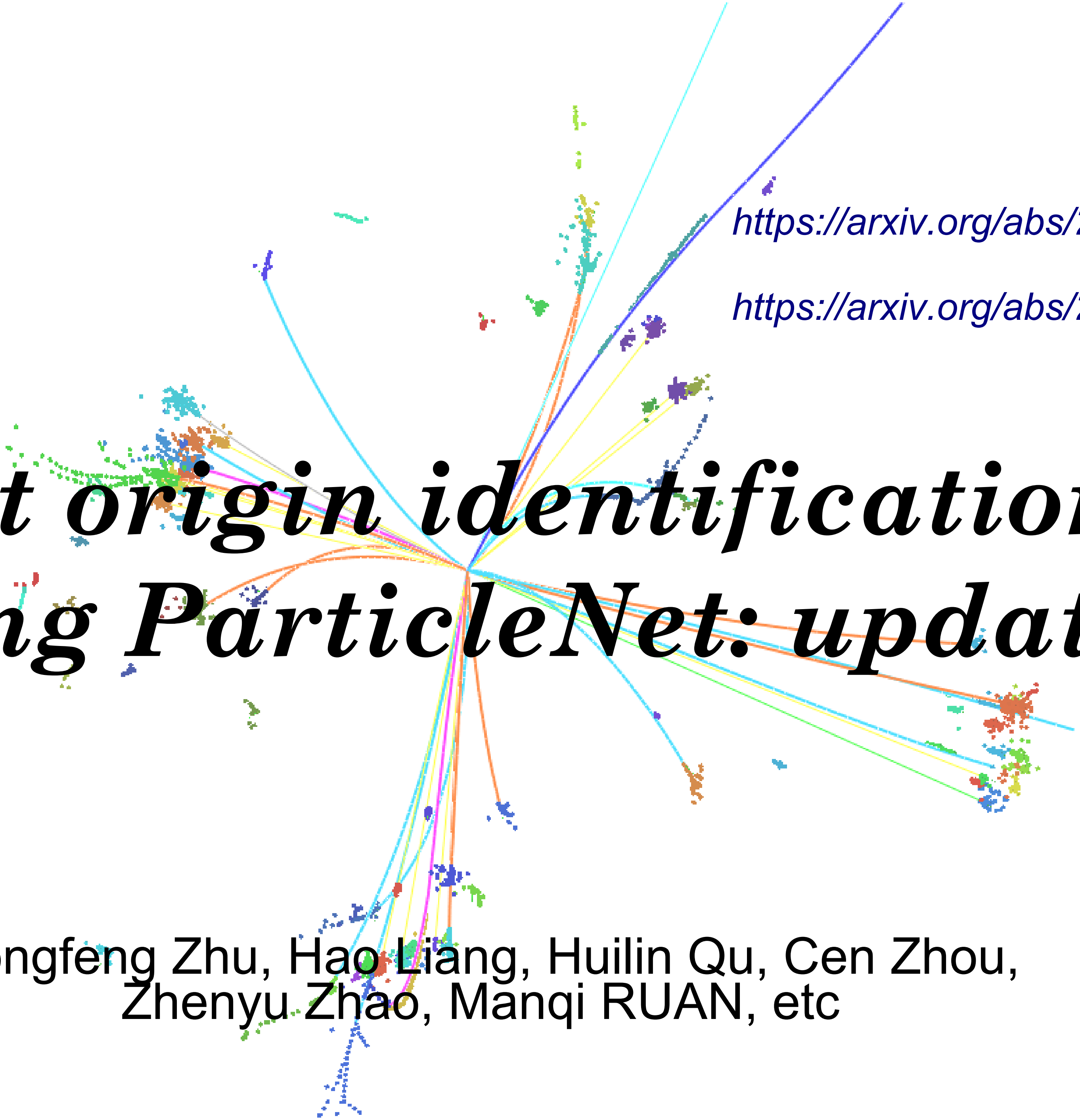
Jet origin identification using ParticleNet: updates

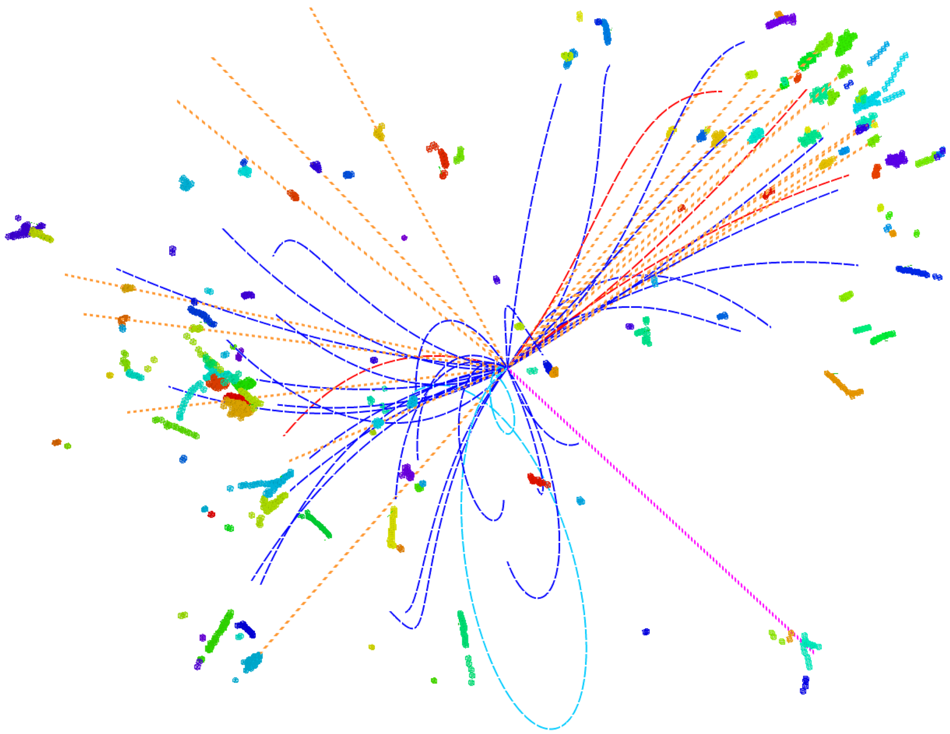
https://arxiv.org/abs/2310.03440

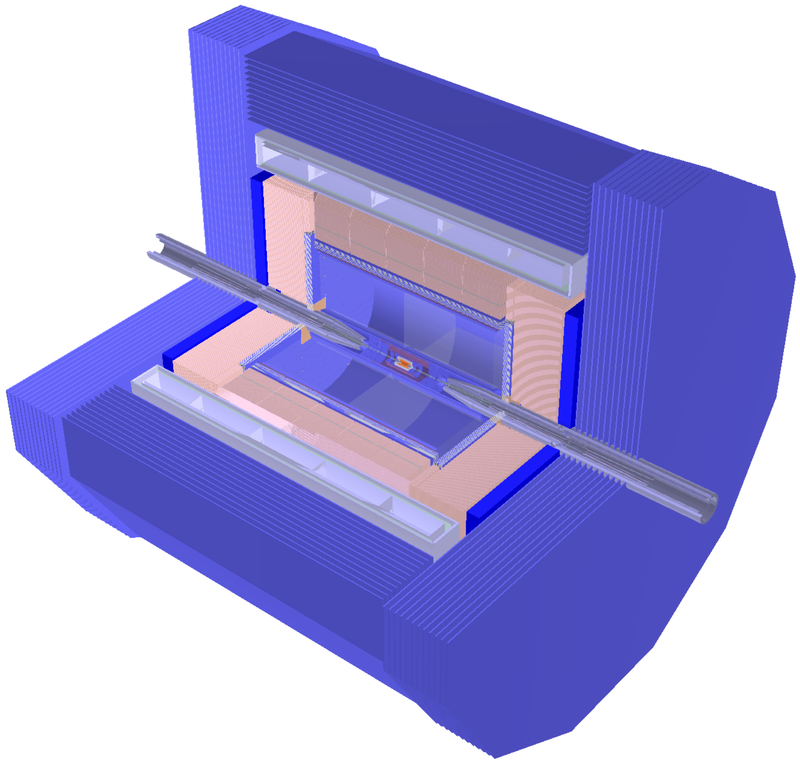
https://arxiv.org/abs/2309.13231



Yongfeng Zhu, Hao Liang, Huilin Qu, Cen Zhou, Zhenyu Zhao, Manqi RUAN, etc

Geo. & Tools

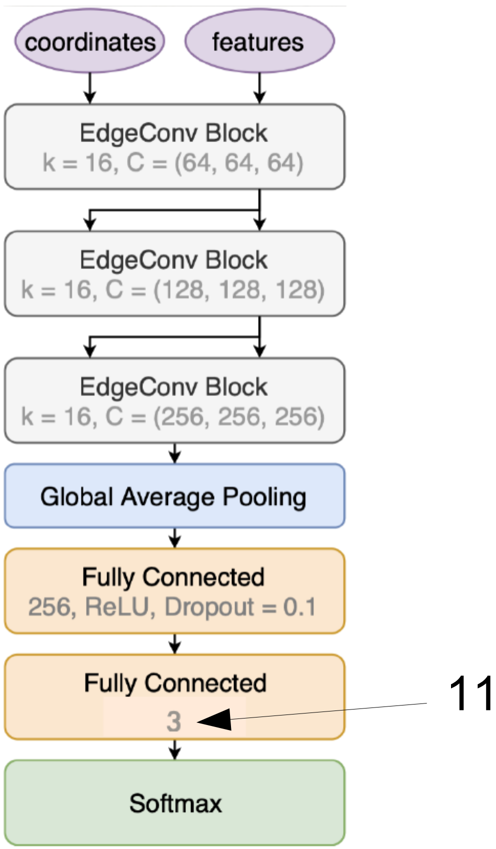


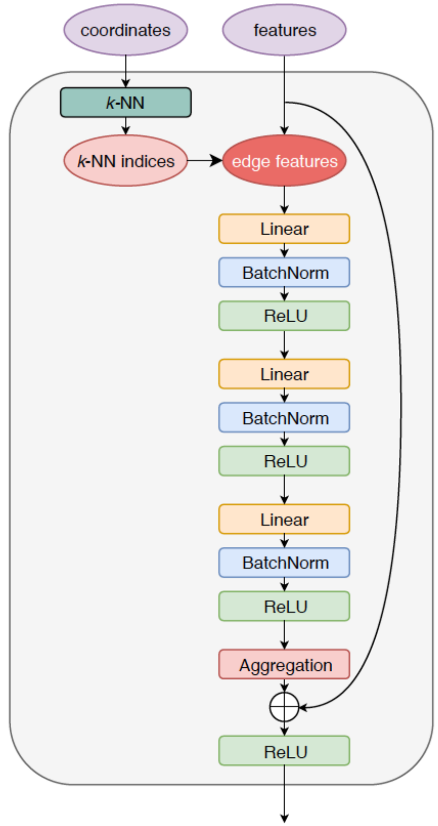


- **Jet origin identification: 11 categories (5 quarks + 5 anti quarks + gluon)**
  - Jet Flavor Tagging + Jet Charge measurements + s-tagging + gluon tagging...

- Full Simulated vvH, Higgs to two jets sample at CEPC baseline configuration: CEPC-v4 detector, reconstructed with Arbor.

Particle Net: IO





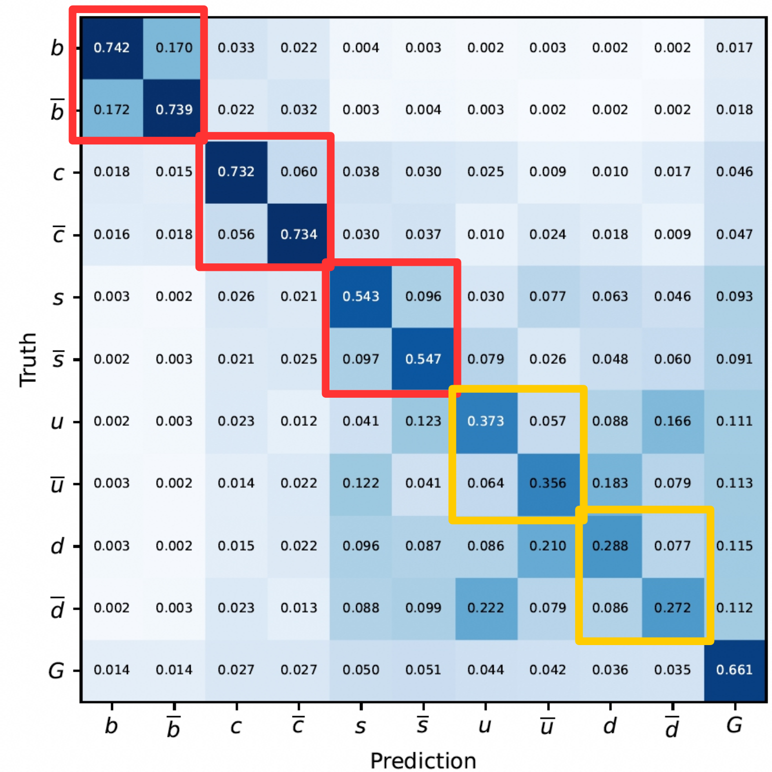
Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$log p_T$	logarithm of the particle's p_T
$log E$	logarithm of the particle's energy
$log \frac{p_T}{p_T(jet)}$	logarithm of the particle's p_T relative to the jet p_T
$log \frac{E}{E(jet)}$	logarithm of the particle's energy relative to the jet energy
ΔR	angular separation between the particle and the jet axis $(\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2})$
d0	transverse impact parameter of the track
d0err	uncertainty associated with the measurement of the d0
z0	longitudinal impact parameter of the track
z0err	uncertainty associated with the measurement of the z0
charge	electric charge of the particle
isElectron	if the particle is an electron
isMuon	if the particle is a muon
isChragedKaon	if the particle is a charged Kaon
isChragedPion	if the particle is a charged Pion
isProton	if the particle is a proton
isNeutralHadron	if the particle is a neutral hadron
isPhoton	if the particle is a photon

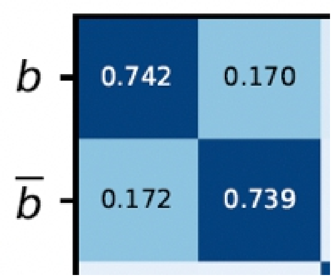
Table 3. The input variables used in ParticleNet for jet flavor tagging at the CEPC.

- Input: reco particles corresponding to 1 jet...
- Output: likelihoods to 11 different categories (sum =1)

Jet origin id: 11 categories

- vvH sample, with Higgs decays into different species of colored particle: 5 quark, 5 antiquark & gluon
  - **1 Million** of each type
  - **60/20/20%** for training, validating, and testing, result corresponding to testing sample
- Pid: ideal Pid – three scenarios
  - Lepton identification
  - + Charged hadron identification
  - + Neutral Kaons identification
- Patterns:
  - ~ Diagonal at quark sector...
  - P(g→q) < P(q→g)...
  - Light jet id...

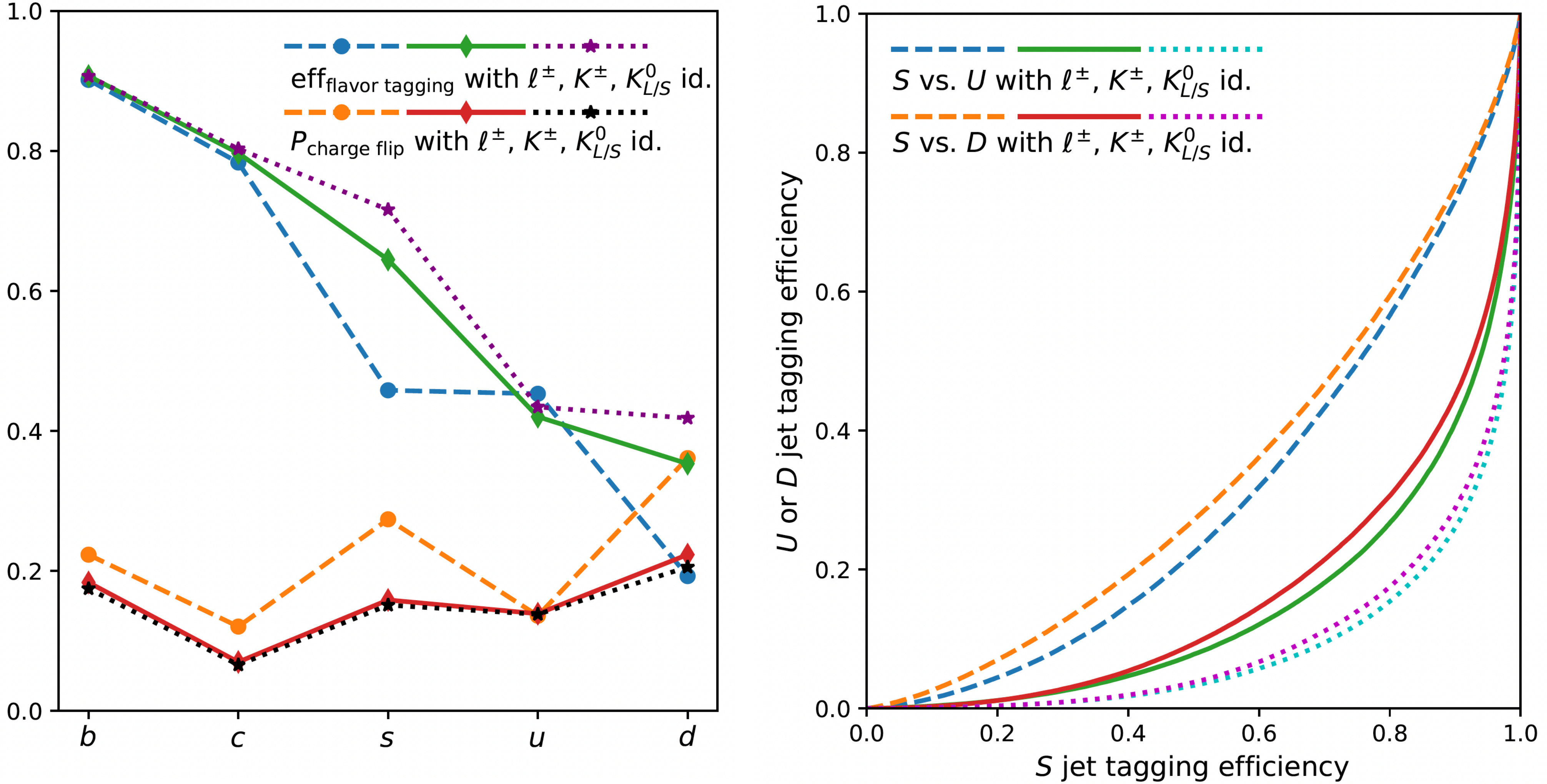




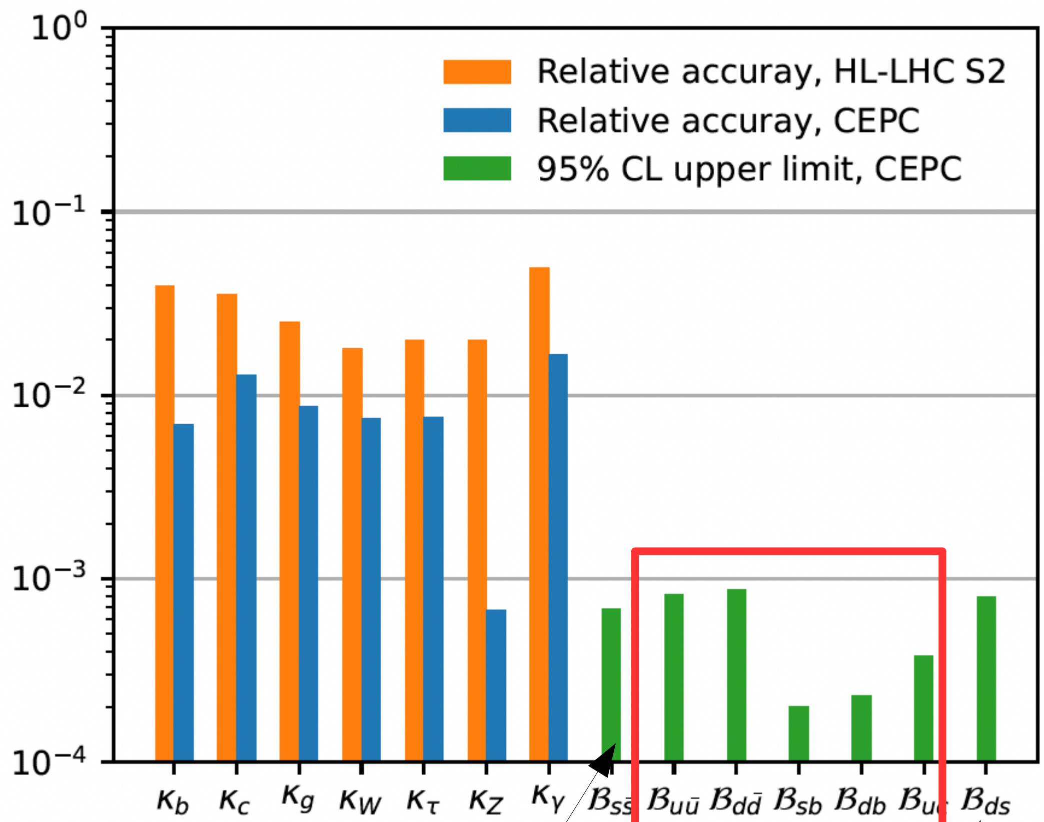
Eff = (0.74 + 0.17 + 0.74 + 0.17)/2 = 0.91

Charge flip rate = 0.17/0.91 = 0.19

Performance with different PID scenarios



Benchmark analyses using Jet origin ID



Improved by ~3 times

Improved by 1-2 orders of magnitudes

Presumably... firstly quantified

TABLE I: Summary of background events of $H \to b\bar{b}/c\bar{c}/gg$, Z, and W prior to flavor-based event selection, along with the expected upper limits on Higgs decay branching ratios at 95% CL. Expectations are derived based on the background-only hypothesis.

	Bkg. (10^3)			Upper limit (10^{-3})						
	H	Z	W	$s\bar{s}$	$u\bar{u}$	$d\bar{d}$	sb	db	uc	ds
$\nu\bar{\nu}H$	151	20	2.1	0.81	0.95	0.99	0.26	0.27	0.46	0.93
$\mu^+\mu^-H$	50	25	0	2.6	3.0	3.2	0.5	0.6	1.0	3.0
e^+e^-H	26	16	0	4.1	4.6	4.8	0.7	0.9	1.6	4.3
Comb.	-	-	-	0.75	0.91	0.95	0.22	0.23	0.39	0.86

[28] J. Duarte-Campderros, G. Perez, M. Schlaffer, and A. Soffer. Probing the Higgs–strange-quark coupling at e^+e^- colliders using light-jet flavor tagging. *Phys. Rev. D*, 101(11):115005, 2020.

[50] Alexander Albert et al. Strange quark as a probe for new physics in the Higgs sector. In *Snowmass 2021*, 3 2022.

[59] J. de Blas et al. Higgs Boson Studies at Future Particle Colliders. *JHEP*, 01:139, 2020.

[60] Jorge De Blas, Gauthier Durieux, Christophe Grojean, Jiayin Gu, and Ayan Paul. On the future of Higgs, electroweak and diboson measurements at lepton colliders. *JHEP*, 12:117, 2019.

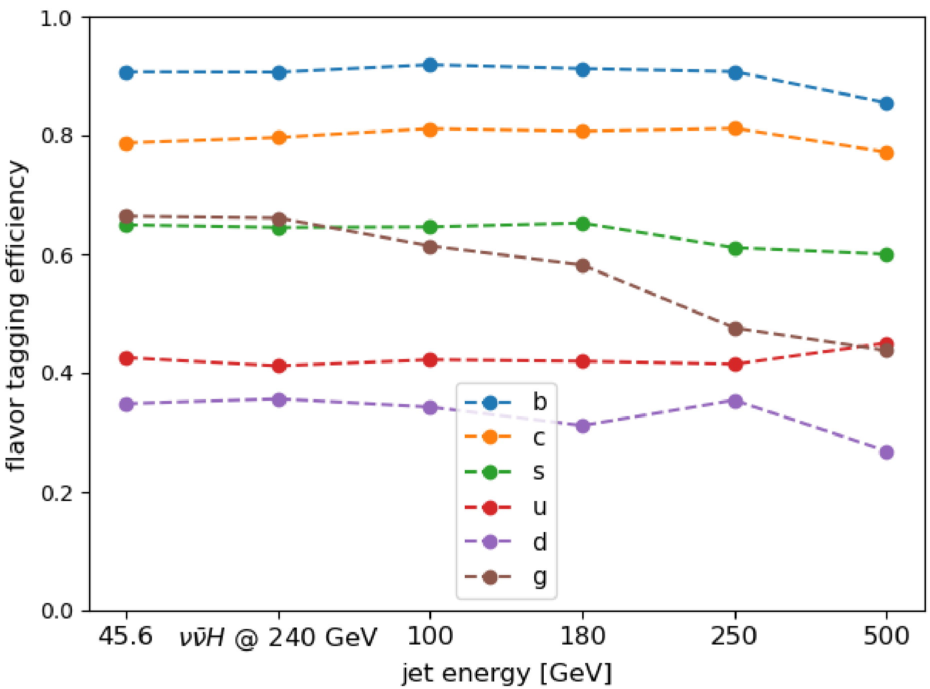
For H->bb, cc, gg: results in 20 – 40% improvement in relative accuracies (preliminary)...

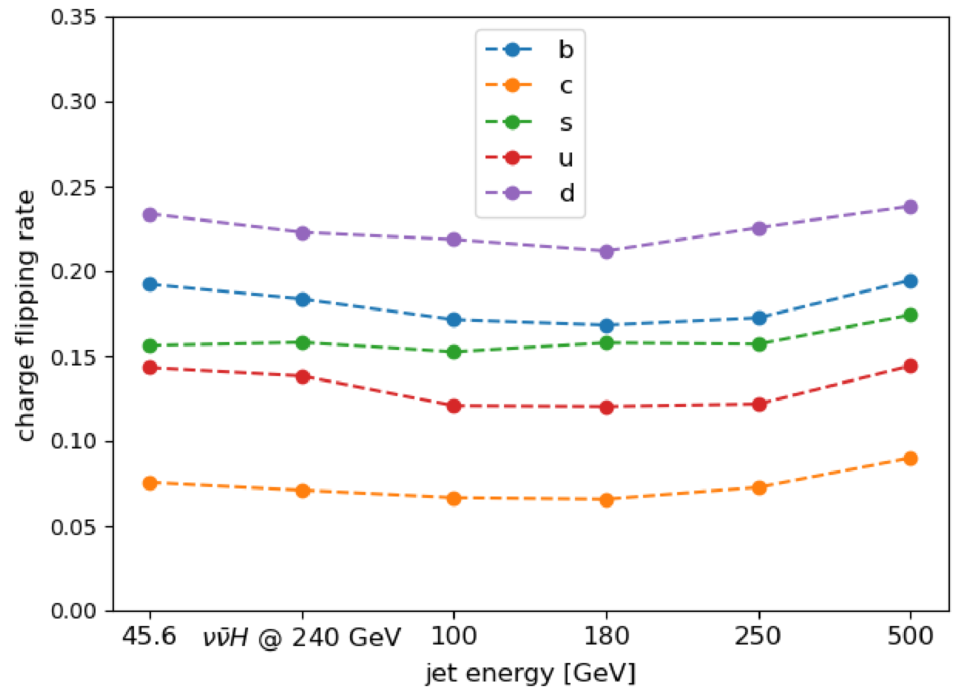
Updates

A lot to scan!!

- A lot to be understood...
  - V.S. Scaling of Jet energy, Polar angle/eta,
  - V.S. Collision environment: beam background, # PU
  - V.S. Detector geometry: VTX configuration, acceptance, etc
  - V.S. Jet Clustering algorithm, interactions with jet finding & Color Singlet identification
  - V.S. Different hadronization & fragmentation modes...
  -
  - V.S. algorithm architecture
  - V.S. training & implementation procedure...

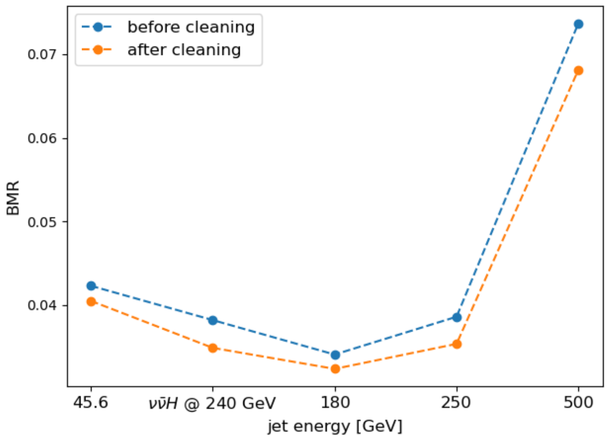
Stability V.S. Jet energy



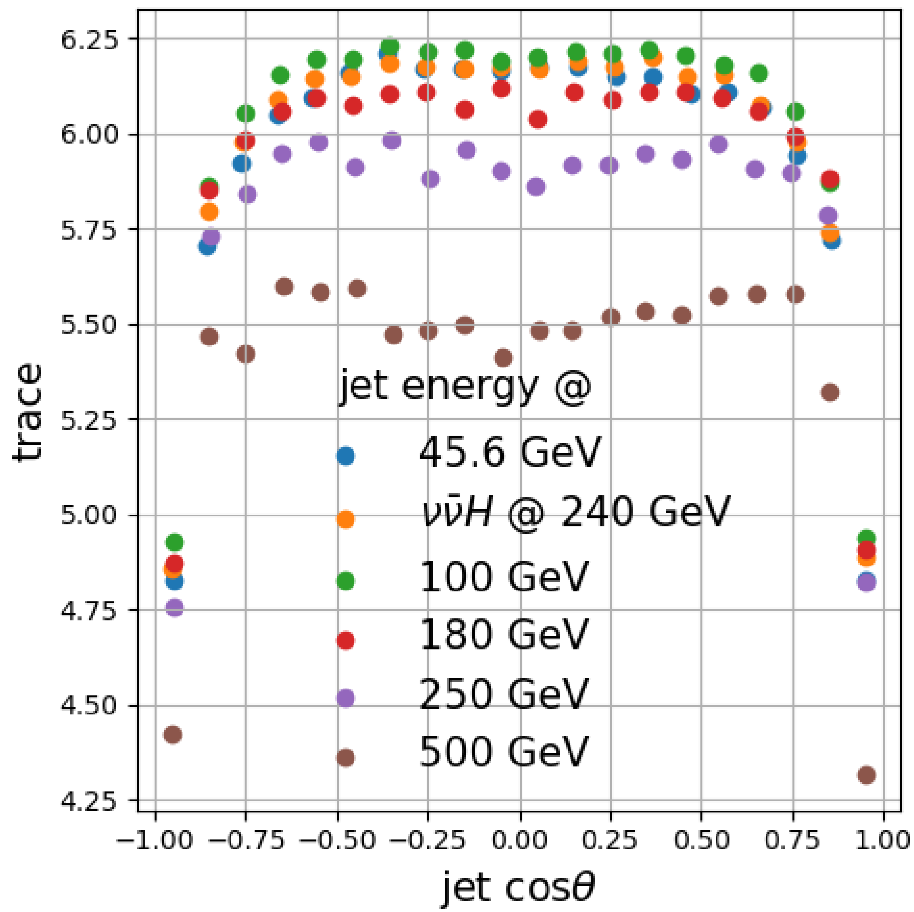


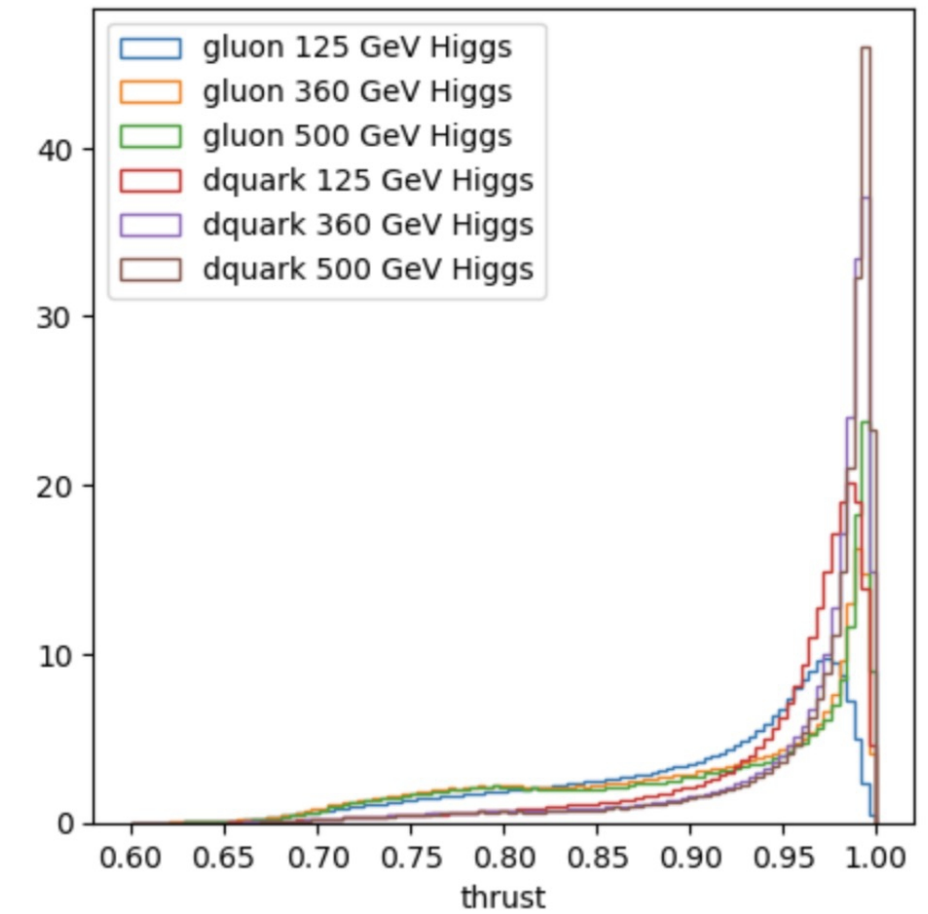
... smooth ...

... Degrading for b/g id at high energy, relevant with BMR blow up...



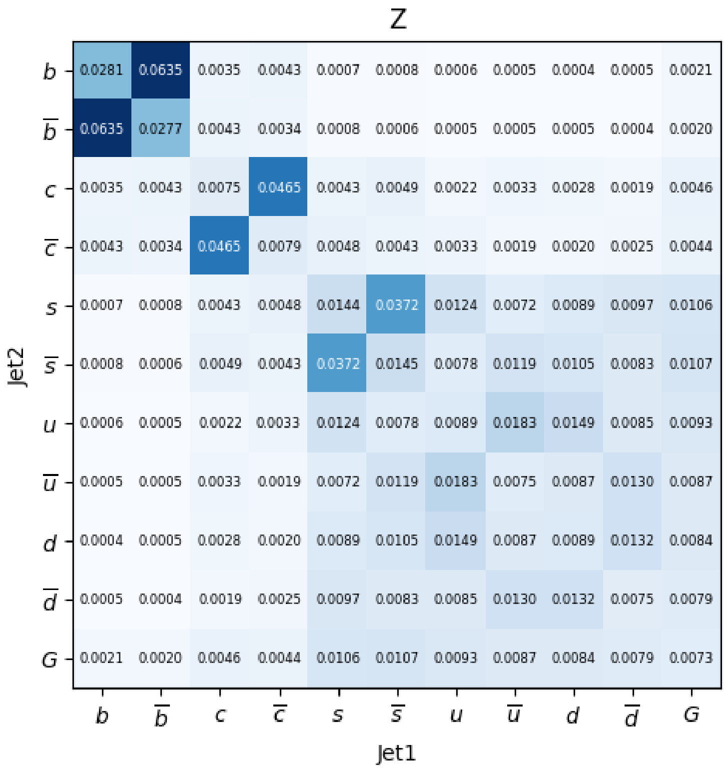
V.S. Polar angle

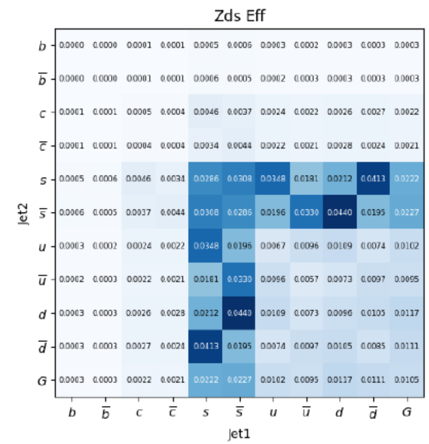


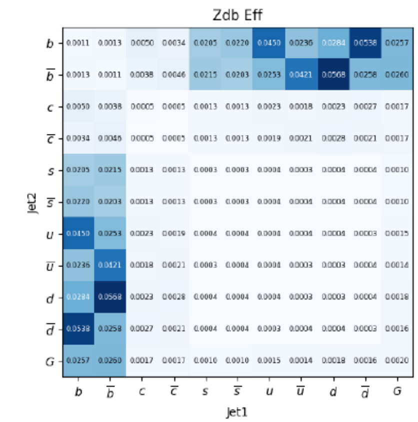


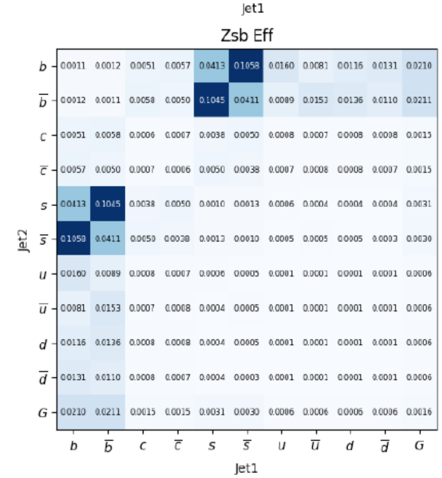
... And event topology thrust deformation ...

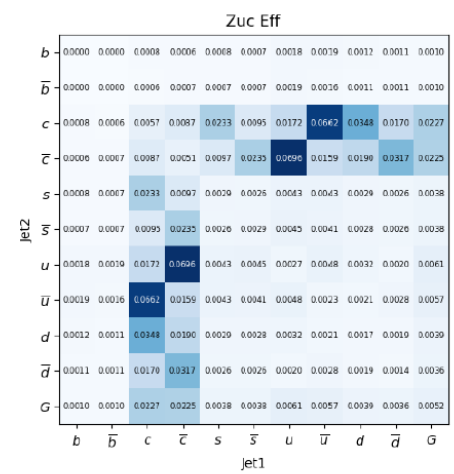
Applied to Z FCNC (Preliminary)











	SM Br	95% Upper limit on Br (statistical only)	FCC (2306.17520)
Z->bs	4.2E-8	2.3e-07	~1E-6
Z->bd	1.8E-9	2.5e-07	6E-8
Z->cu	1.4E-18	6.3e-07	4E-7
Z->sd	-	1.3e-06	-

- @ Tera Z using template fit
- Order of magnitudes ~ J.F.K
- **Calibration & Systematic control is critical**

Applied to Forward-backward...

Sensitivity and Tagging power

$$\boldsymbol{sensitivity: S = S^{phy} * Det}$$

$$S^{phy} = \frac{\partial A_{FB}^{phy}}{\partial \sin^2 \theta_{eff}}$$

$$Det = \frac{1}{1-2f} \cdot \sqrt{\frac{1}{\epsilon_{tagging}}} = \sqrt{\frac{1}{\text{tagging power}}}$$

$\epsilon_{tagging}$	overall efficiency of events observation
f	charge mis-identification probability (event-level)

tagging power: $\boldsymbol{\epsilon * (1-2f)^2}$

Lepton	Quarks
$\epsilon \sim 100\%$ $f \sim 0$	tagging power: $\epsilon * (1-2f)^2$ $= 0.088$ (for b quarks)

Previous work by Hanhua Cui *et al.*, perform jet tagging and charge measurement.

This selection is **event-level**

Estimation of tagging power

	Purity P	Efficiency ϵ	Mis-id f	Tagging power
b	**~100%**	**0.577**	**0.05**	**0.467**
c	**99%**	**0.546**	**0.00056**	**0.528**
s	**90.5%**	**0.338**	**0.086**	**0.232**
u	62.6%	0.219	0.342	0.022
d	71.4%	0.119	0.269	0.025

- With a high-purity sample, b/c channel can be used to measure $\sin^2 \theta_{eff}^{l}$.
- (Maybe) after adjusting the working point, s, even u/d channel can also be utilized.
- (Maybe) a joint measurement of u/d.

- Compared to previous study, tagging power enhanced significantly (i.e., *5 for b)

Zhao Zhenyu https://iopscience.iop.org/article/10.1088/1674-1137/acf91f

Updated result on $\sin^2\theta_{eff}^{l}$ measurement

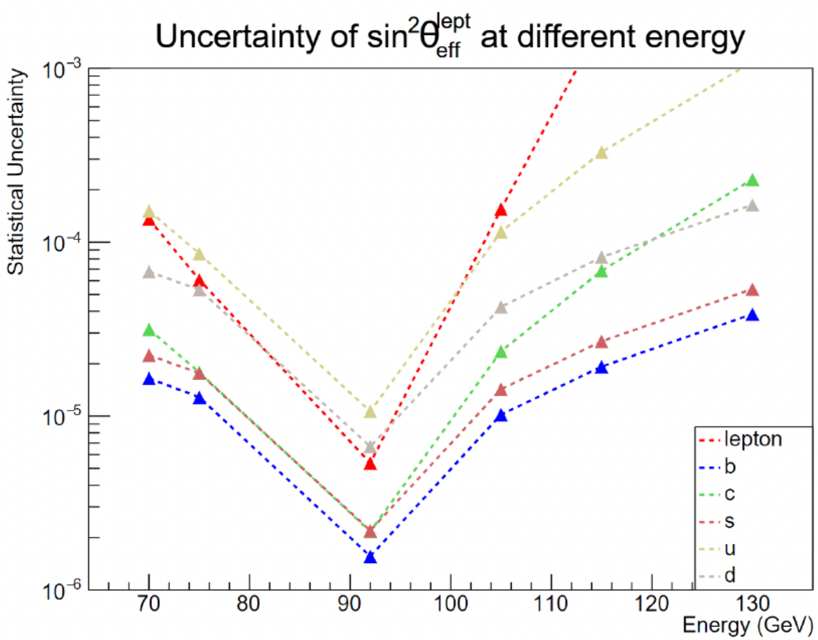
Table 2. Sensitivity S of different final state particles.

$\sqrt{s}$/GeV	S of $A_{FB}^{e/\mu}$	S of A_{FB}^{d}	S of A_{FB}^{u}	S of A_{FB}^{s}	S of A_{FB}^{c}	S of A_{FB}^{b}
70	0.224	4.396	1.435	4.403	1.445	4.352
75	0.530	5.264	2.598	5.269	2.616	5.237
92	1.644	5.553	4.200	5.553	4.201	5.549
105	0.269	4.597	1.993	4.598	1.994	4.586
115	0.035	3.956	1.091	3.958	1.087	3.942
130	0.027	3.279	0.531	3.280	0.520	3.261

Table 3. Cross section of process $e^+e^- \to f\bar{f}$ calculated using the ZFITTER package. Values of the fundamental parameters are set as $m_Z = 91.1875$ GeV, $m_t = 173.2$ GeV, $m_H = 125$ GeV, $\alpha_s = 0.118$ and $m_W = 80.38$ GeV.

$\sqrt{s}$/GeV	σ_μ/mb	σ_d/mb	σ_u/mb	σ_s/mb	σ_c/mb	σ_b/mb
70	0.039	0.032	0.066	0.031	0.058	0.028
75	0.039	0.047	0.073	0.046	0.065	0.043
92	1.196	5.366	4.228	5.366	4.222	5.268
105	0.075	0.271	0.231	0.271	0.227	0.265
115	0.042	0.135	0.122	0.135	0.118	0.132
130	0.026	0.071	0.068	0.071	0.066	0.069

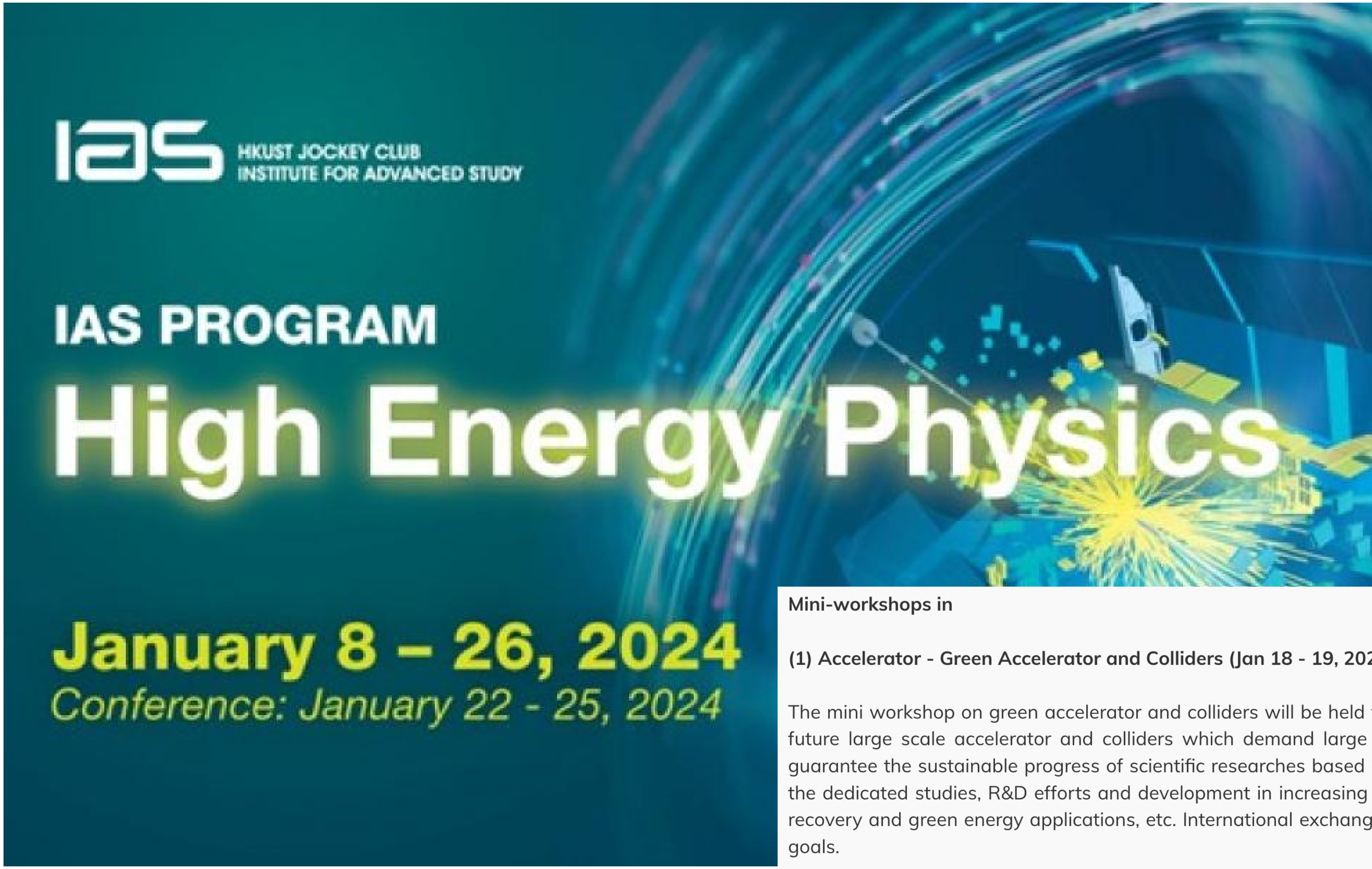
Expected statistical uncertainties on $\sin^2\theta_{eff}^{l}$ measurement.
(Using one-month data collection, ~ *4e12/24 Z events* at Z pole)



$\sqrt{s}$	b	c	s	u	d
70	1.6×10^{-5}	3.2×10^{-5}	2.2×10^{-5}	1.5×10^{-4}	6.8×10^{-5}
75	1.3×10^{-5}	1.8×10^{-5}	1.8×10^{-5}	8.6×10^{-5}	5.3×10^{-5}
92	1.6×10^{-6}	2.2×10^{-6}	2.2×10^{-6}	1.1×10^{-5}	6.7×10^{-6}
105	1.0×10^{-5}	2.4×10^{-5}	1.4×10^{-5}	1.1×10^{-4}	4.2×10^{-5}
115	1.9×10^{-5}	6.8×10^{-5}	2.7×10^{-5}	3.3×10^{-4}	8.2×10^{-5}
130	3.9×10^{-5}	2.3×10^{-4}	5.4×10^{-5}	1.1×10^{-3}	1.6×10^{-4}

Summary

- Vision (long term): **Jet origin id as Pid** + Access to g(Hss) at future Higgs factory
- Performance check/scan on going...
  - Smooth & interpretable behavior V.S. Energy & polar angle... Looks OK ...
  - A lot more to be covered... man/computing power intensive
- Multiple applications with significant impact... not surprising
- Challenges in Systematical control & Calibration...



Mini-workshops in

(1) Accelerator - Green Accelerator and Colliders (Jan 18 - 19, 2024)

The mini workshop on green accelerator and colliders will be held from Jan. 18-19, 2023 at HKUST IAS. Facing to the future large scale accelerator and colliders which demand large amount of electricity power during operations, to guarantee the sustainable progress of scientific researches based on these kinds of machines, it is vital to strengthen the dedicated studies, R&D efforts and development in increasing the efficiency of electcity consumptions, wast heat recovery and green energy applications, etc. International exchanges and collaborations are important towards these goals.

(2) Experiment and Detector - Tracking Detectors and Reconstruction for Future Colliders (Jan 17 - 19, 2024)

A Higgs factory is universally recognised as the next large future collider to be realised. Tracking detectors, ranging from vertex detectors, large gas tracker to muon detectors, play an essential role in any detector concept conceived for a Higgs factory. Information from all tracking detectors, in conjunction with information coming from calorimeters and PID, is the input to the reconstruction algorithms employed. Modern reconstruction algorithms make an extensive use of machine learning and artificial intelligence tools.

(3) Theory (Jan 15 - 16, 2024)

Conference: Jan 22 - 25, 2024

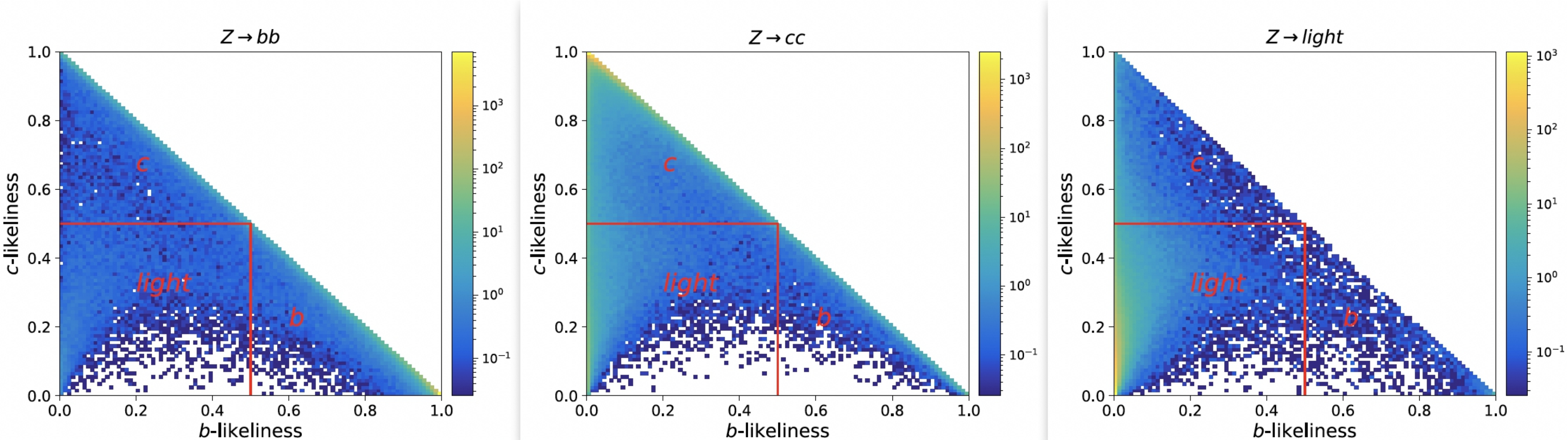
Hope to see you in person!

Backup

Summary

- PFA oriented detector design ~ CALICE laid solid foundation for the excellent reco/measurement at high energy frontier, especially with hadronic events at electron positron Higgs factories.
  - Better BMR shall always be pursues,
  - To be in cope with beam background & event rates,
  - Provide Pid: charged & even neutral hadron,
  - New AI tool... inject new momentum
  - ...
- At current baseline detector & ParticleNet, jet origin identification is possible and has encouraging performances
  - Flavor Tagging of 91%/80%/64% & Charge Flip Rate of 18%/7%/16% for b/c/s jets
  - Gluon tagging at efficiency of 67%; slight distinguish power between u & d.
  - Higgs exotic/FCNC processes with hadronic final states limited to the BRs of 1E-3 to 1E-4; $H\rightarrow ss$ limited to 3 times SM prediction (vvH + llH only)
  - *Yet, it cannot figure out some Ks decays into 2 pion...*
- Vision (long term): **Jet origin id as Pid** + Access to g(Hss) at future Higgs factory

Three categories: b, c, & light



Hadronic Z pole sample

1 M Z→bb, cc, (uds) each

60/20/20% for

training/validating/testing.

Result on Testing sample

ParticleNet (left):

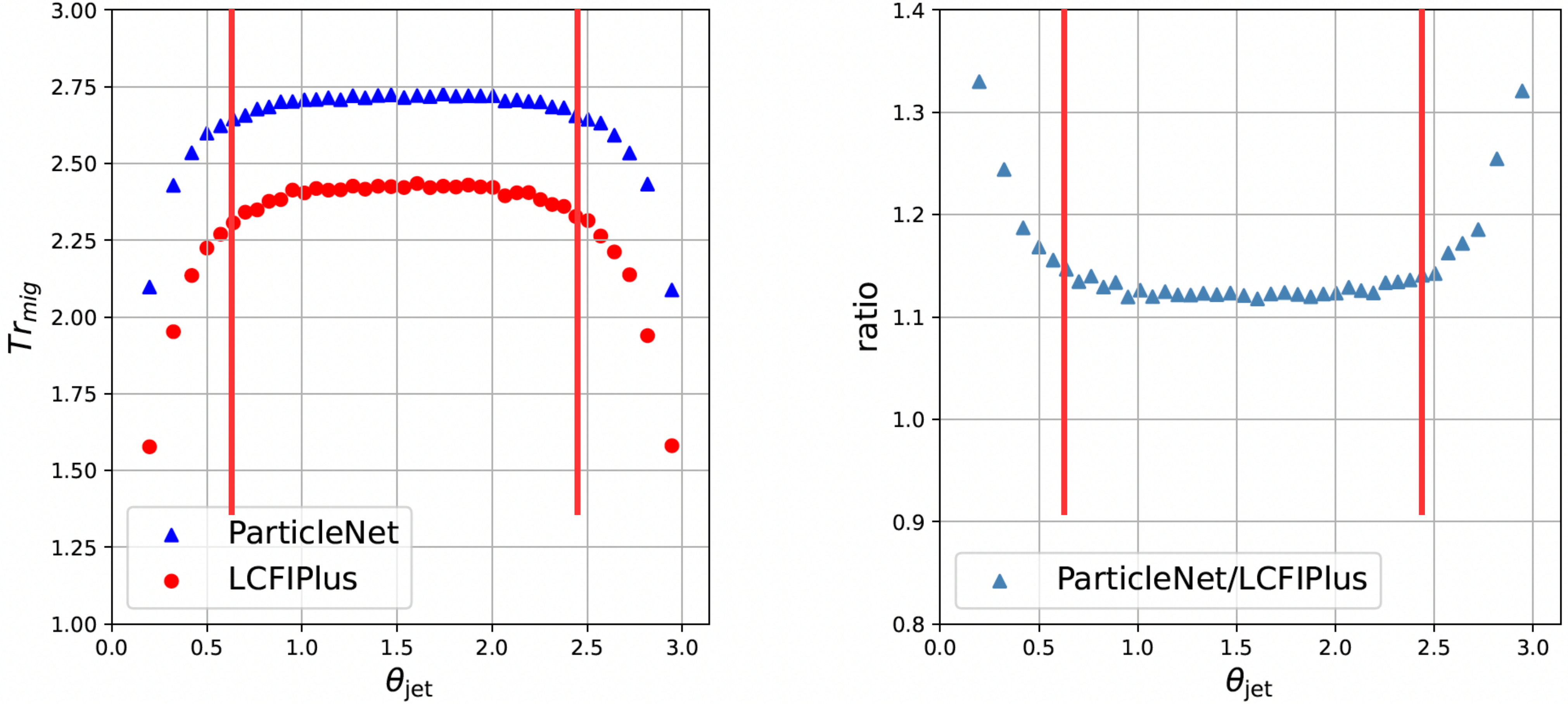
truth \ predicted	b	c	uds
b	0.911	0.059	0.031
c	0.039	0.784	0.177
uds	0.005	0.051	0.944

LCFIPlus (right):

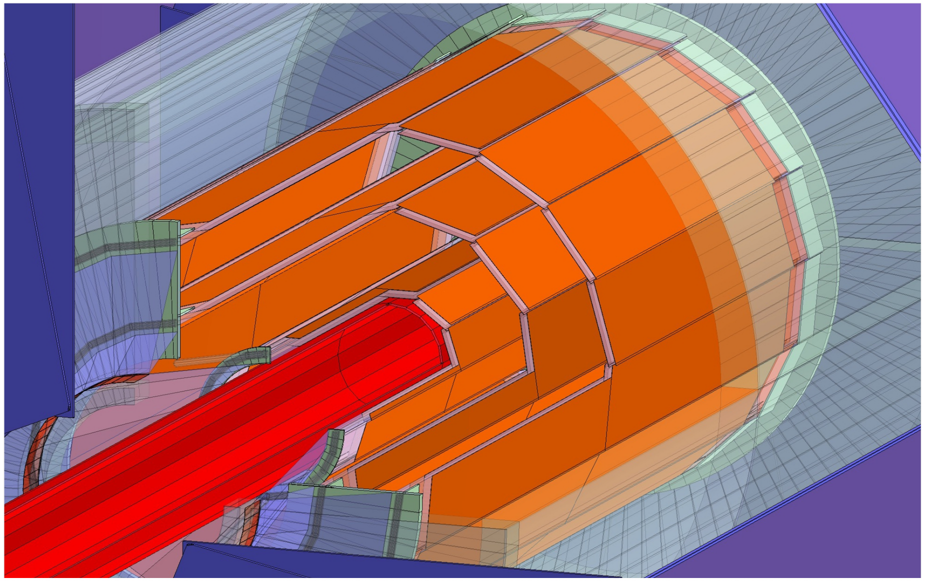
truth \ predicted	b	c	uds
b	0.789	0.126	0.085
c	0.084	0.582	0.334
uds	0.008	0.06	0.933

Figure 7. The migration matrix of ParticleNet (left) and LCFIPlus (right) at the CEPC.

Dependence on polar angle

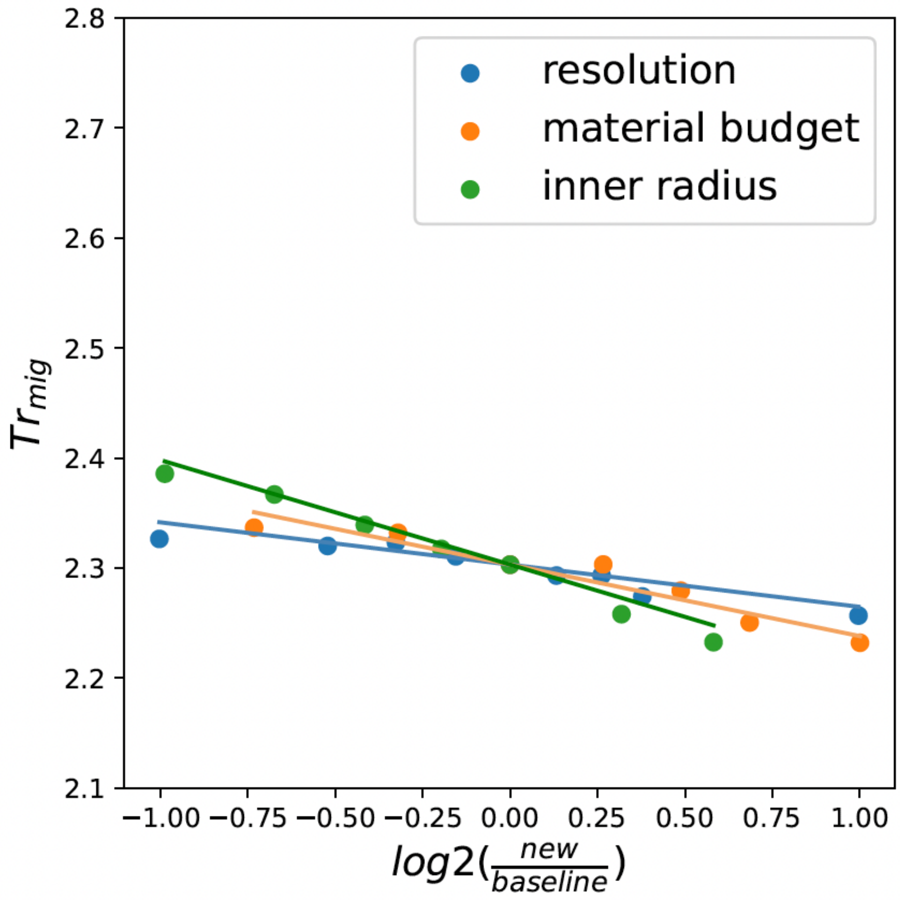


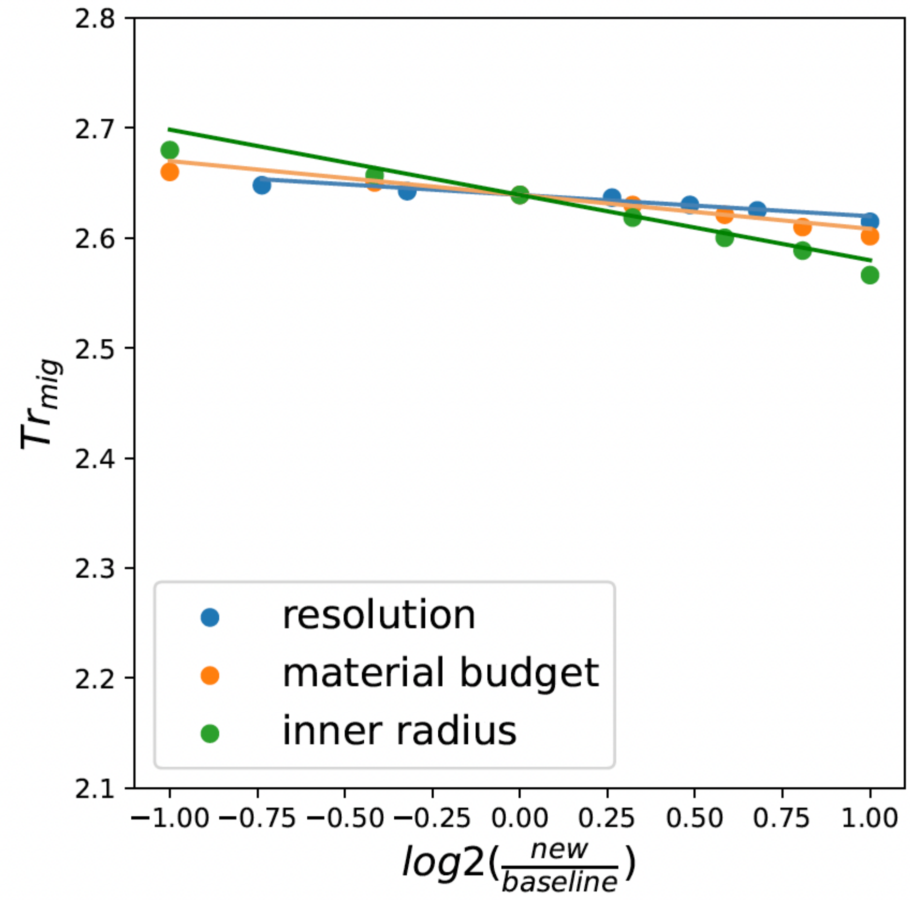
Comparison on Det. Optimization



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

Comparison on Det. Optimization

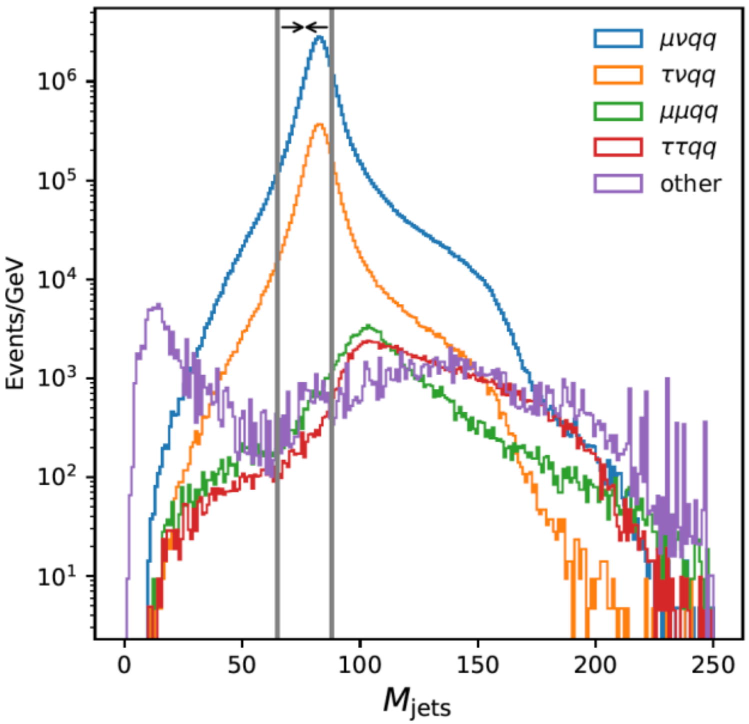




$$Tr_{mig} = 2.30 + 0.06 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2 \frac{R^0_{radius}}{R_{radius}} \quad (4.1)$$

$$Tr_{mig} = 2.64 + 0.03 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.02 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.06 \cdot log_2 \frac{R^0_{radius}}{R_{radius}} \quad (4.2)$$

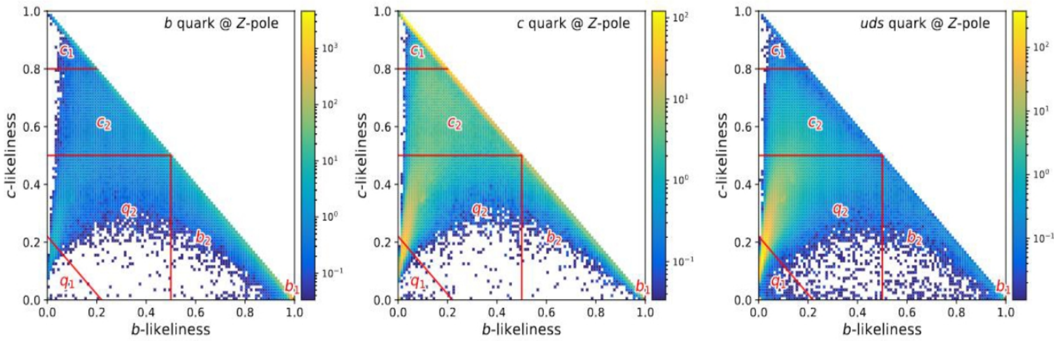
Vcb from W decay



	$\mu\nu W, W \to$				$\tau(\mu\nu)\nu_\tau W, W \to$				$\tau\nu_\tau qq, \tau \to$					
	cb	ub	$c(d/s)$	$u(d/s)$	cb	ub	$c(d/s)$	$u(d/s)$	$e2\nu$	had.ν_τ	$\tau\tau qq$	$\mu\mu qq$	Higgs	others
w/o slections	40.3K	363	24.2M	24.2M	7.73K	74	4.2M	4.2M	8.66M	31.4M	2.18M	4.47M	4.07M	2.06G
$E_{\text{L}\mu} > 12\text{GeV}$	37.9K	330	22.6M	22.6M	5.59K	56	2.98M	2.97M	133K	687K	422K	2.82M	645K	186.3M
$R_{\text{L}\mu} > 0.85$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$\cos(\theta_{\text{L}\mu})$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$q_{\text{L}\mu} \cos(\theta_{\text{L}\mu}) < 0.20$	32.8K	283	19.6M	19.6M	4.7K	42	2.57M	2.57M	1.26K	39.9K	156K	1.03M	183K	92.6M
2nd isolation ℓ veto	32.8K	283	19.5M	19.6M	4.7K	42	2.57M	2.57M	1.26K	39.9K	154K	526K	138K	43.9M
multiplicity ≥ 15	32.8K	283	19.5M	19.4M	4.7K	42	2.56M	2.55M	1.23K	39.6K	153K	522K	118K	185K
Missing $P_T > 9.5$ GeV/c	31.5K	264	18.7M	18.6M	4.38K	37	2.4M	2.39M	1.18K	37.2K	136K	118K	92.6K	97.7K
$M_{\text{jets}} > 65$ GeV/c^2	29.4K	254	18.1M	18.3M	4.15K	32	2.33M	2.35M	978	36.0K	132K	112K	85.3K	24.5K
$M_{\text{jets}} < 88$ GeV/c^2	24.1K	193	14.3M	14.1M	3.49K	23	1.87M	1.85M	641	24.7K	5.62K	11.5K	6.76K	4.31K
$M_{\text{jets,recoil}} < 115$ GeV/c^2	20.2K	184	13.0M	13.1M	2.96K	23	1.72M	1.73M	505	22.6K	3.57K	6.86K	536	3.02K
$M_{\text{L}\mu\text{S}\mu} < 75$ GeV/c^2	19.6K	184	12.9M	13.0M	2.95K	23	1.72M	1.73M	505	22.6K	3.56K	5.78K	414	3.0K
$M_{\ell\nu} > 12$ GeV/c^2	19.6K	184	12.9M	13.0M	2.7K	18	1.54M	1.55M	416	19.5K	2.08K	5.16K	390	1.81K
ϵ_{kin} (%)	48.8 (0.7)	50.6 (8.1)	53.5 (0.0)	53.7 (0.0)	34.9 (1.5)	25.0 (12.5)	36.7 (0.1)	36.9 (0.1)	0.0 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.0 (0.0)	0.0 (0.0)
$b_1 c_{1,2}$	5.14K	4	2.79K	571	632	0	407	65	0	14	67	228	0	0
$\epsilon_{b_1 c_{1,2}}$ (%)	12.8 (0.4)	1.3 (1.3)	0.0 (0.0)	0.0 (0.0)	8.2 (0.7)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

- Purity > 99.5% at Eff. 50% for $\mu\nu qq$ and 34% for $\tau(\mu 2\nu)\nu qq$
- Main backgrounds include:
  - $W \to c(d/s)$
  - $\mu\mu qq$

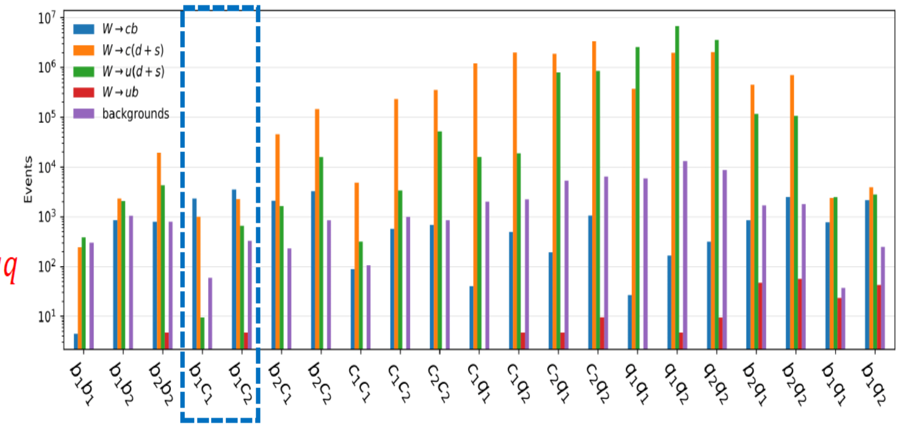
Vcb from W decay



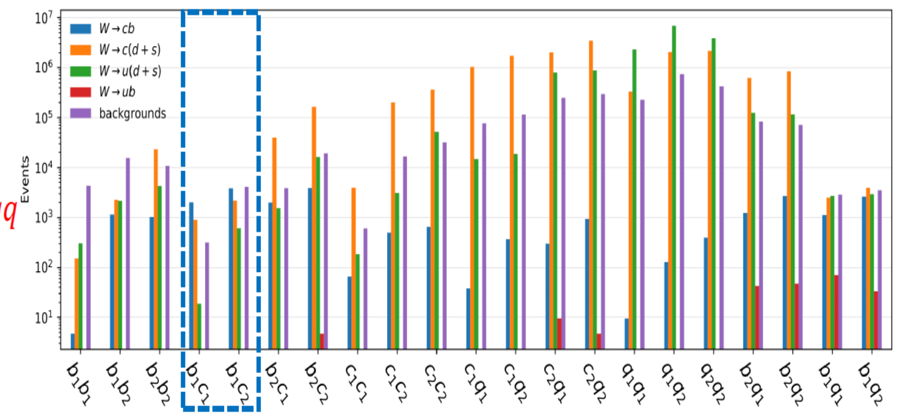
quark \ tag	b_1	b_2	c_1	c_2	q_1	q_2
b	0.47	0.378	0.0197	0.0965	0.00397	0.0315
c	0.00042	0.078	0.298	0.373	0.0682	0.182
uds	0.000104	0.00477	0.00145	0.054	0.538	0.401

- $\mu\nu qq$
  - Statistical (relative) error: 1.5%, 3.4E-4, 3.4E-4
  - $|V_{cb}|$ Statistical error: 0.75%
- $e\nu qq$
  - statistical (relative) error: 1.7%, 3.7E-4, 3.7E-4
  - $|V_{cb}|$ Statistical error: 0.85%

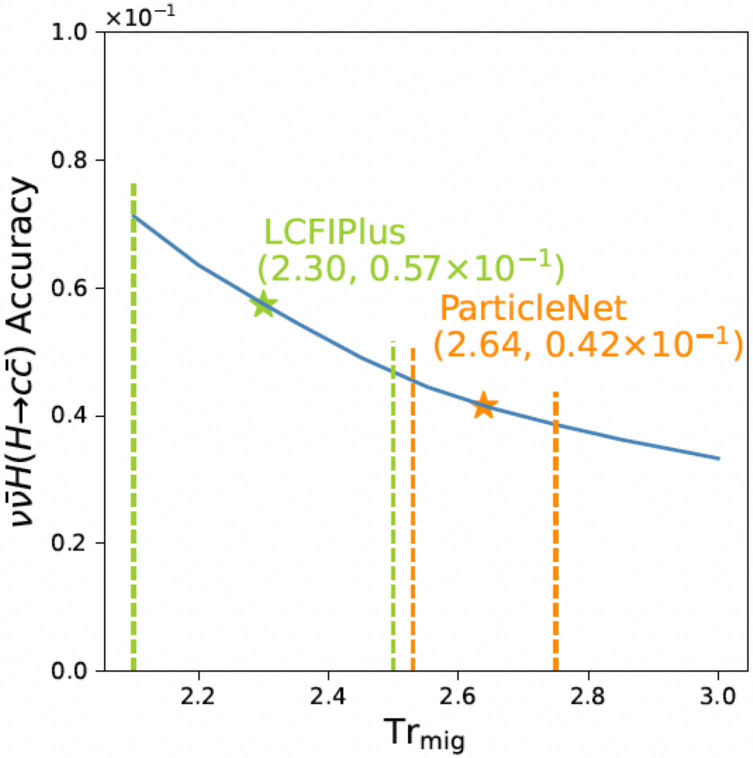
$\mu\nu qq$

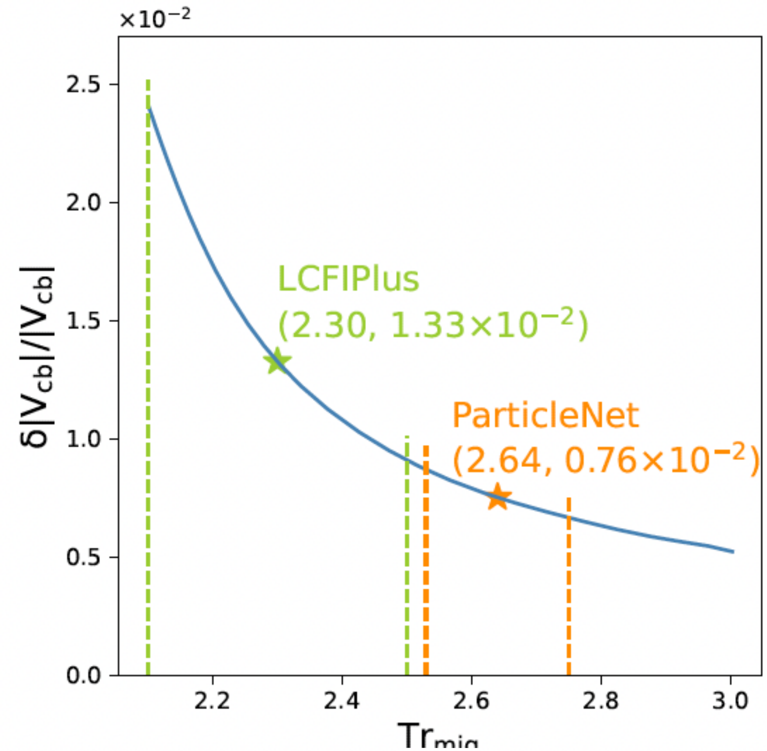


$e\nu qq$



Impact on physics benchmarks





Conservative/Aggressive:

all three parameters 2/0.5*Baseline

		conservative	baseline	optimal
$\nu\nu Hc\bar{c}$	LCFIPlus	0.071	0.057	0.047
	ParticleNet	0.045	0.042	0.038
	$\frac{\text{LCFIPlus}}{\text{ParticleNet}}$	1.58	1.38	1.26
$\lvert V_{cb}\rvert$	LCFIPlus	0.0241	0.0133	0.0091
	ParticleNet	0.0086	0.0076	0.0067
	$\frac{\text{LCFIPlus}}{\text{ParticleNet}}$	2.80	1.75	1.36