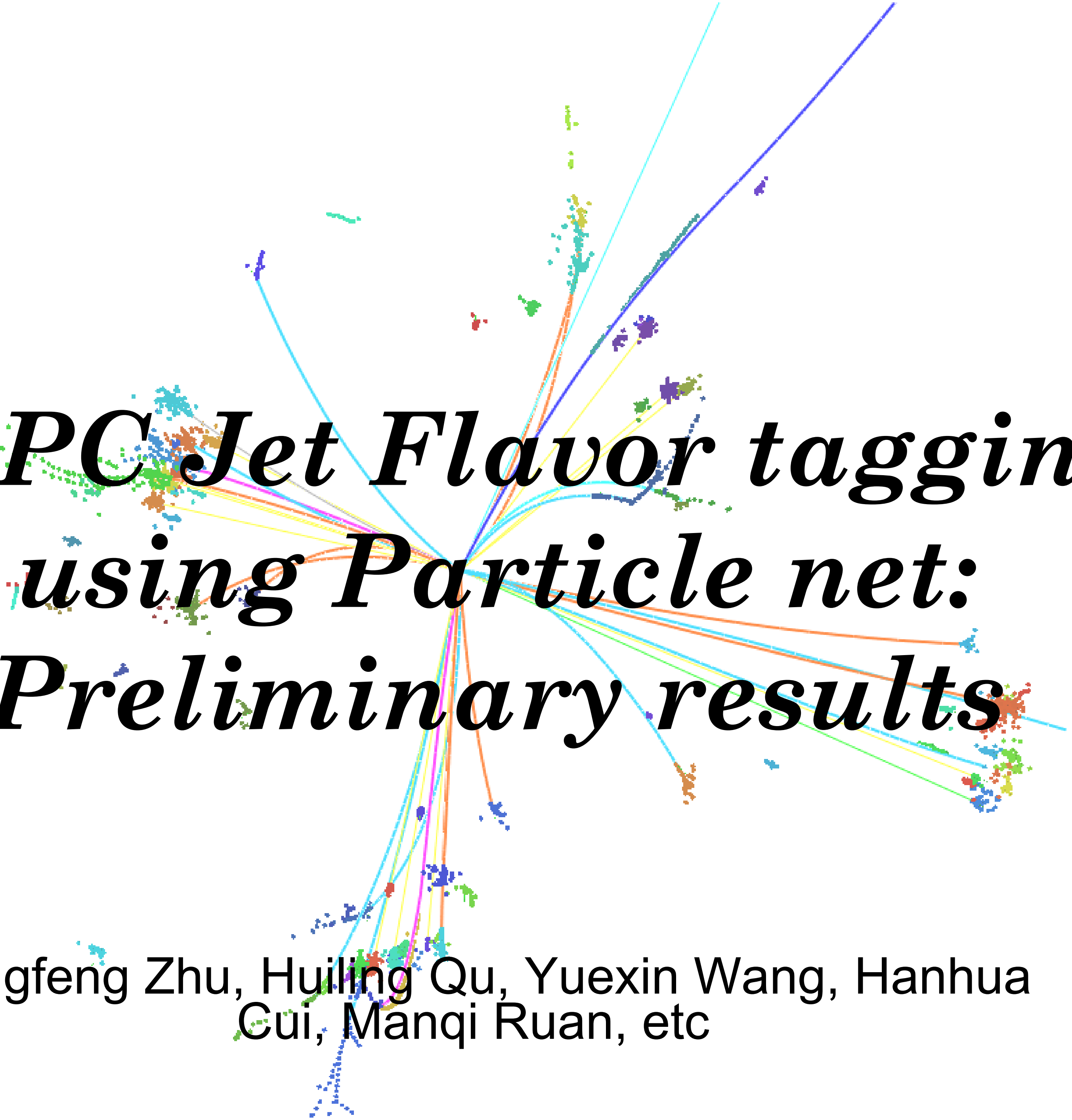
CEPC Jet Flavor tagging using Particle net: Preliminary results



Yongfeng Zhu, Huiling Qu, Yuexin Wang, Hanhua Cui, Manqi Ruan, etc

Flavor tagging

- Critical for High Energy Frontier, especially Future Higgs factory
- Sample: Z->di jet, Higgs->di jet
- Methods:
  - LCFIPlus: baseline (BDT based)
  - LCFIPlus*:
  - Particle Net
- Input:
  - 4 momentum, Pid, Track impact parameters for Charged Particle for all reconstructed final state particles
- Output: b/c-likelihoods, etc

```
- [part_isElectron, null]
- [part_isMuon, null]
- [part_isNeutralHadron, null]
- [part_isPhoton, null]
- [part_d0, null]
- [part_d0err, 0, 1, 0, 1]
- [part_dz, null]
- [part_dzerr, 0, 1, 0, 1]
- [part_deta, null]
- [part_dphi, null]
```

```
- [part_pt_log, -1.5, 1.0]
- [part_e_log, -0.687, 1.0]
- [part_logptrel, -4.7, 1.0]
- [part_logerel, -4.473, 1.0]
- [part_deltaR, 2.1, 2.3]
- [part_charge, null]
- [part_isChargedHadron, null]
```

[part_isKaon, null]

Methods: more details

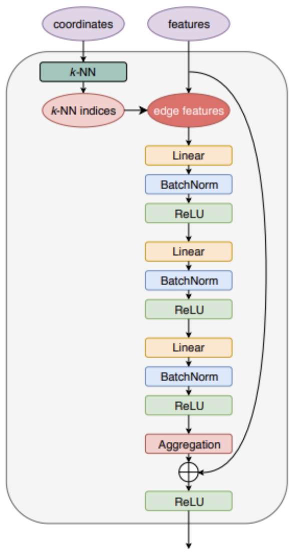
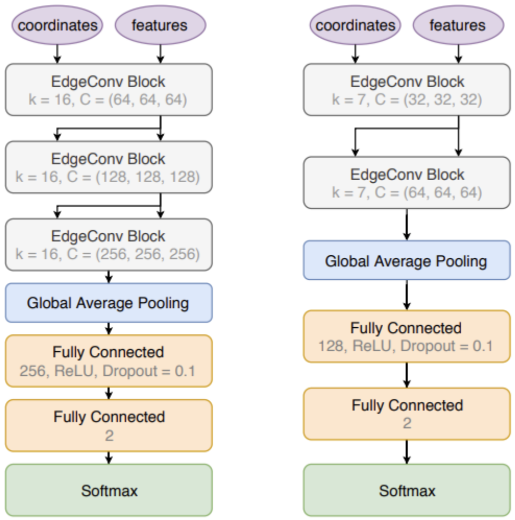


FIG. 1: The structure of the EdgeConv block.



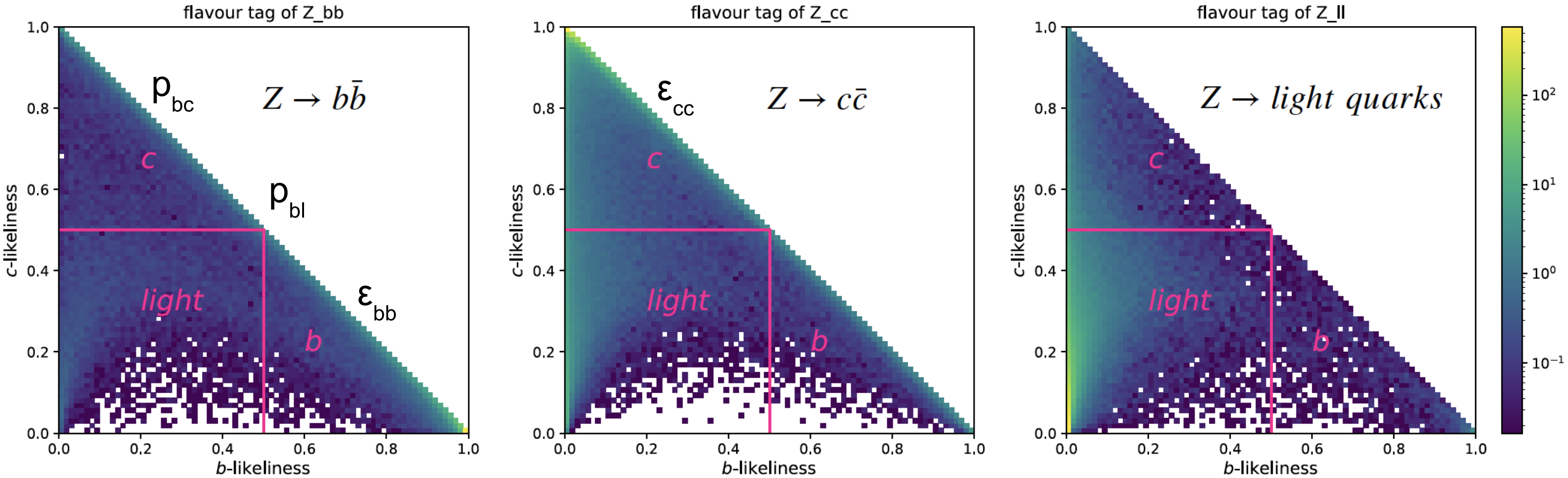
(a) ParticleNet (b) ParticleNet-Lite

FIG. 2: The architectures of the ParticleNet and the ParticleNet-Lite networks.

https://arxiv.org/pdf/1902.08570.pdf

- LCFI:
  - VTX finding;
  - Categories according to VTX number
  - BDT developed for different categories
- LCFI*: BDT (using lightgbm framework to acc. processing)
- PN: Machine learning using Particle Cloud

Performance: Migration Matrix



- For simplicity: line parallel to axis from (0.5, 0.5)

predicted

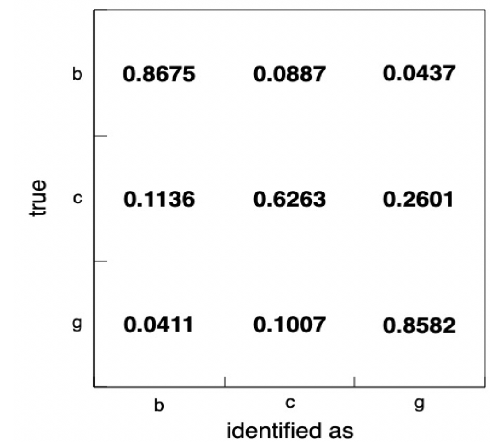
truth	b	c	uds
b	0.905	0.063	0.031
c	0.04	0.768	0.192
uds	0.006	0.058	0.936

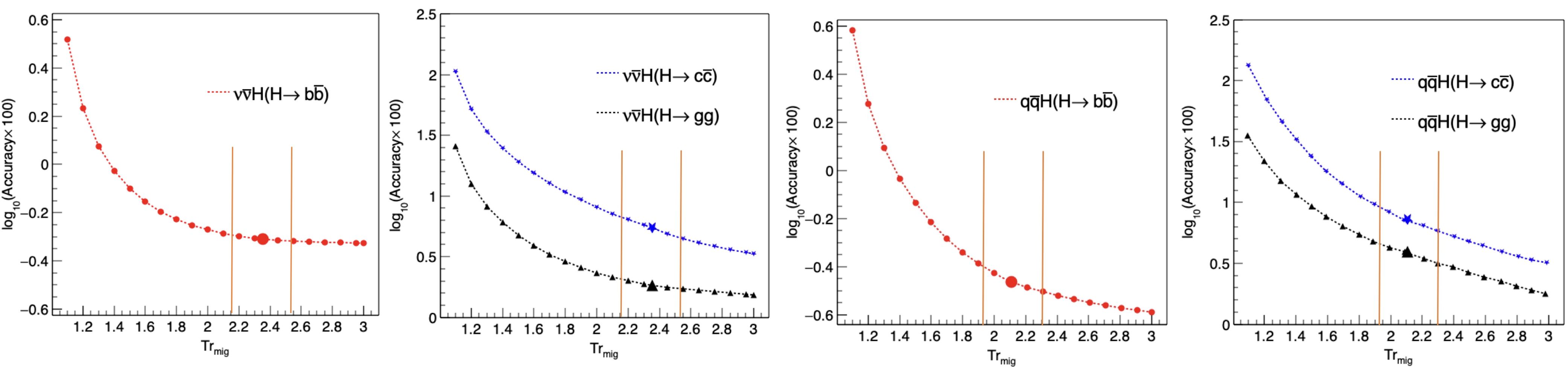
Flavor Tagging: impact on H→2jet

$$M_{mig} = \frac{Tr_{mig} - Tr_{opt}}{Tr_{l} - Tr_{opt}} \cdot (M_{l} - M_{opt}) + M_{opt}$$

$$M_{mig} = \frac{Tr_{mig} - Tr_{opt}}{Tr_{1/3} - Tr_{opt}} \cdot (M_{1/3} - M_{opt}) + M_{opt}$$

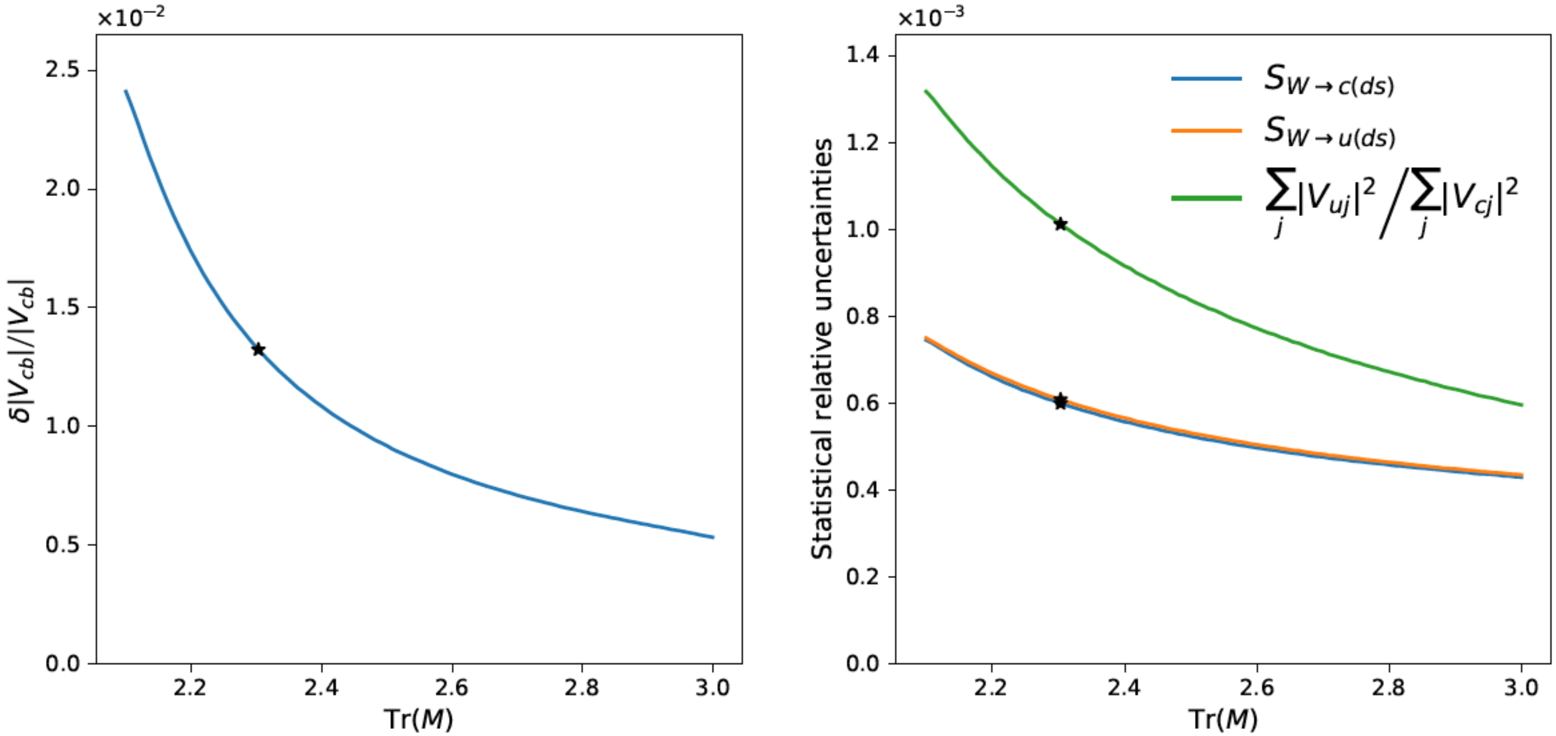
LCFI@Baseline





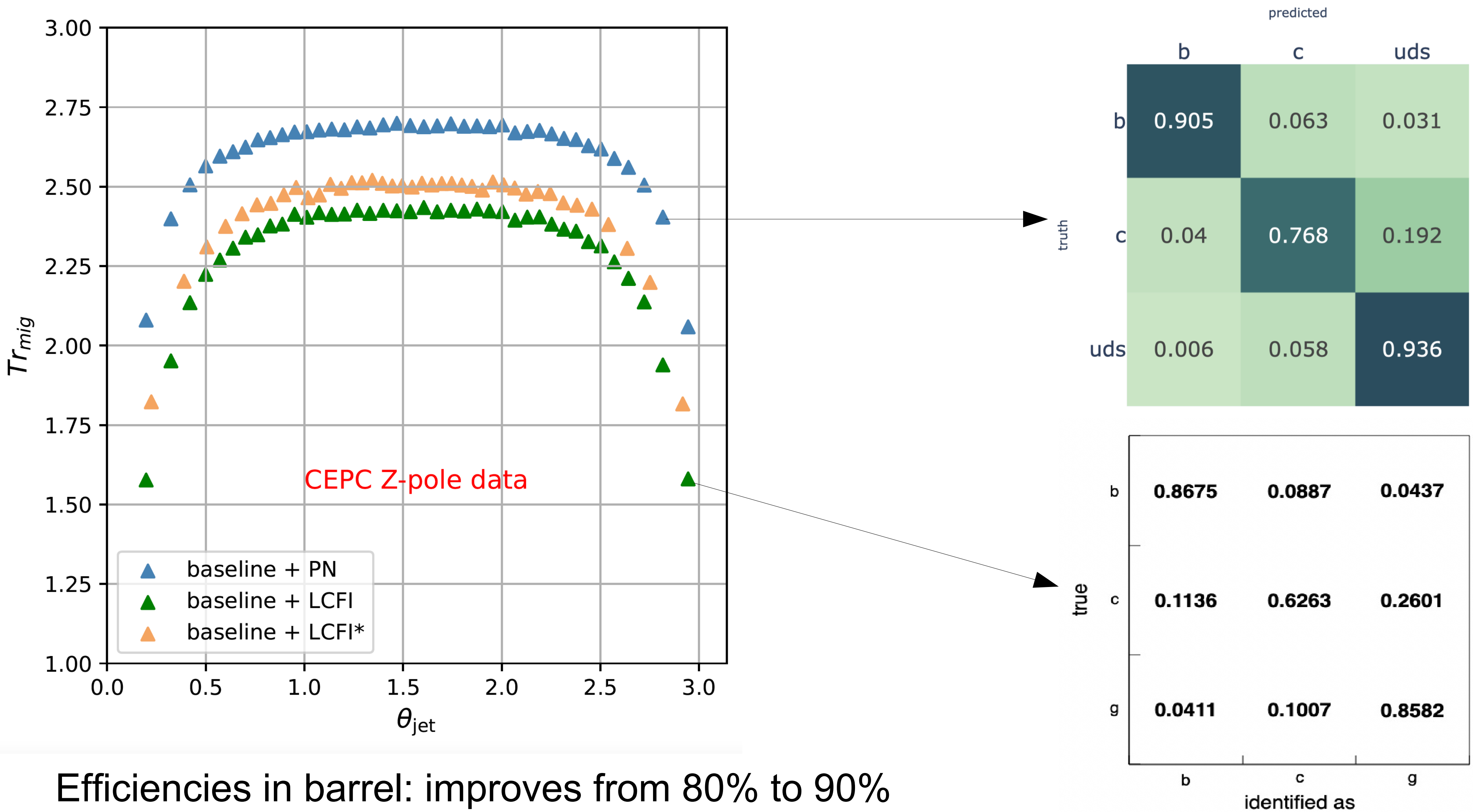
- Compared to baseline, perfect Flavor tagging improves the accuracy by 2%/63%/13% for vvH and 35%/120%/180% for qqH channels (bb, cc, gg)

@ Vcb



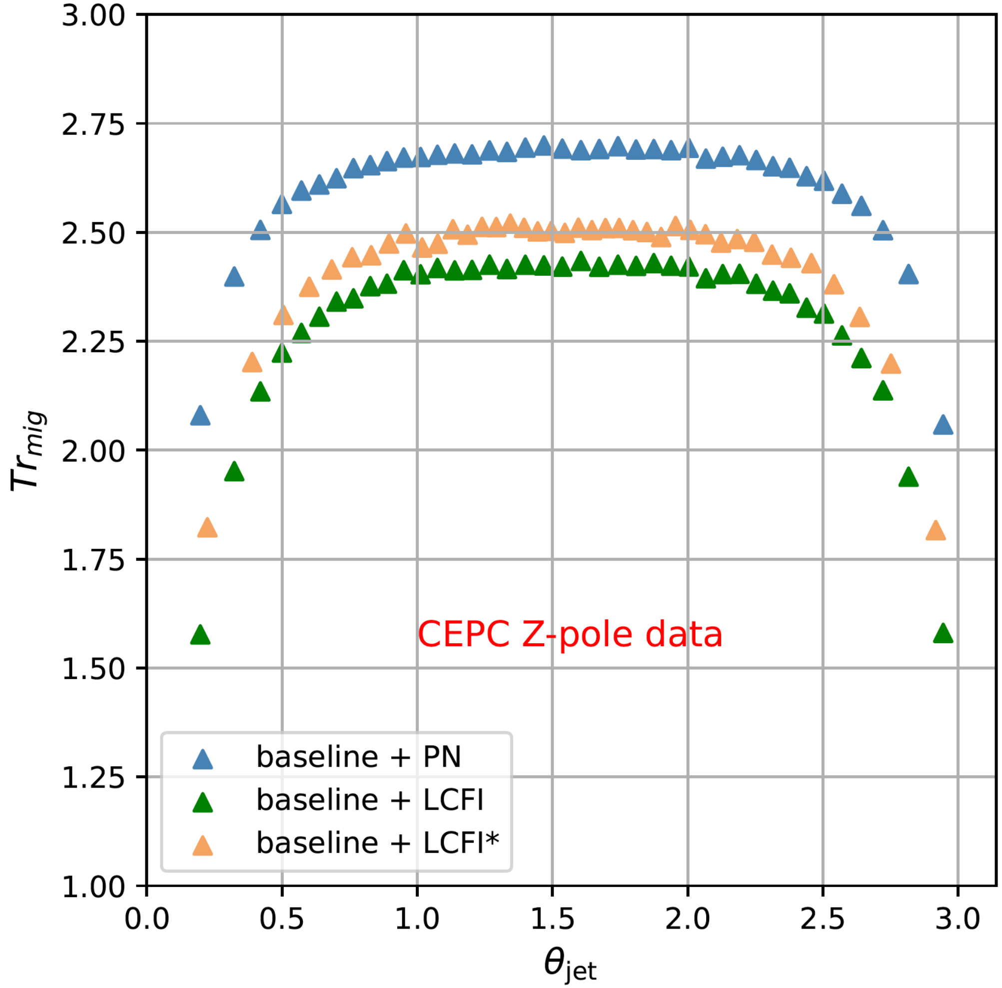
Similar performance dependence on CKM measurements at 240 GeV using semi-leptonic WW events

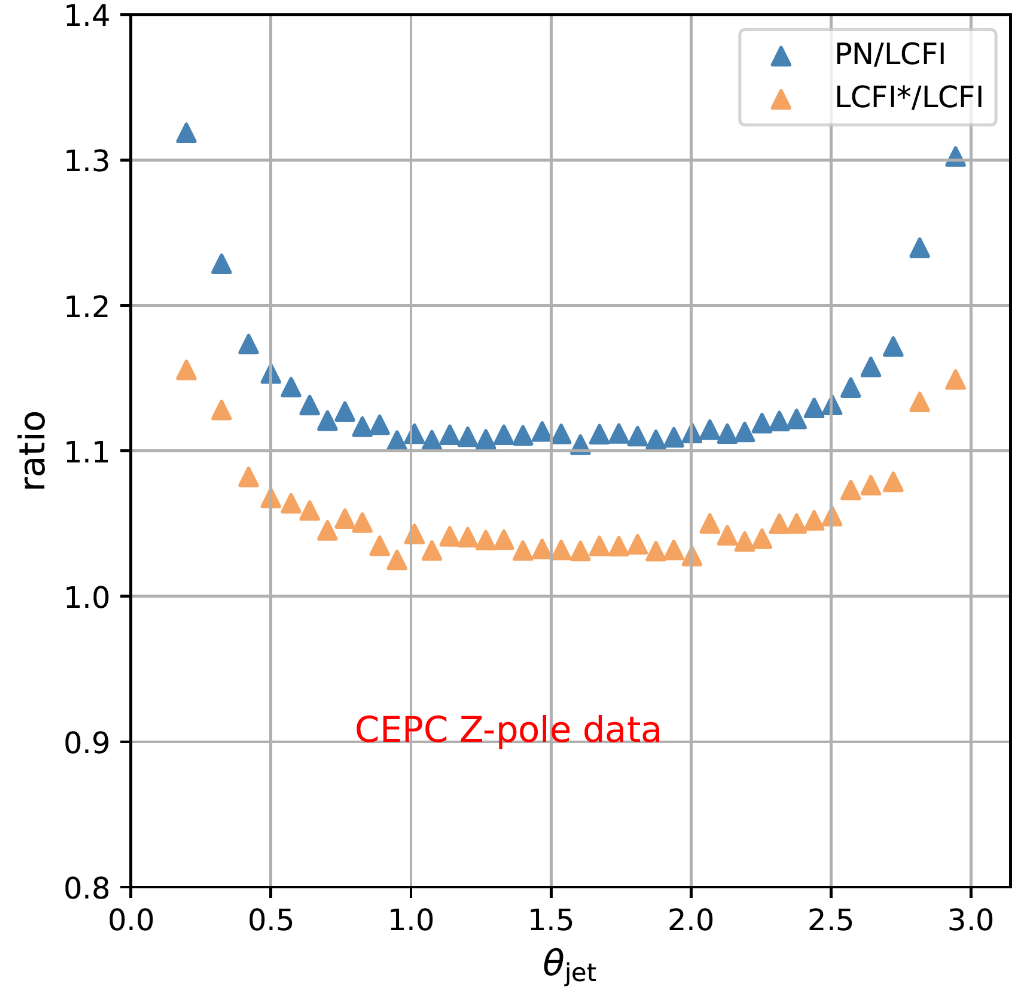
Performance Comparison



Efficiencies in barrel: improves from 80% to 90%

Performance Comparison





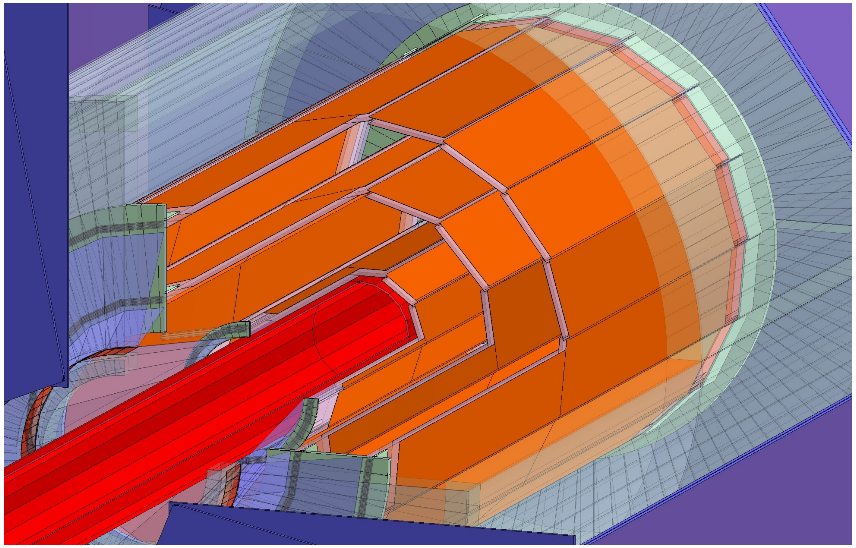
Improves much significant in forward region...

Remark: No Pid

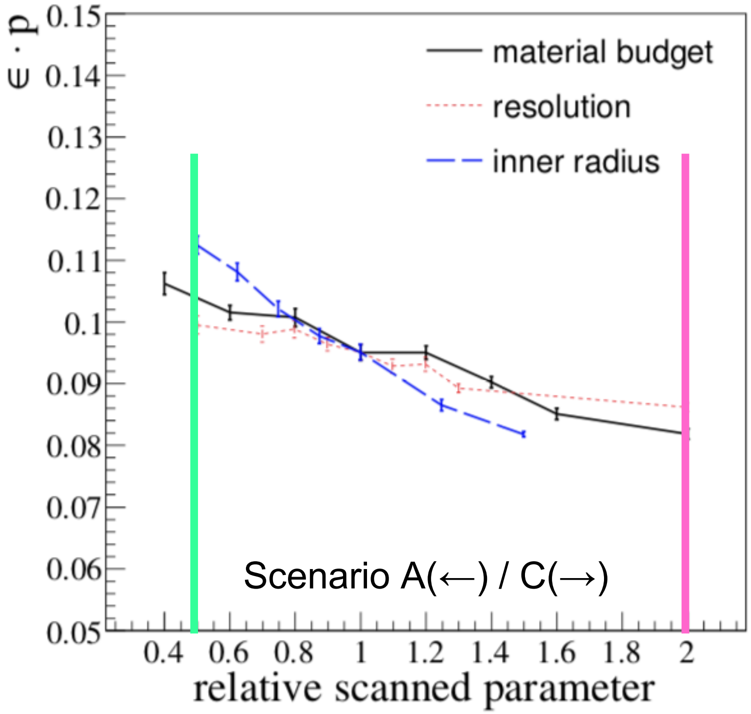
Vertex

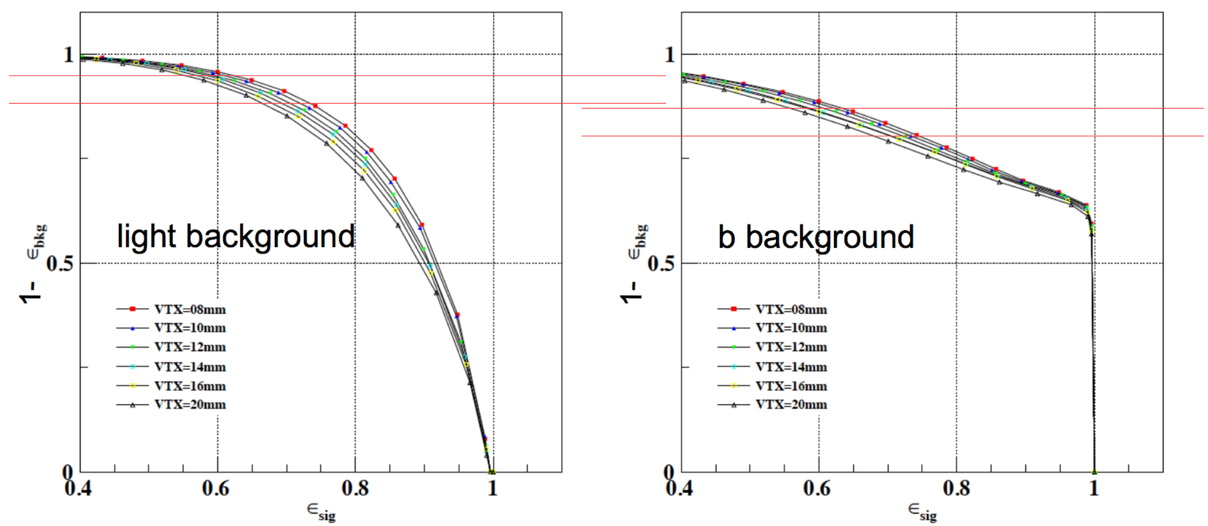
- Closer, Thinner, Preciser w.r.t. Baseline
- 3 Scenarios
  - Smaller radius: 10 mm inner radius
  - **Vin (Vertex inside): innermost layer inside Beam pipe**
  - Vin portable: Movable innermost layers inside beam pipe

	R(mm)	Z(mm)	single-point resolution(μm)	material budget
Layer 1	16	62.5	2.8	$0.15\%/X_0$
Layer 2	18	62.5	6	$0.15\%/X_0$
Layer 3	37	125.0	4	$0.15\%/X_0$
Layer 4	39	125.0	4	$0.15\%/X_0$
Layer 5	58	125.0	4	$0.15\%/X_0$
Layer 6	60	125.0	4	$0.15\%/X_0$



Flavor tagging V.S. VTX geometry



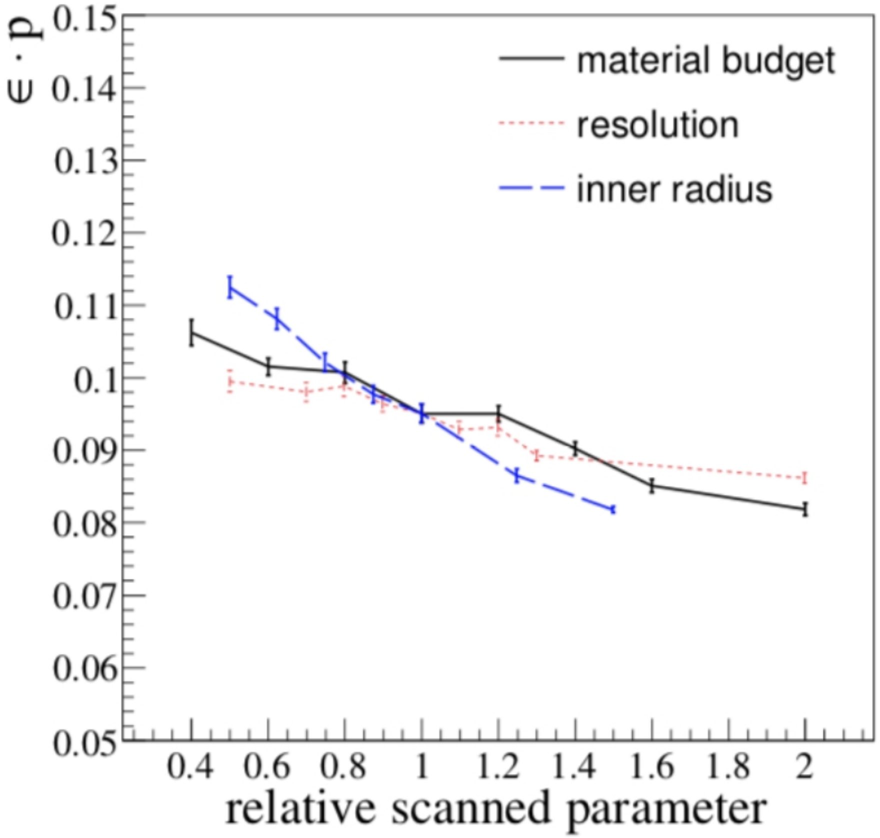


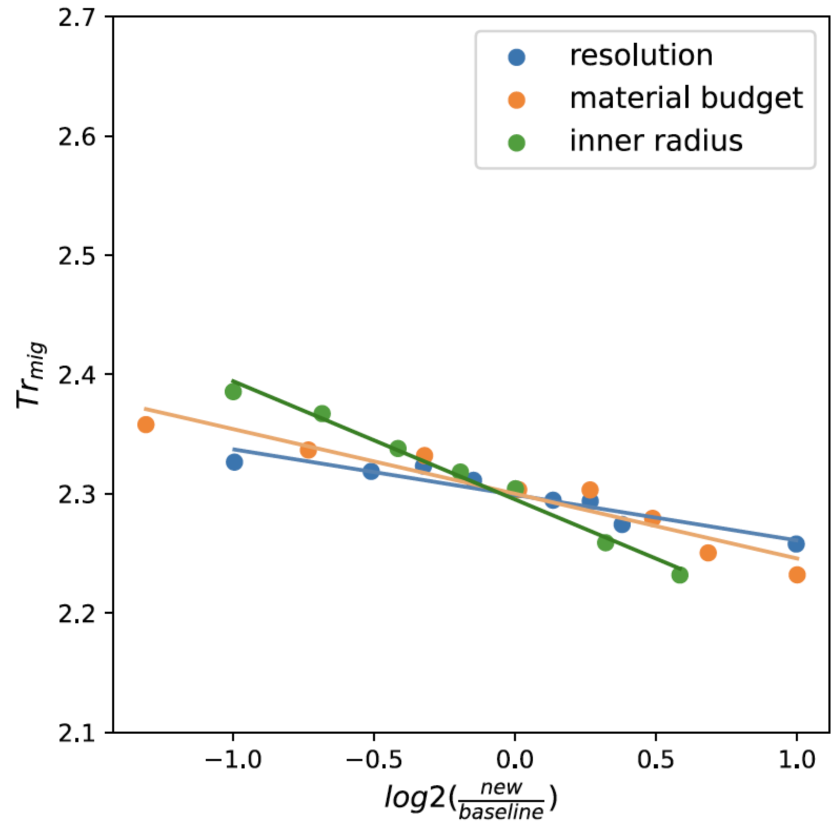
$$\epsilon \cdot p = 0.095(1 - 0.14\frac{\Delta x_{\mathrm{material}}}{x_{\mathrm{material}}})(1 - 0.09\frac{\Delta x_{\mathrm{resolution}}}{x_{\mathrm{resolution}}})(1 - 0.23\frac{\Delta x_{\mathrm{radius}}}{x_{\mathrm{radius}}})$$

Table 2. Reference geometries.

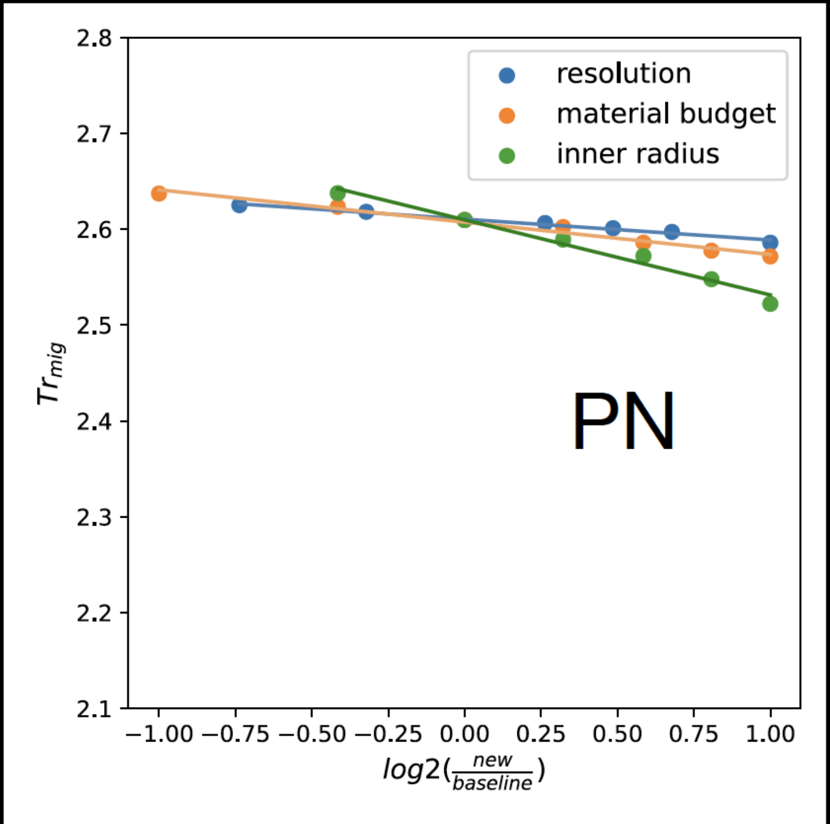
	Scenario A (Aggressive)	Scenario B (Baseline)	Scenario C (Conservative)
Material per layer/X_0	0.075	0.15	0.3
Spatial resolution/μm	1.4 - 3	2.8 - 6	5 - 10.7
R_{in}/mm	8	16	23

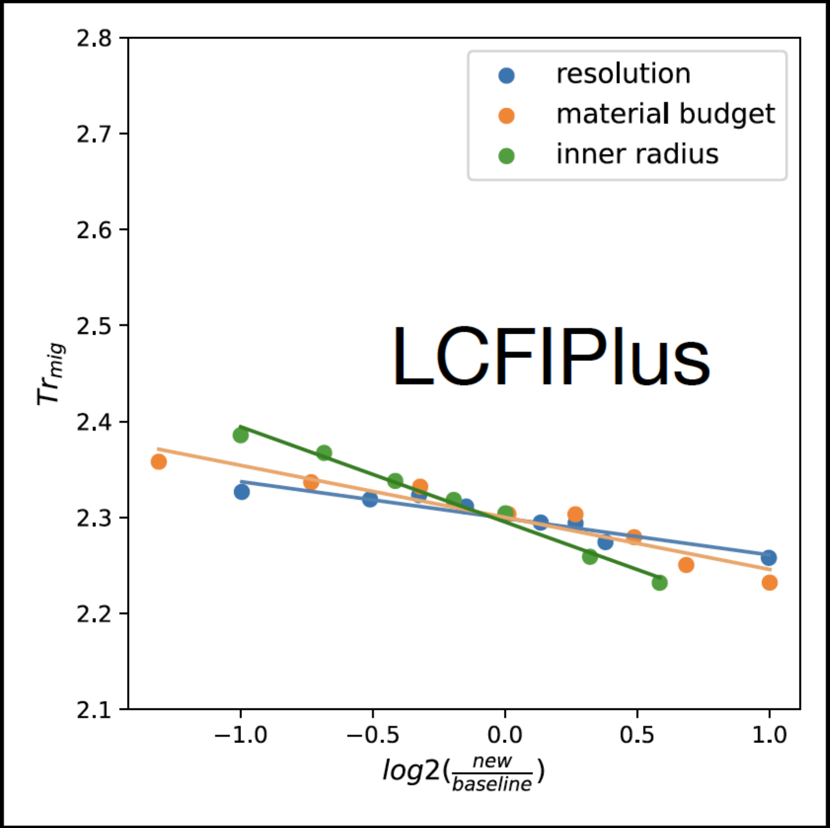
From eff*purity → Tr(MM)





the dependence of JFT on vertex detector configuration

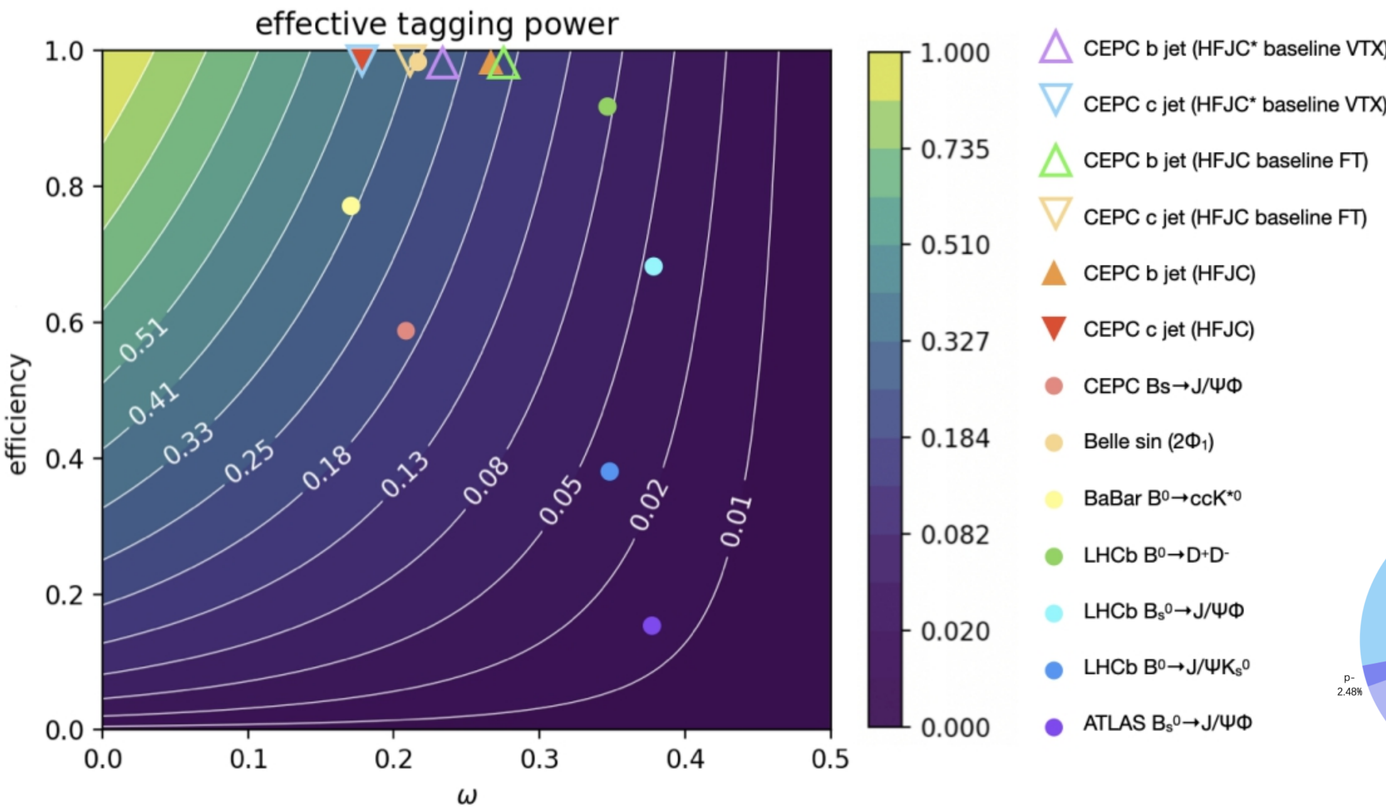


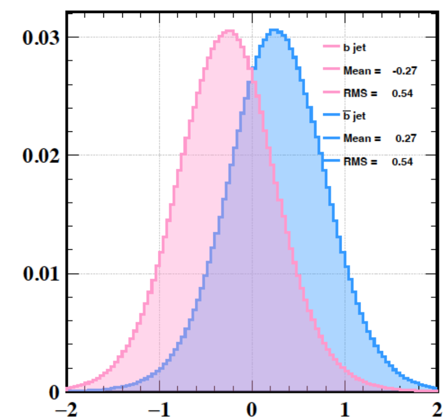


$$\text{PN}: Tr_{mig} = 2.609 + 0.034 \cdot log_2\frac{R^0_{material}}{R_{material}} + 0.022 \cdot log_2\frac{R^0_{resolution}}{R_{resolution}} + 0.078 \cdot log_2\frac{R^0_{radius}}{R_{radius}}$$

$$\text{LCFIPlus}: Tr_{mig} = 2.303 + 0.05 \cdot log_2\frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2\frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2\frac{R^0_{radius}}{R_{radius}}$$

Jet Charge @ Conventional





https://arxiv.org/pdf/2306.14089.pdf

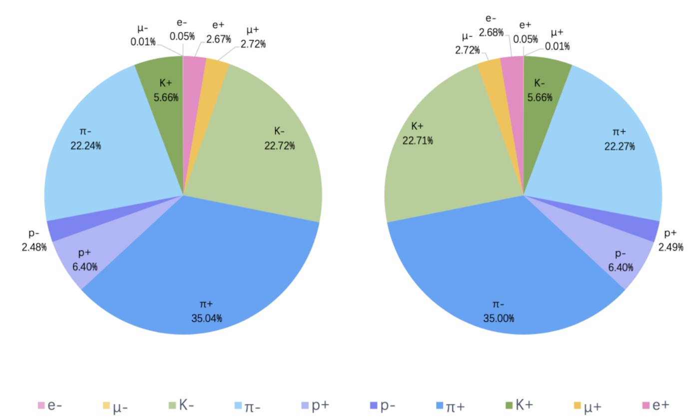
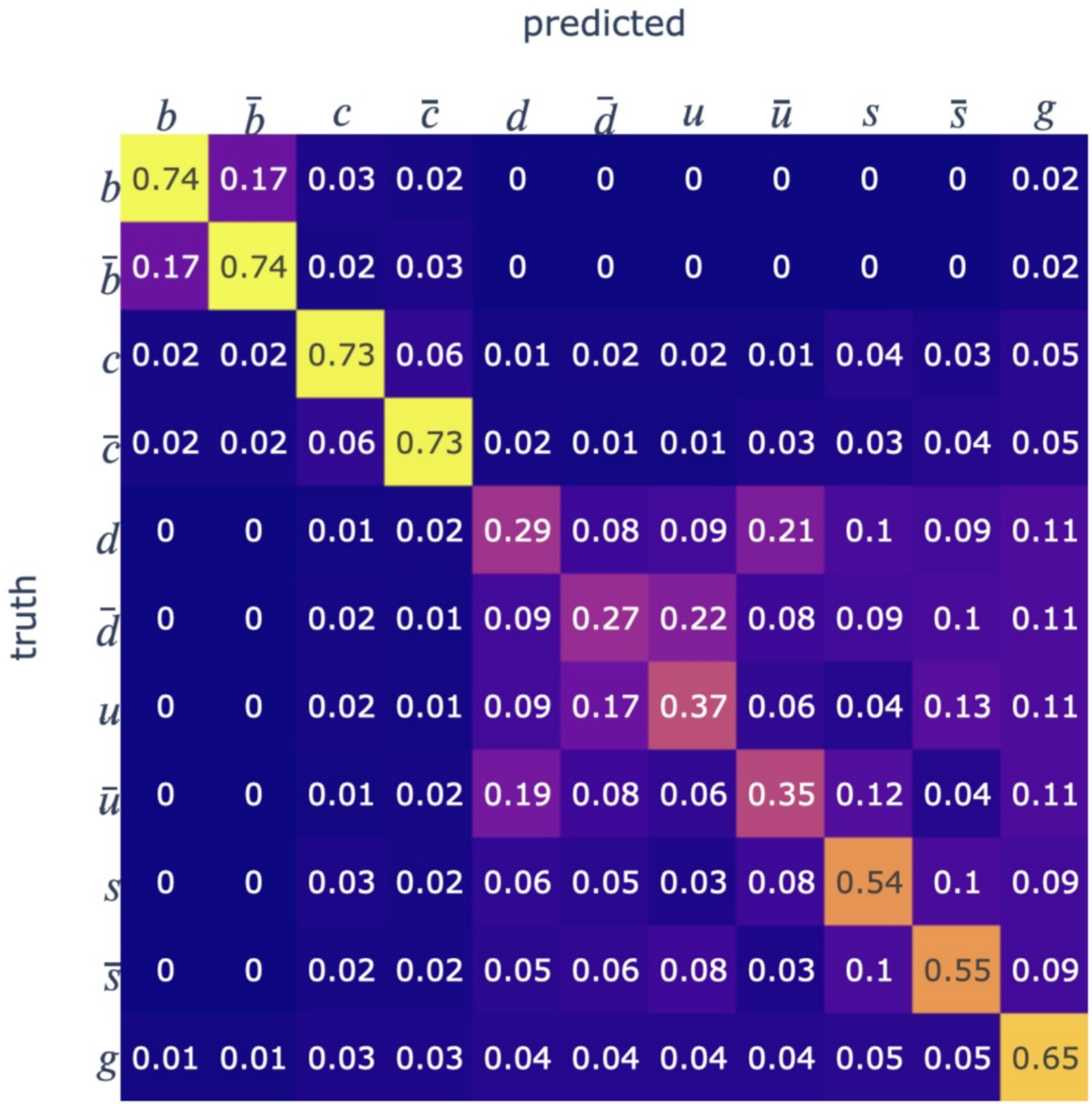


Figure 3. The percentages of species of final state leading charged particles within the c jet (left) and $\bar{c}$ jet (right).

- Eff. Tagging power (eff*(1-2*omega)^2): ~ 20%/40% for c/b jet (@ Z pole)



Remark: Higgs to dijet sample, Ideal Pid

Summary

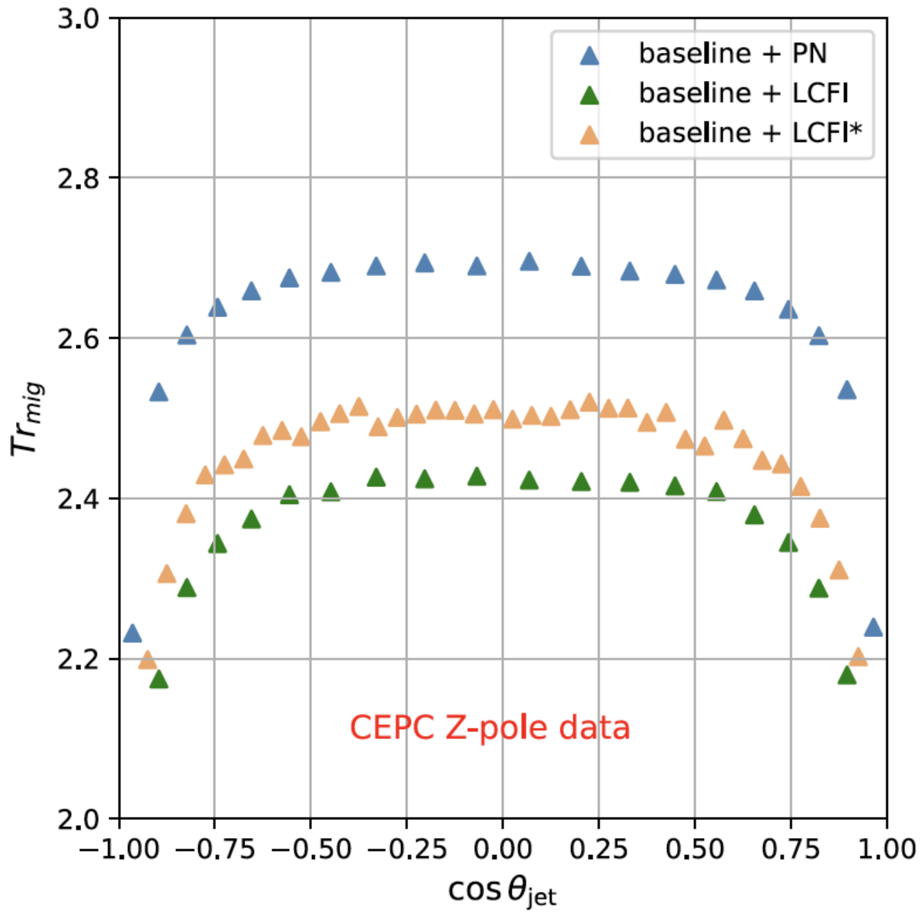
- Compared to conventional methods (LCFI), Machine Learning (Particle Net)
  - Improves significantly: Tr(MM) from 2.4 → 2.7 at baseline (barrel)
    - ~ 50% improvements on H→gg/cc measurements
  - Gives consistent conclusion on VTX optimization
    - Inner Radius > Material budget > Position Resolution

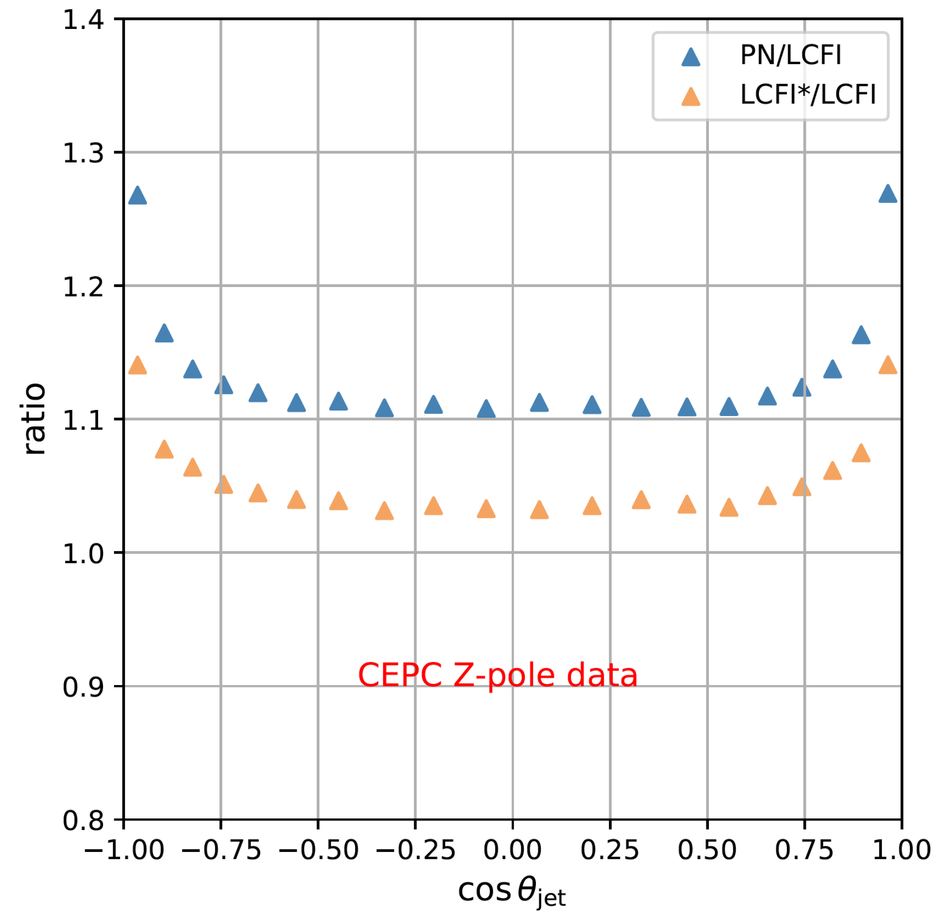
- Vision: jet identification (quark species & gluon) as particle identification
  - New paradigm that combines jet charge, flavor & quark-gluon jet id
  - Huge impact on Physics
  - *Entanglement with Jet Clustering, CSI...*

- To do: lots of scan, comparisons, application...

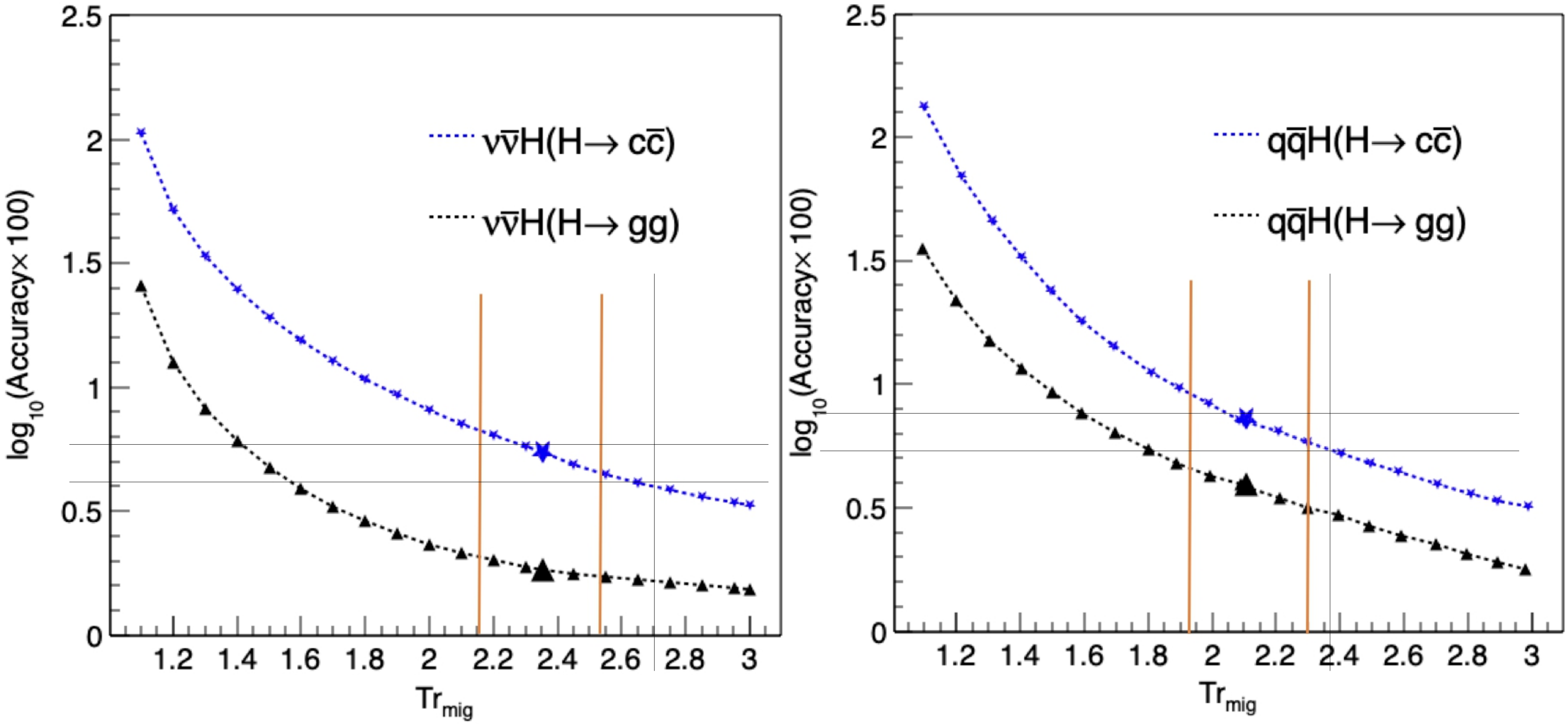
Back up

Performance Comparison cos(theta)





Improves much significant in forward region...



2.5
2
1.5
1
0.5
0
log10(Accuracy× 100)
νν̄H(H→ cc̄)
νν̄H(H→ gg)
1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Trmig
2.5
2
1.5
1
0.5
0
log10(Accuracy× 100)
qq̄H(H→ cc̄)
qq̄H(H→ gg)
1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Trmig