Quick look on Tau tagger

Jets, samples and Wednesday working meeting

Kaili Zhang

zhangkl@ihep.ac.cn

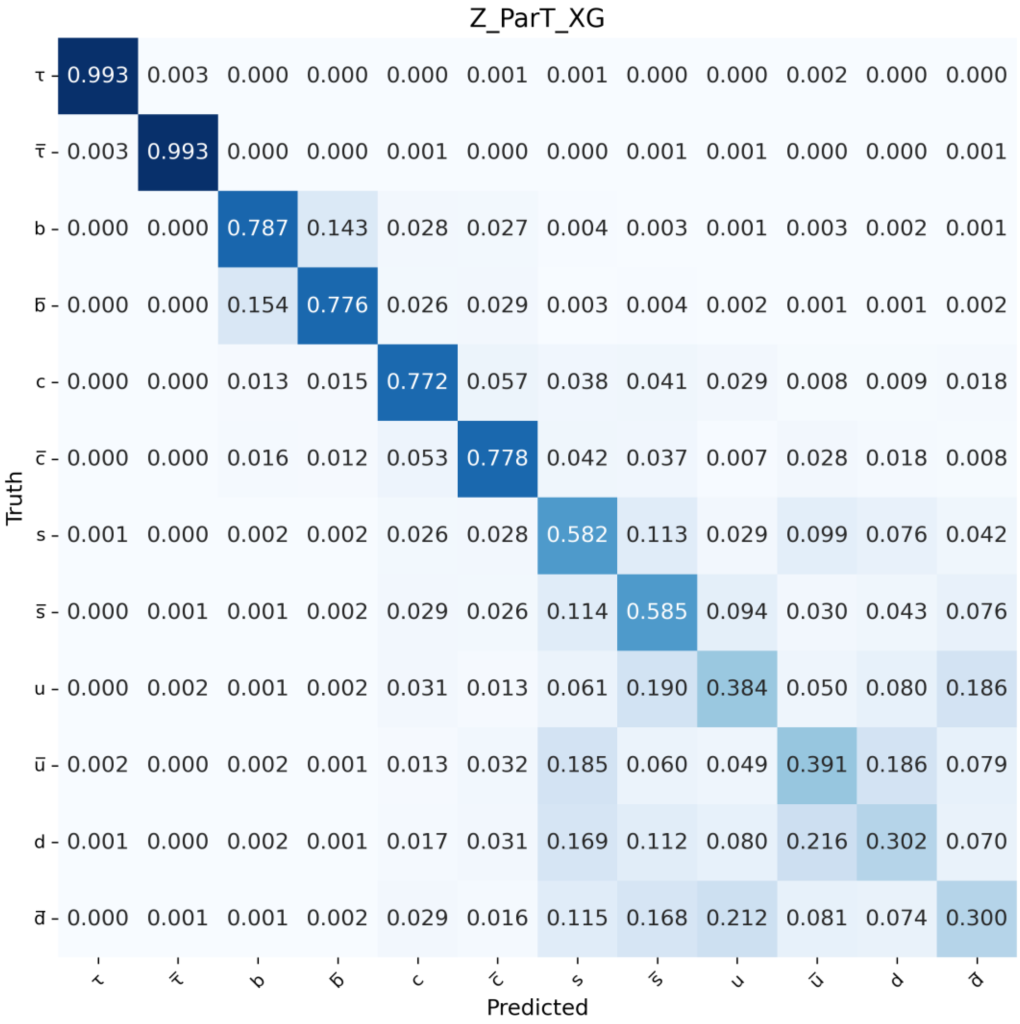
- Tag tau in jets.   Candidates: ee-kt algorithum jets.

- Environment: Zpole E91.2, tautau/bb/cc/uu/dd/ss, each 100k in training.

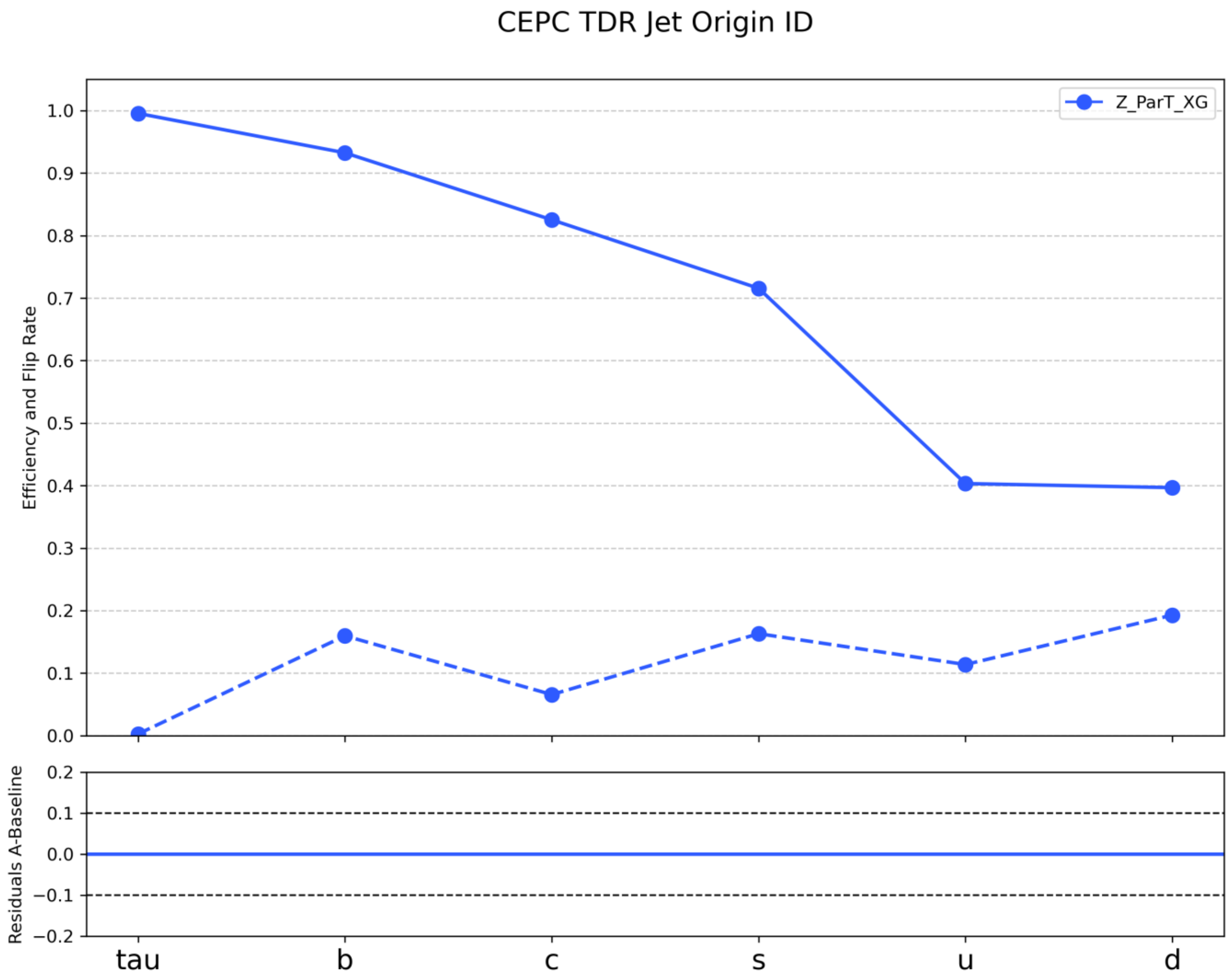
- Tau can be hardonic and leptonic.

- Trained in V100 4 cards, 10 hours.

M12 Matrix


Z_ParT_XG
τ - 0.993 0.003 0.000 0.000 0.000 0.001 0.001 0.000 0.000 0.002 0.000 0.000
τ̄ - 0.003 0.993 0.000 0.000 0.001 0.000 0.000 0.001 0.001 0.000 0.000 0.001
b - 0.000 0.000 0.787 0.143 0.028 0.027 0.004 0.003 0.001 0.003 0.002 0.001
b̄ - 0.000 0.000 0.154 0.776 0.026 0.029 0.003 0.004 0.002 0.001 0.001 0.002
c - 0.000 0.000 0.013 0.015 0.772 0.057 0.038 0.041 0.029 0.008 0.009 0.018
c̄ - 0.000 0.000 0.016 0.012 0.053 0.778 0.042 0.037 0.007 0.028 0.018 0.008
s - 0.001 0.000 0.002 0.002 0.026 0.028 0.582 0.113 0.029 0.099 0.076 0.042
s̄ - 0.000 0.001 0.001 0.002 0.029 0.026 0.114 0.585 0.094 0.030 0.043 0.076
u - 0.000 0.002 0.001 0.002 0.031 0.013 0.061 0.190 0.384 0.050 0.080 0.186
ū - 0.002 0.000 0.002 0.001 0.013 0.032 0.185 0.060 0.049 0.391 0.186 0.079
d - 0.001 0.000 0.002 0.001 0.017 0.031 0.169 0.112 0.080 0.216 0.302 0.070
d̄ - 0.000 0.001 0.001 0.002 0.029 0.016 0.115 0.168 0.212 0.081 0.074 0.300
Truth
τ τ̄ b b̄ c c̄ s s̄ u ū d d̄
Predicted

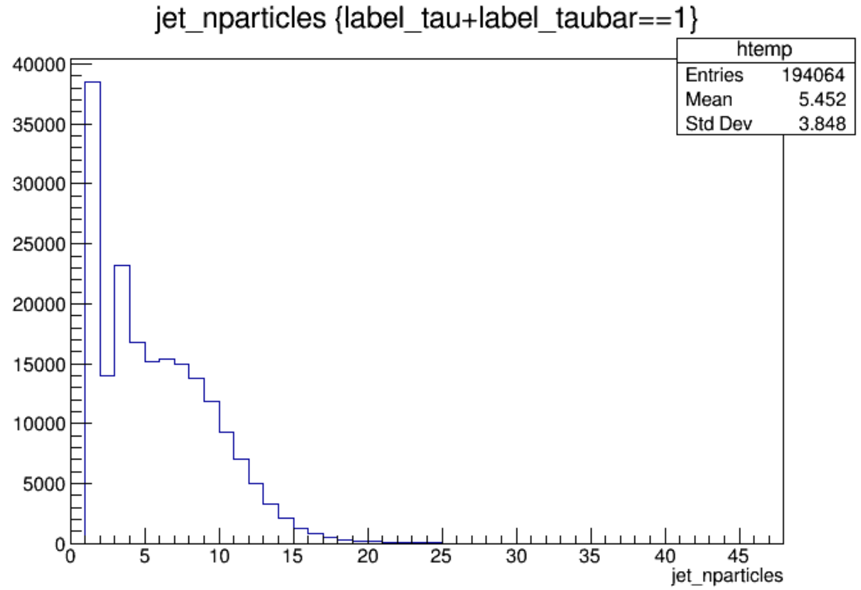
Performance

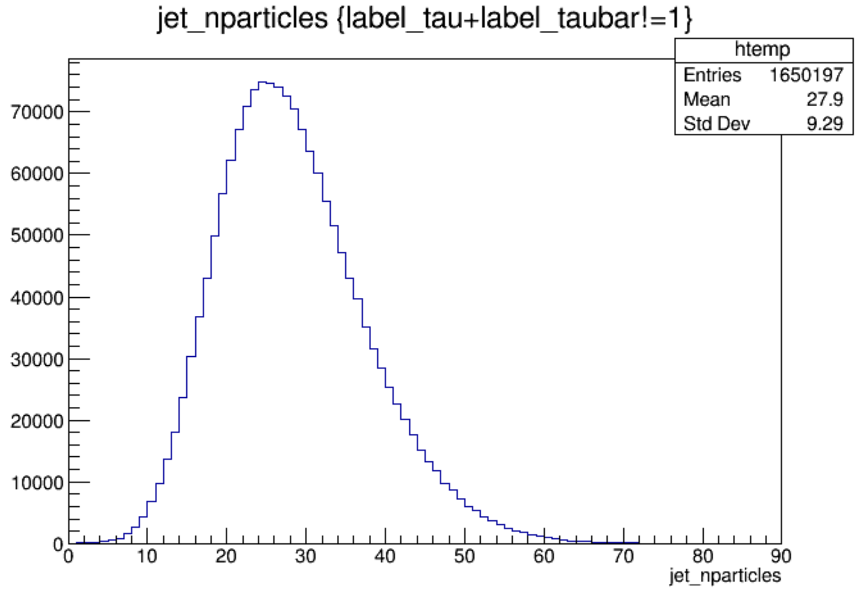


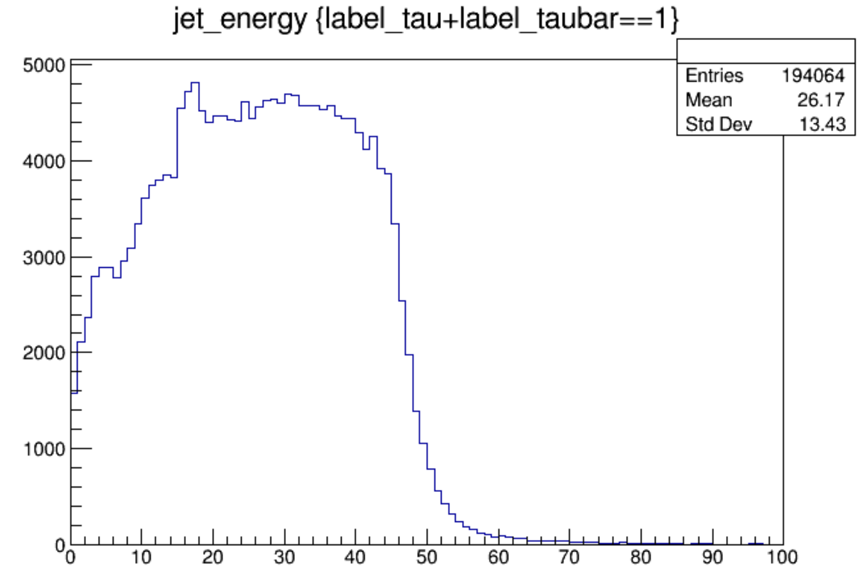
- Tau tagging efficacy ~99.6%
- Tau charge flip rate <1%.
- No need for leptonic and hardonic – enough power.
  - In principle should first remove isolated lepton first…

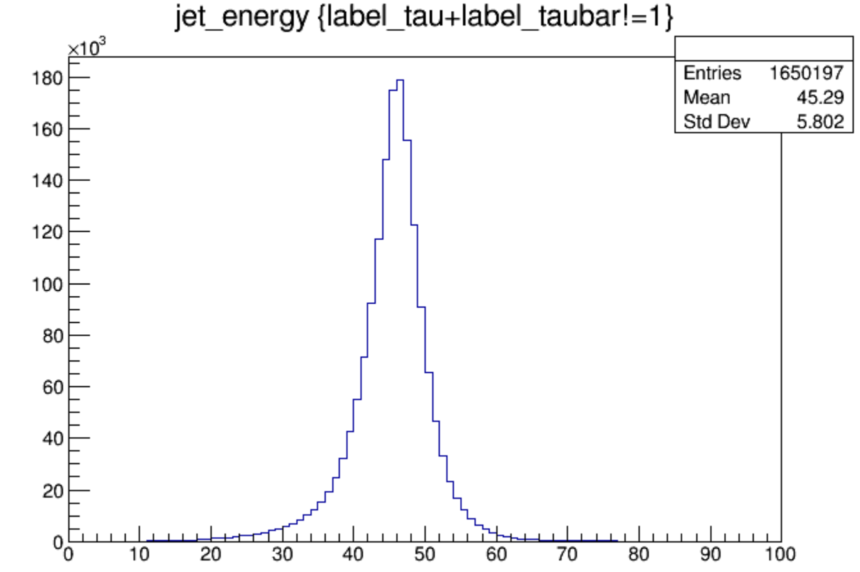
Distributions:

From jet energy and nPFO distribution, tau jets and non-tau jets are very different.



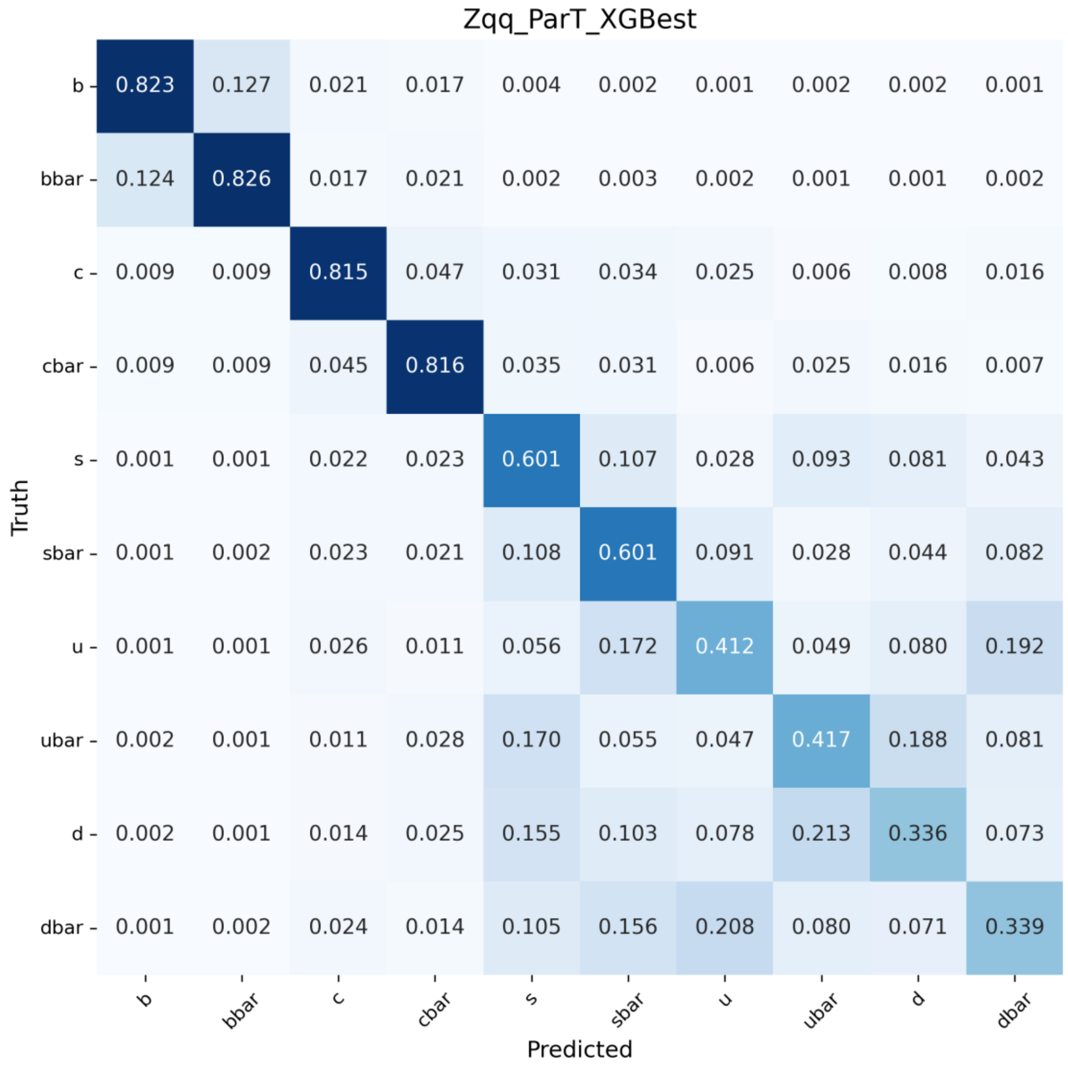






Other performance degrading?

- No. right plots use 10X samples training M10. Performance improved in 3% level, assistant with scale conclusion.
- Adding tau in categories no harm for other tagging.



Conclusion

- Current tests shows good tau tagging performance expected.
  - As long as eekt clustering grab the tau clustering in jets.
  - Should be difficult for merging jets, overlap and so on. Need further study.
  - Anything need for validation. Ready for implement in TDR.

- Yuexin’s 1-1 report
- Fcc-ee status
- CEPC JOI RecoID in Zpole

Fcc-ee Status

- Third Annual US Higgs Factory Future Circular Collider Workshop
- arXiv:2505.00272 Future Circular Collider Feasibility Study Report
- arXiv:2504.11103 Impact of tracker- and calorimeter-detector performance on jet flavor identification and Higgs physics analyses
- https://repository.cern/communities/fcc-ped-sub/records
    - CDS Record for Fcc physics. Recommended to read for analyzers.
    - Hinclusive Z(qq, comb): https://repository.cern/records/c5dn3-c0s73
    - Hmass, inclusive Z(ll): https://repository.cern/records/e9wsh-tb178
    - H(bb,cc,gg): https://repository.cern/records/3jjdh-6fz97
    - Hinvisible: https://repository.cern/records/9b128-qqc43/

Fcc-ee: Current setup in analysis

Fcc-ee:	10.8iab@ZH240, 3.1iab@ttbar
CEPC refTDR:	20iab@ZH240, 1iab@ttbar

- Generator: Whizard3.1.2+Pythia6
- Fast simulation: Delphes+IDEA detector
- Vertex resolution: 3um
- In concept, Fcc has better vertex and Hcal detector compared to CEPC.
- Most their analysis done in fast simulation.

Table 1: The baseline FCC-ee operation model with four interaction points, showing the centre-of-mass energies, design instantaneous luminosities for each IP, and integrated luminosity per year summed over 4 IPs. The integrated luminosity values correspond to 185 days of physics per year and a 75% operational efficiency (i.e., 1.2×10^7 seconds per year) [15], in the Z, WW, ZH, and $t\bar{t}$ baseline sequence. The last two rows indicate the total integrated luminosity and number of events expected to be produced in the four detectors. The number of WW events includes all $\sqrt{s}$ values from 157.5 GeV up.

Working point	Z pole	WW thresh.	ZH	$t\bar{t}$	
$\sqrt{s}$ (GeV)	88, 91, 94	157, 163	240	340–350	365
Lumi/IP (10^{34} cm^{-2}s^{-1})	140	20	7.5	1.8	1.4
Lumi/year (ab^{-1})	68	9.6	3.6	0.83	0.67
Run time (year)	4	2	3	1	4
Integrated lumi. (ab^{-1})	205	19.2	10.8	0.42	2.70
Number of events	6×10^{12} Z	2.4×10^8 WW	2.2×10^6 ZH + 65k WW $\to$ H	2×10^6 $t\bar{t}$ + 370k ZH + 92k WW $\to$ H	

The nominal energy resolution for the IDEA prototype calorimeters has been assumed from Ref. [30]. In the electromagnetic crystal calorimeter:

$$\sqrt{a^2E^2 + b^2E + c^2} \qquad (1)$$

with $a = 0.005$, $b = 0.03$ and $c = 0.002$.

In the hadronic dual-readout calorimeter:

$$\sqrt{d^2E^2 + e^2E + f^2} \qquad (2)$$

Fcc-ee: Higgs mass

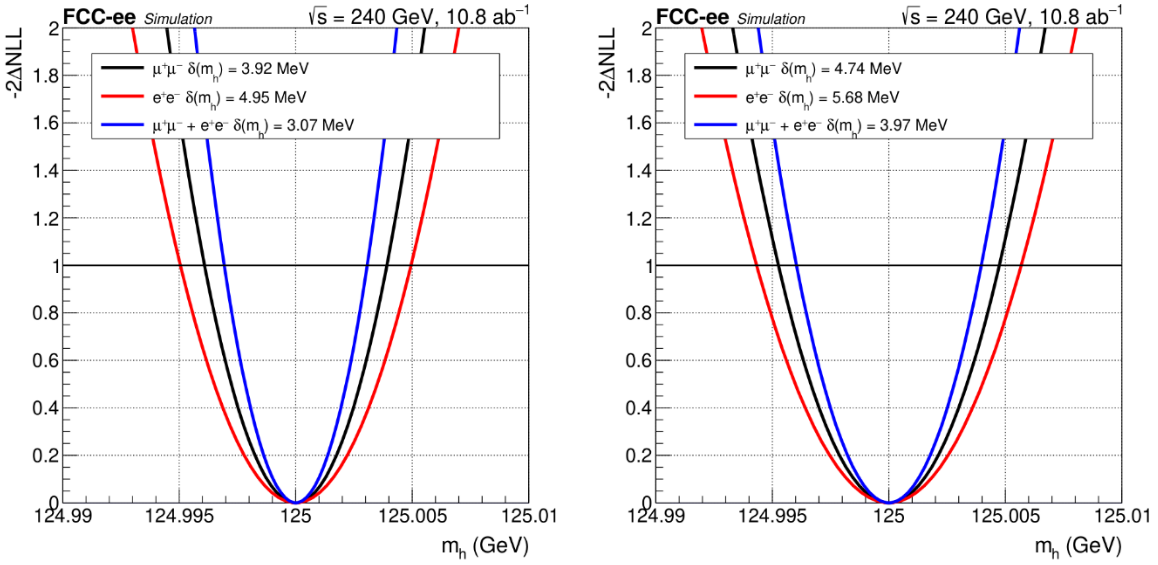


Fig. 23 Likelihood scan statistical-only (left) and statistical+systematics (right).

Table 5 Statistical (stat+syst) uncertainty on the Higgs mass (MeV) for various fit configurations. Fits are performed using the nominal categorization unless stated otherwise. The values in brackets represent the statistical+systematic uncertainties. Values are normalized to an integrated luminosity of 10.8 ab^{-1}.

Fit configuration	$\mu^+\mu^-$ channel	e^+e^- channel	combination
Nominal	3.92 (4.74)	4.95 (5.68)	3.07 (3.97)
Degradation electron resolution	3.92 (4.74)	5.79 (6.33)	3.24 (4.12)
Magnetic field 3T	3.22 (4.14)	4.11 (4.83)	2.54 (3.52)
CLD 2T (silicon tracker)	5.11 (5.73)	5.89 (6.42)	3.86 (4.55)
BES 6% uncertainty	3.92 (4.79)	4.95 (5.92)	3.07 (3.98)
No beam energy spread	2.11 (3.31)	2.93 (3.88)	1.71 (2.92)
Ideal resolution	3.12 (3.95)	3.58 (4.52)	2.42 (3.40)
Freeze backgrounds	3.91 (4.74)	4.95 (5.67)	3.07 (3.96)
Remove backgrounds	3.08 (4.13)	3.51 (4.58)	2.31 (3.45)

Hadronic decays:

Table 23: Expected Precision (%) for the $\sigma(ZH) \times \mathrm{BR}(H \to jj)$ at 68 % CL

	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to s\bar{s}$	$H \to \tau\tau$	$H \to ZZ$	$H \to WW$
$Z \to ll$ $(l = \mu, e)$	0.60	3.47	1.93	220	2.54	7.65	1.49
$Z \to qq$	0.32	3.52	3.07	410	110	50	8.74
$Z \to \nu\bar{\nu}$ (I)	0.35	2.06	1.01	100	10.6	11.4	1.28
$Z \to \nu\bar{\nu}$ (II)	0.33	2.27	0.94	140	21.8	19.8	1.89
Combined ($Z \to \nu\nu$ (I))	0.21	1.56	0.85	89	2.46	6.24	0.95
Combined ($Z \to \nu\nu$ (II))	0.21	1.66	0.80	105	3.97	10.1	1.16

Table 24: Relative Uncertainty (%) for the $\sigma_{ZH} \times \mathcal{B}(H \to XX)$ and $\sigma_{\nu_e\bar{\nu}_e} \times \mathcal{B}(H \to XX)$ at 68 % CL

$\sqrt{s}$	240 GeV		365 GeV	
Integrated luminosity	10.8 ab^{-1}		3.0 ab^{-1}	
Channel	ZH	$\nu_e\bar{\nu}_e H$	ZH	$\nu_e\bar{\nu}_e H$
$H \to b\bar{b}$	±0.21	±1.89	±0.41	±0.67
$H \to c\bar{c}$	±1.61	±19.4	±3.13	±3.49
$H \to s\bar{s}$	±120	±990	±360	±290
$H \to gg$	±0.80	±5.50	±2.21	±2.66
$H \to WW$	±1.17	±15.6	±3.18	±5.36
$H \to ZZ$	±9.94	±130	±26.0	±37.1
$H \to \tau^+\tau^-$	±3.67	∞	±11.0	±24.2

Hinclusive:

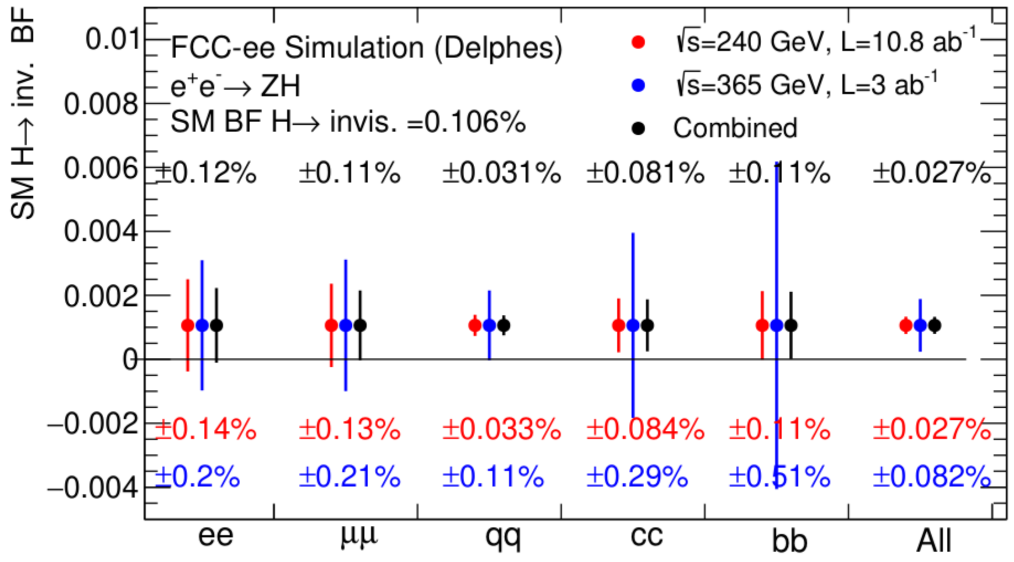
Presented path to model-independent ZH cross-section at center-of-mass energy of 240(365) GeV

- Hadronic channel 0.38(0.57)%
- Hadronic+leptonic combined uncertainty of 0.31(0.53)%
- Proven to be model-independent within quoted uncertainties

Channel	Accelerator	Lumi (fb^{-1})	ZH uncertainty (%)	Scaled to FCC (10.8 ab^{-1})
Hadronic	CLIC [1]	500	3.65	0.79
	ILC [2]	250	2.6/2.4 (+/–)	0.40/0.37
	FCC (this work)	10800	0.41	0.41
Leptonic	ILC [3]	250	2.5/2.9 (+/–)	0.38/0.44
	FCC (this work)	10800	0.52	0.52
Total ZH	ILC/CLIC [4]	250	2.0/2.0 (+/–)	0.30
	FCC CDR	5000	0.50	0.34
	FCC (this work)	10800	0.32	0.32

Full paper ready with systematic study.

Hinvisible:



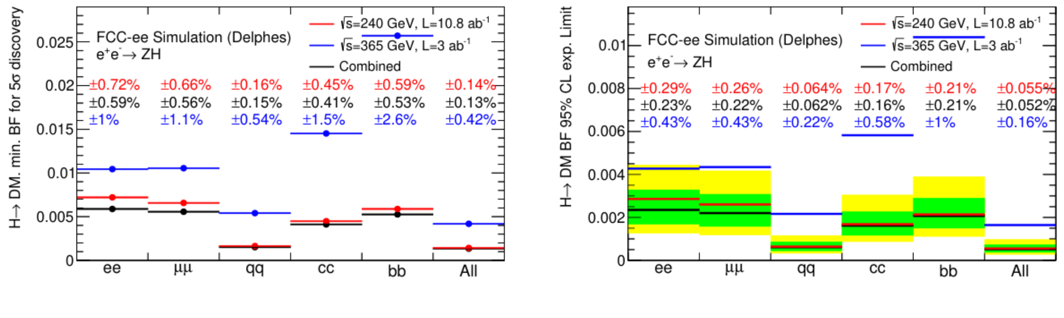


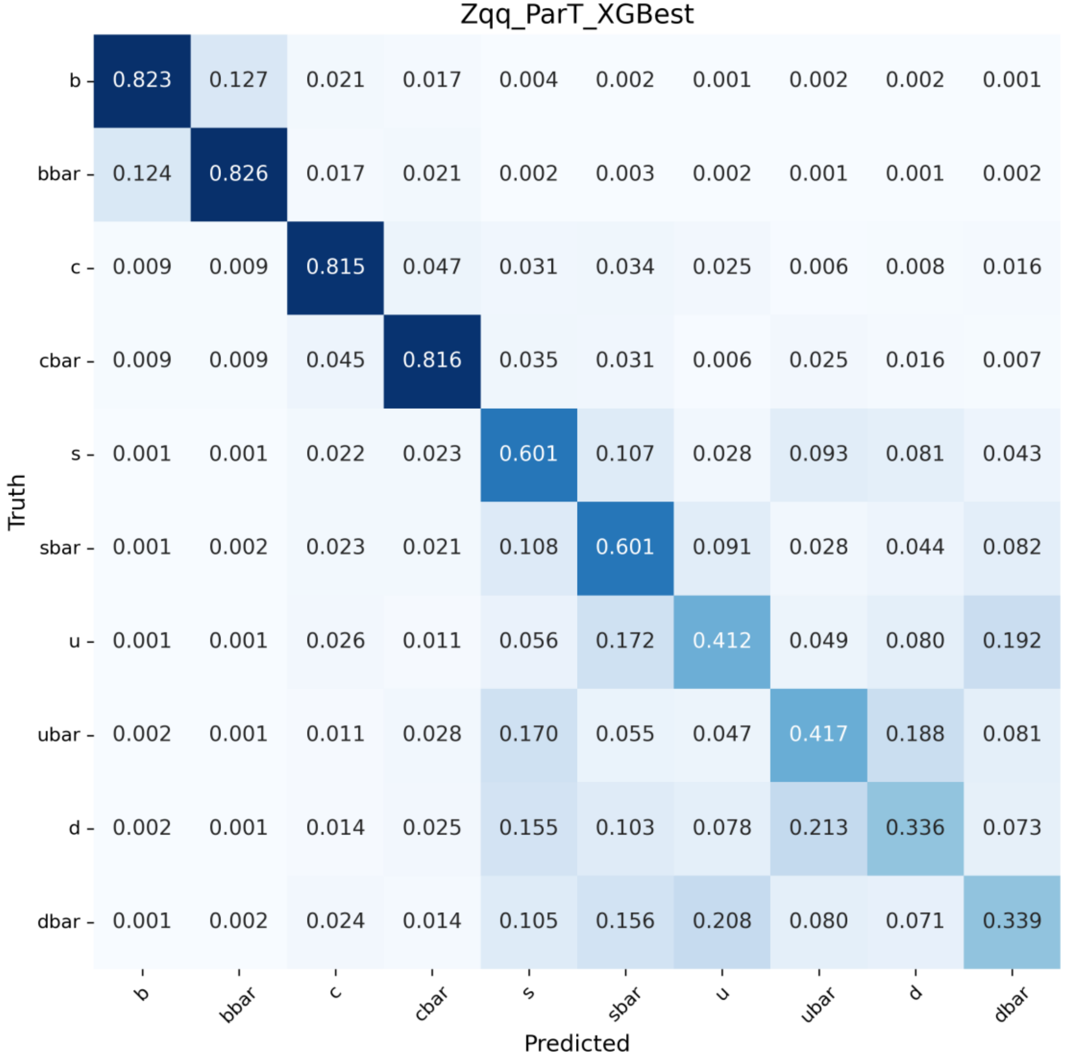
Fig. 11: (a) The minimum for a 5σ discovery and (b) the limit at 95% confidence level on the branching fraction of the Higgs boson to new invisible particles for the individual channels and all channels combined. Results shown are for the data at $\sqrt{s} = 240$ GeV and $\sqrt{s} = 365$ GeV and at the two beam energies combined. The contribution from SM Higgs boson to neutrinos is treated as a background.

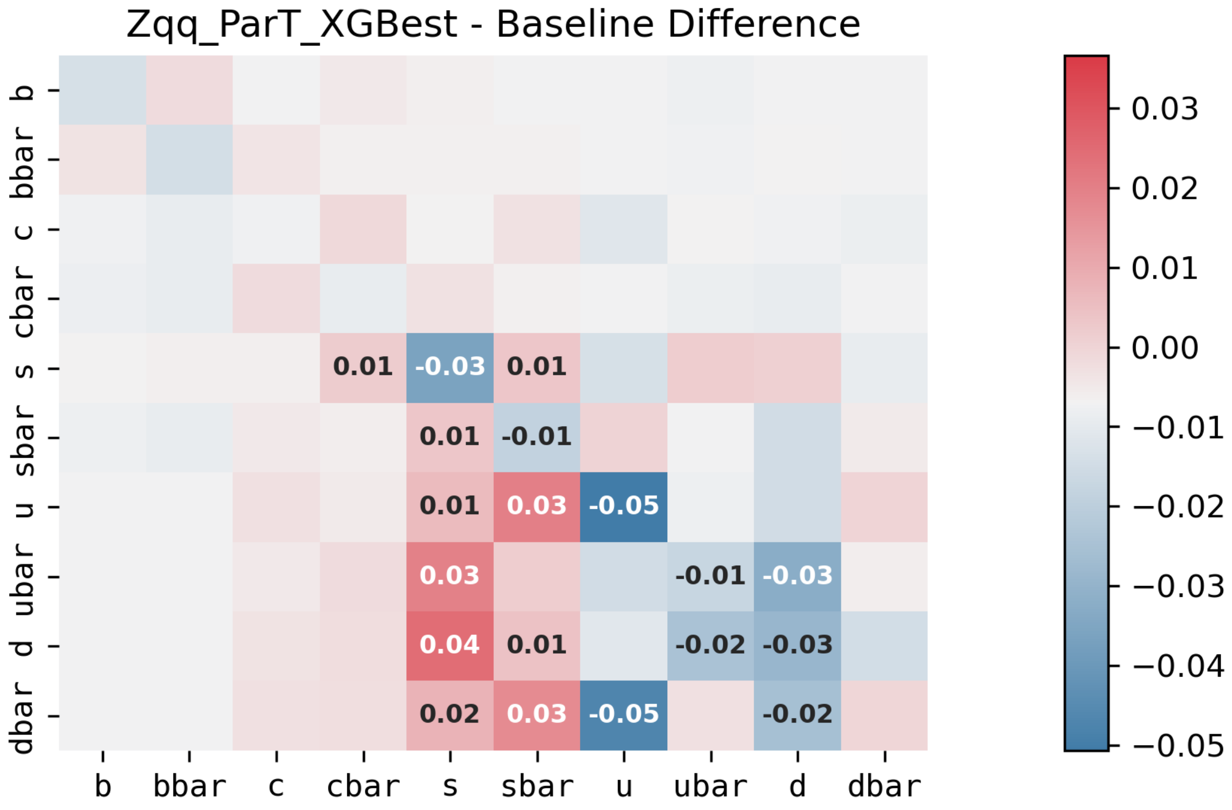
Summary for Fcc status

- Similar hardware + methodology with CEPC.
  - ML-based method widely used like jet tagging.
- Some variation on different detection conditions studied.
- Analysis well prepared including systematics.

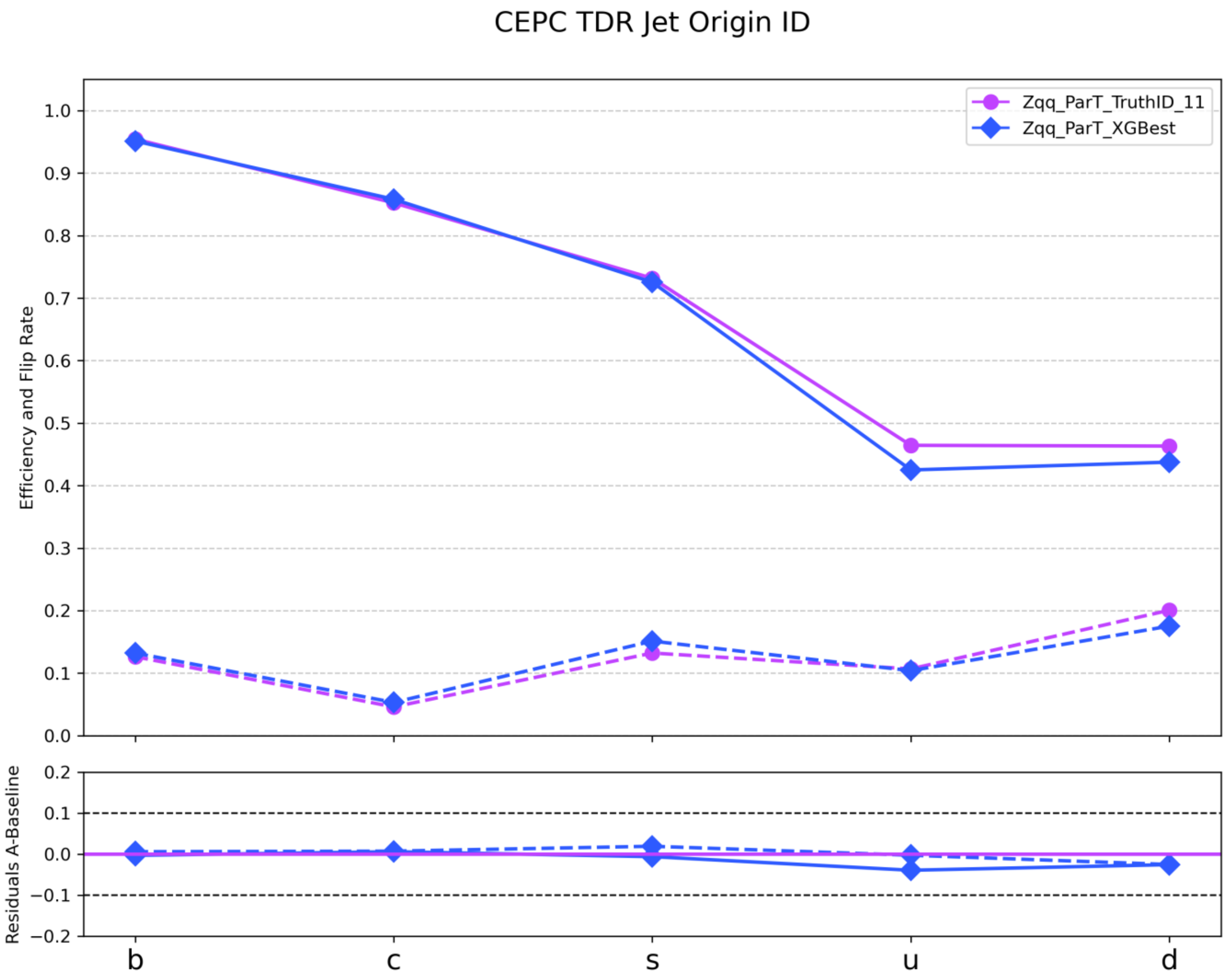
XGBest @Zpole: JOI performance

M10: 0.6134-0.5986





XGBest @Zpole: JOI performance



- However, the reco ID performance for b, c, s is almost kept.
- Good enough for Rb Rc Rs measurement.