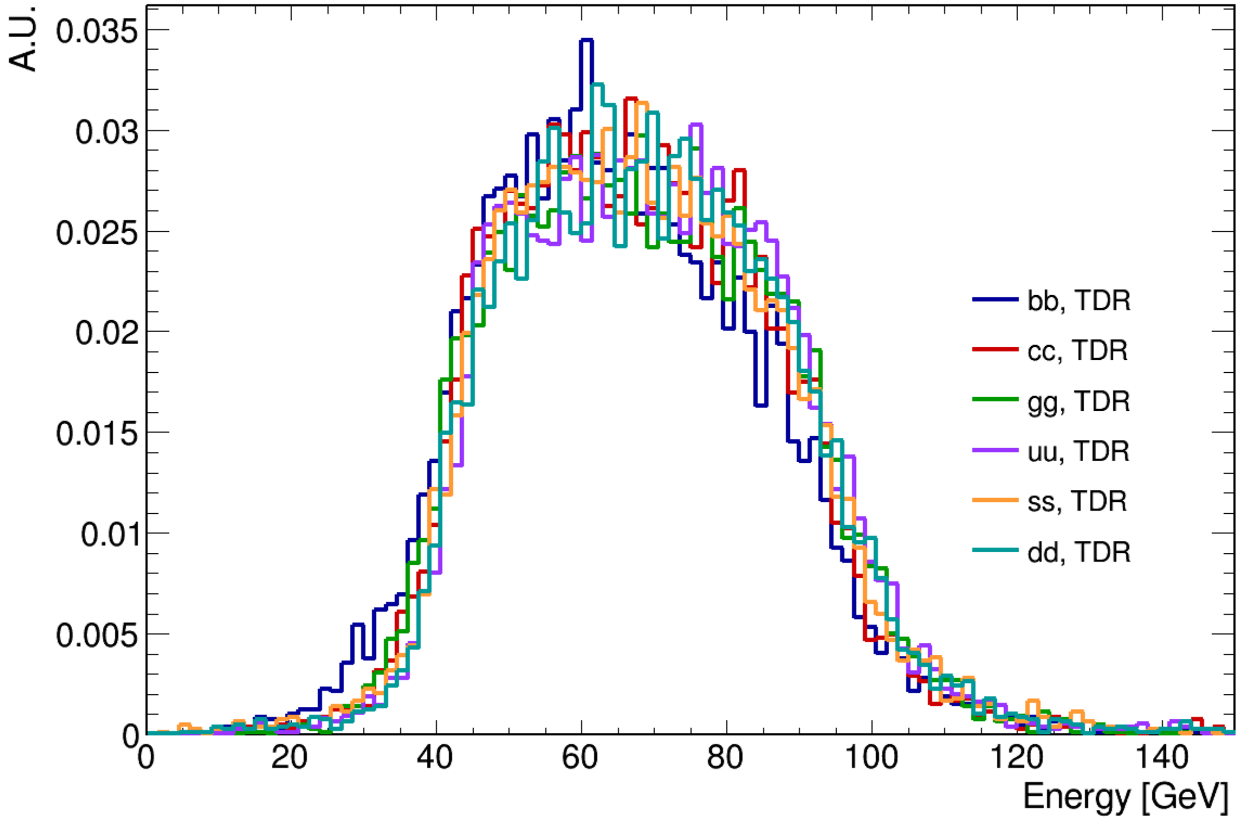
10%

```
bbTDR: 0.230021 ±0.483184
ccTDR: 0.245149 ±0.505041
ggTDR: 0.14169 ±0.472343
uuTDR: 0.245575 ±0.528271
ssTDR: 0.242318 ±0.524427
ddTDR: 0.24227 ±0.520791
```

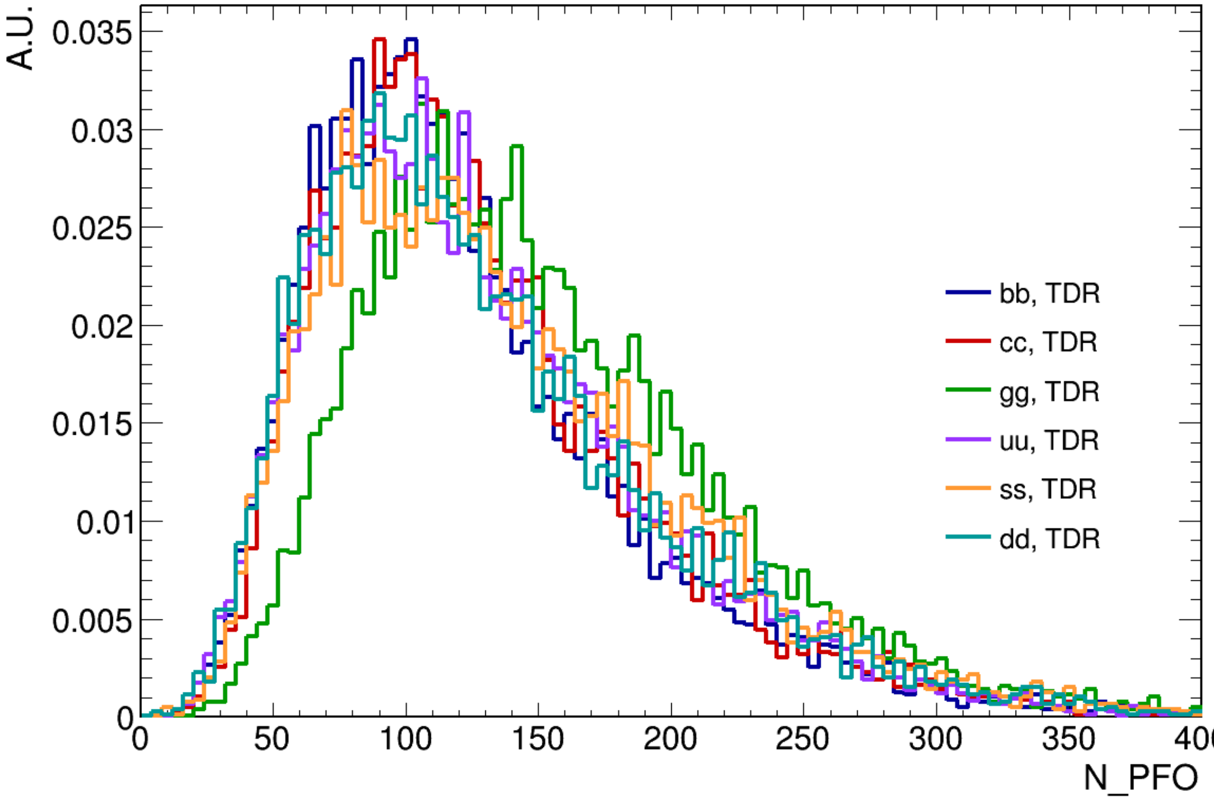


Jet Energy and Jet N_PFO

```
bbTDR: 65.3911 ±18.7989
ccTDR: 67.7635 ±18.5029
ggTDR: 68.5757 ±18.8567
uuTDR: 69.847 ±18.7273
ssTDR: 68.6575 ±19.0577
ddTDR: 69.5257 ±18.7775
```

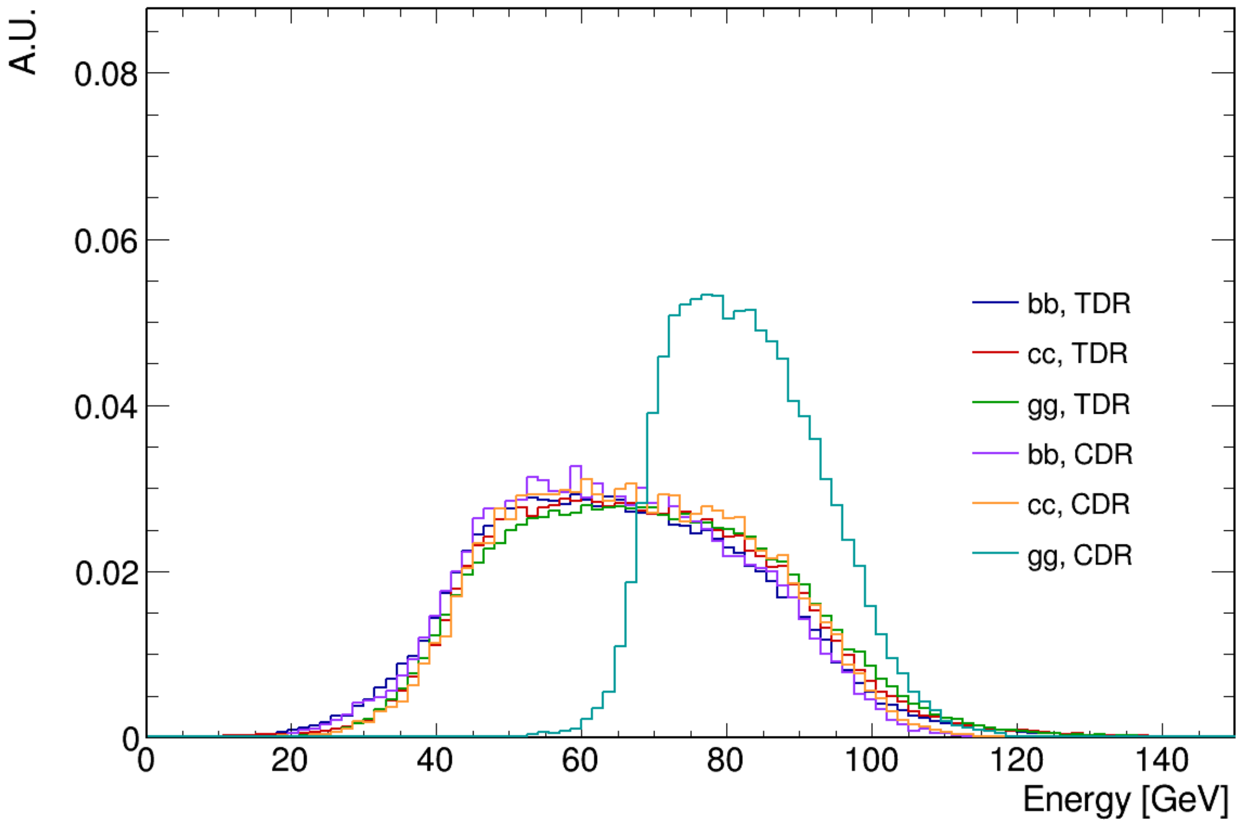



```
bbTDR: 123.738 ±60.9093
ccTDR: 128.538 ±61.7292
ggTDR: 150.318 ±64.4921
uuTDR: 127.983 ±62.172
ssTDR: 135.518 ±65.9203
ddTDR: 129.322 ±65.0416
```



Jet Energy @CDR

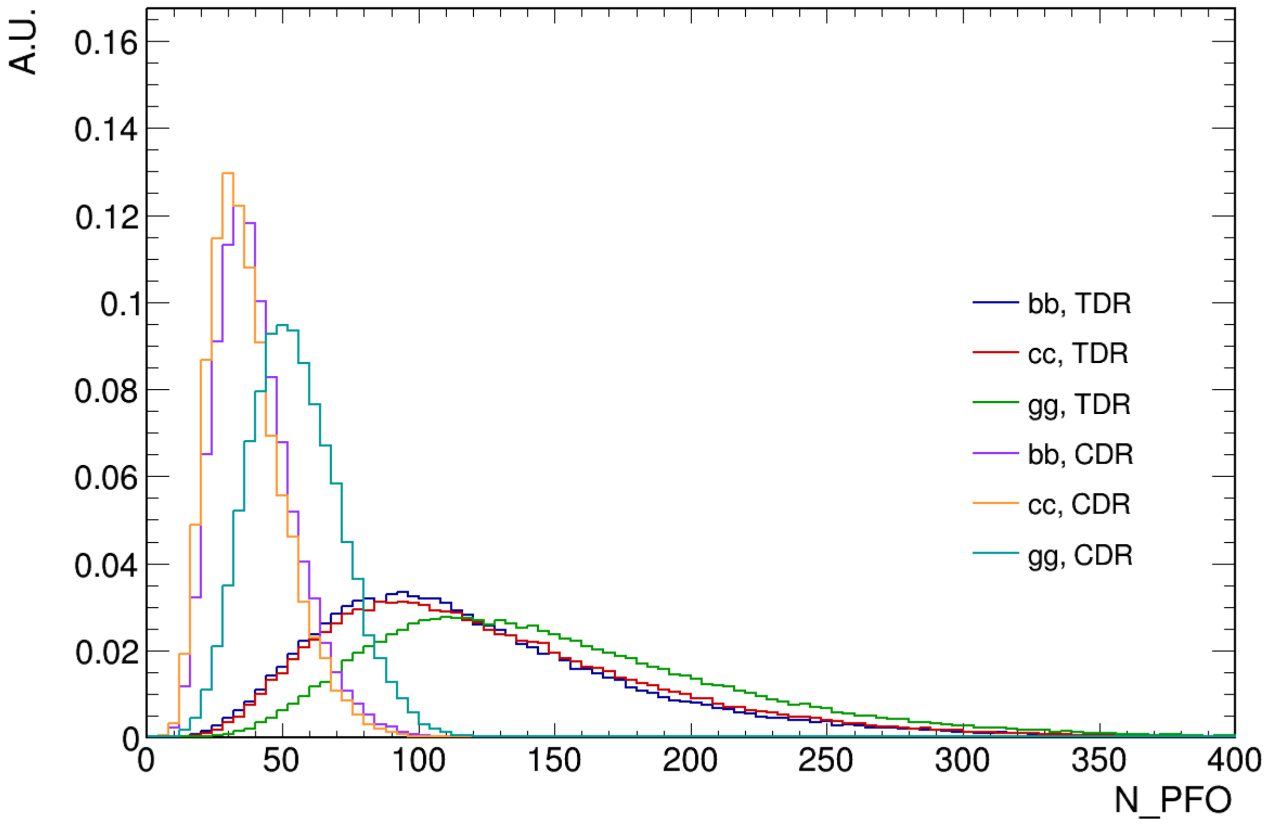
In average, each jet ~66GeV Energy.




```
bbTDR:  65.4526 ±18.8213
ccTDR:  67.7845 ±18.6604
ggTDR:  68.6878 ±18.7631
bbCDR:  64.4942 ±17.0913
ccCDR:  67.0562 ±16.8343
ggCDR:  82.6409 ±10.3267
```

In CDR JOI, one gg event has 2 entries,
choose the leading jet in training. (82GeV) (biased?)

Jet N_PFOs @CDR

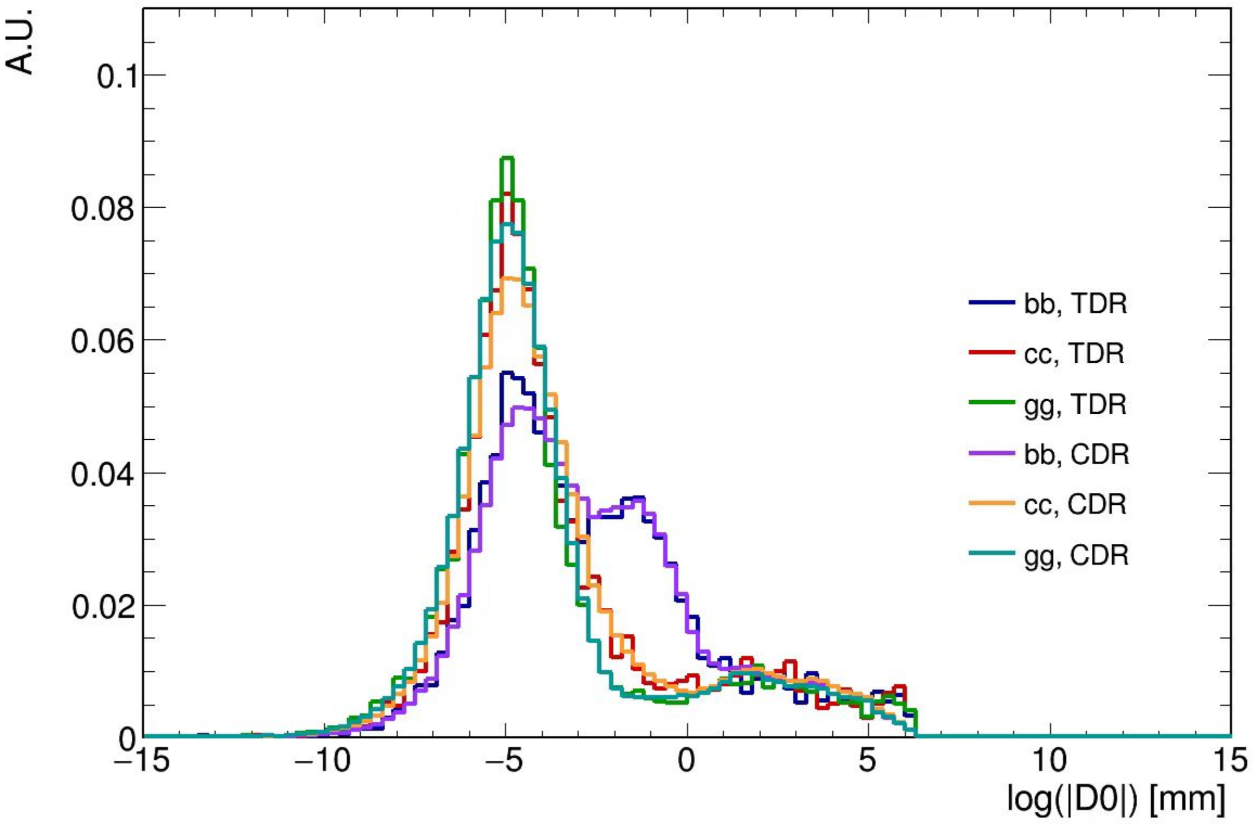


TDR N_PFO with more (broken, nertual)PFOs
TDR cut at 200 (with minimum energy entry ~mev level)
If N_PFO<200, fill with zero.
CDR cut at 128.

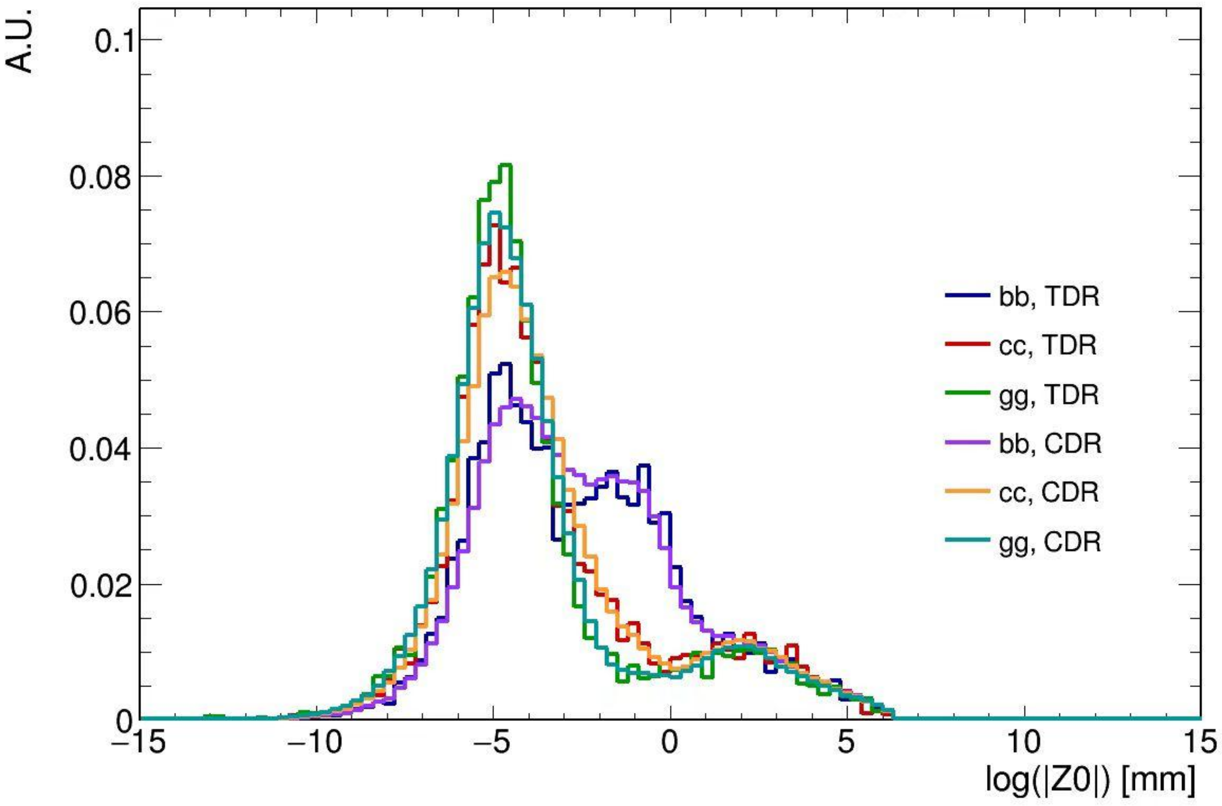
```
bbTDR: 122.982 ±59.9375
ccTDR: 127.902 ±61.5878
ggTDR: 150.687 ±65.2885
bbCDR: 39.7391 ±14.6149
ccCDR: 36.9408 ±14.0868
ggCDR: 54.4543 ±16.8207
```

Among these PFOs, ~10-20 PFOs are charged with tracks.

D0, Z0 @CDR



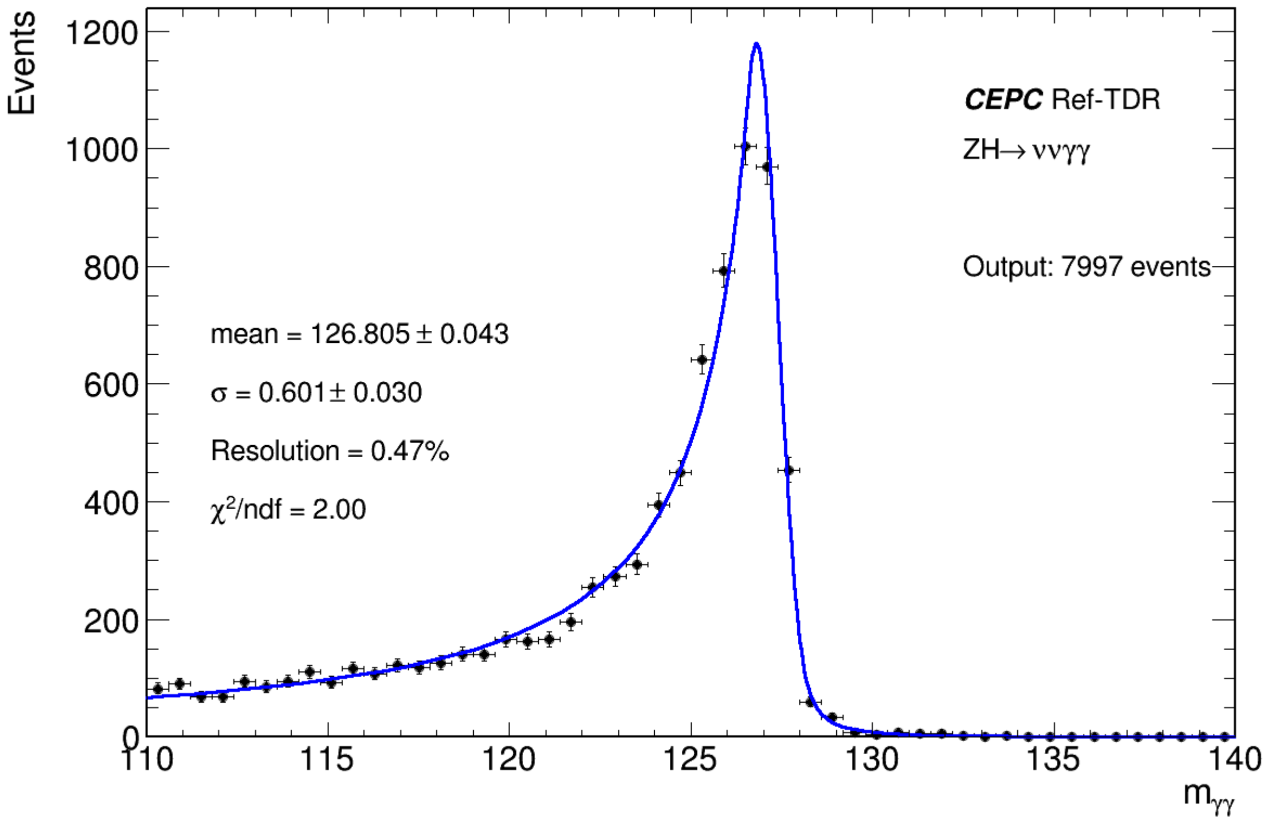
A.U.
0.1
0.08
0.06
0.04
0.02
0
−15
−10
−5
0
5
10
15
log(|D0|) [mm]
bb, TDR
cc, TDR
gg, TDR
bb, CDR
cc, CDR
gg, CDR

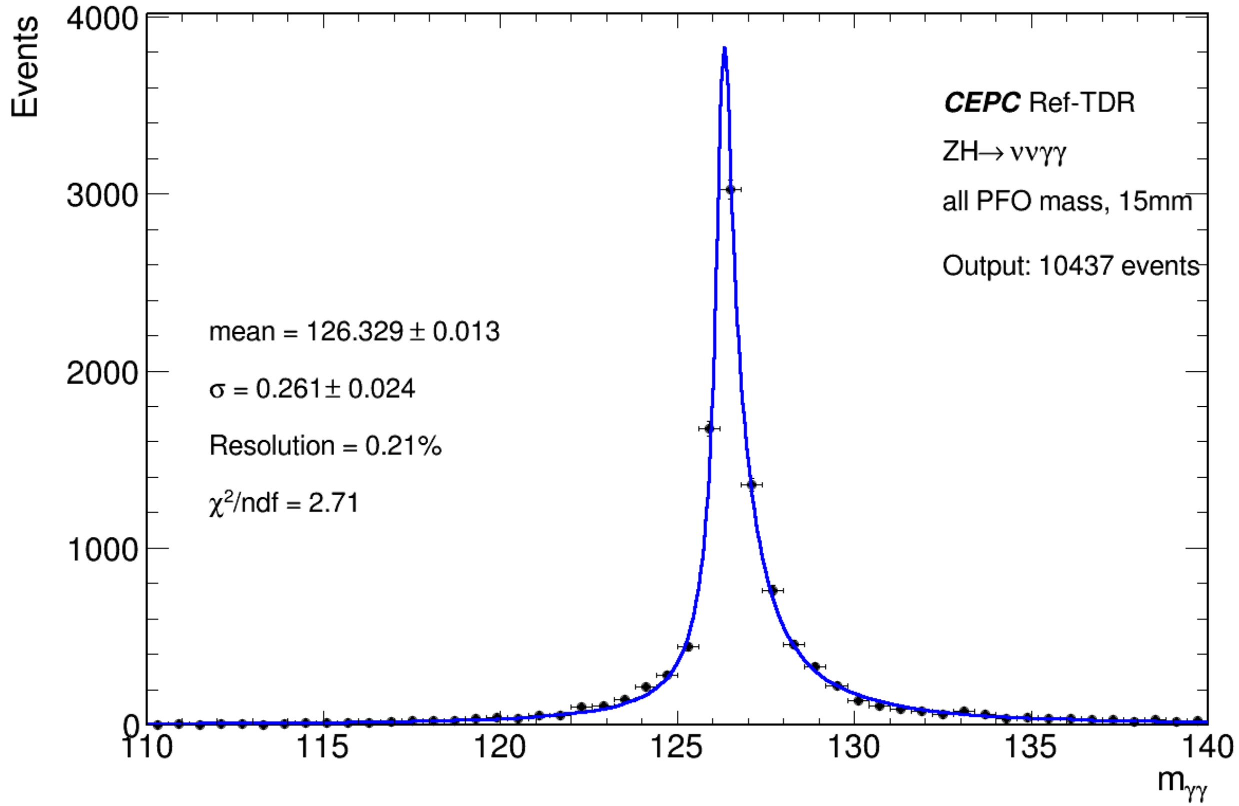


A.U.
0.1
0.08
0.06
0.04
0.02
0
−15
−10
−5
0
5
10
15
log(|Z0|) [mm]
bb, TDR
cc, TDR
gg, TDR
bb, CDR
cc, CDR
gg, CDR

Photon reco

In past H->yy, use leading PFO as the photon. With resolution ~0.47%.
But if use all PFO to reconstruct higgs in vvyy: 0.21% resolution.
->Photon p resolution ~0.1%. Comparable with muon tracks.
->Large improvements can be done in photon reco.





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	$\mathrm{P}_t^{PFO}, \log\frac{P_t^{PFO}}{P_t^{jet}}$	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.
-