Progress on Jet Origin Identification

arxiv: 2310.03440

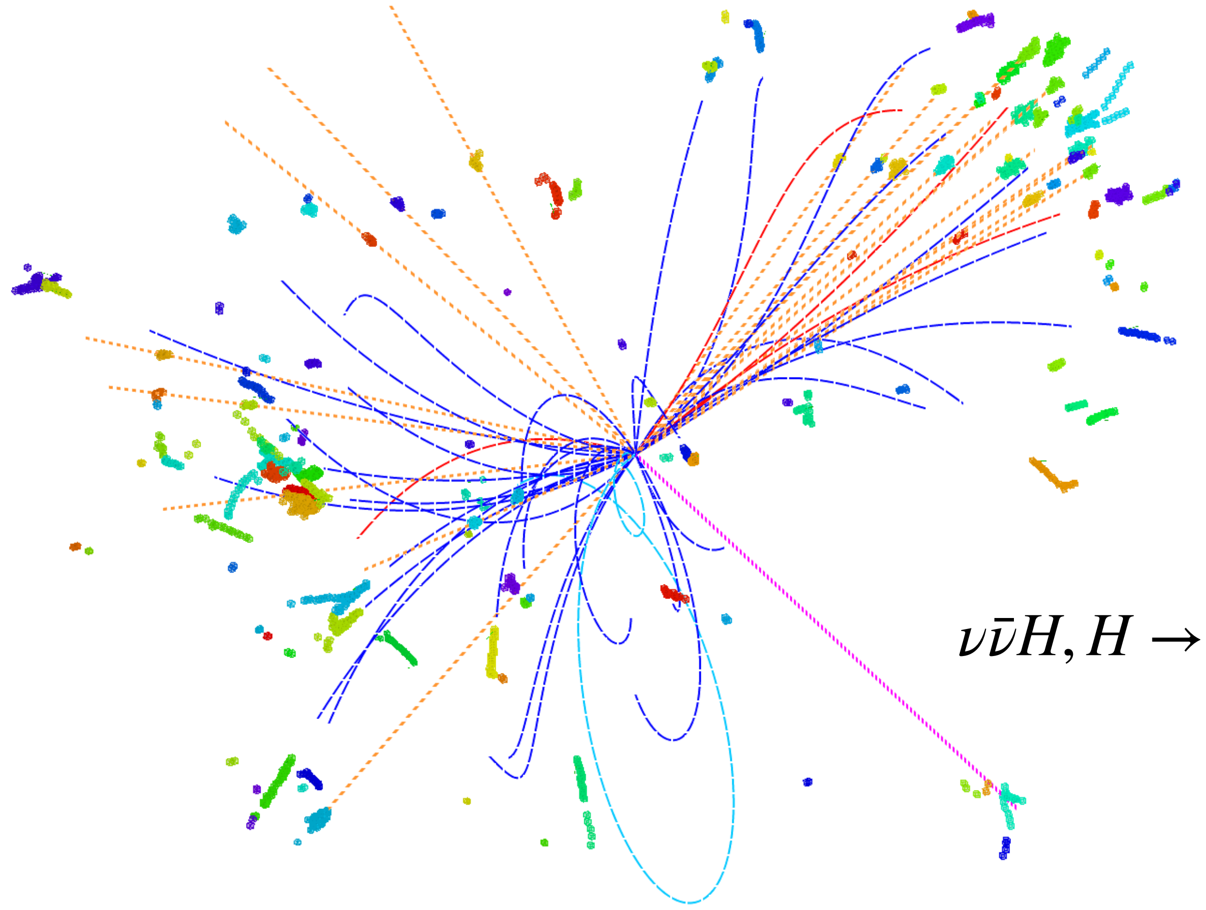
Yongfeng Zhu (PKU), Hao Liang (IHEP), Yuexin Wang (IHEP), Yuzhi Che (IHEP), Chen Zhou (PKU), Huilin Qu (CERN), Manqi Ruan (IHEP)

IAS Program on High Energy Physics (HEP 2024)

1,18,2024

Definition of Jet Origin Identification

Quarks and gluons carry color charge, and they can not travel freely.
Once generated in high-energy collisions, quarks, and gluon would fragment into numerous particles, which are called jet.



$\nu\bar{\nu}H, H \rightarrow gg$ at $\sqrt{s} = 240\ GeV$

Jet Origin Identification: categorizes jets into 5 quarks (b, c, s, u, d), 5 anti-quark ($\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d}$), and gluon

motivation

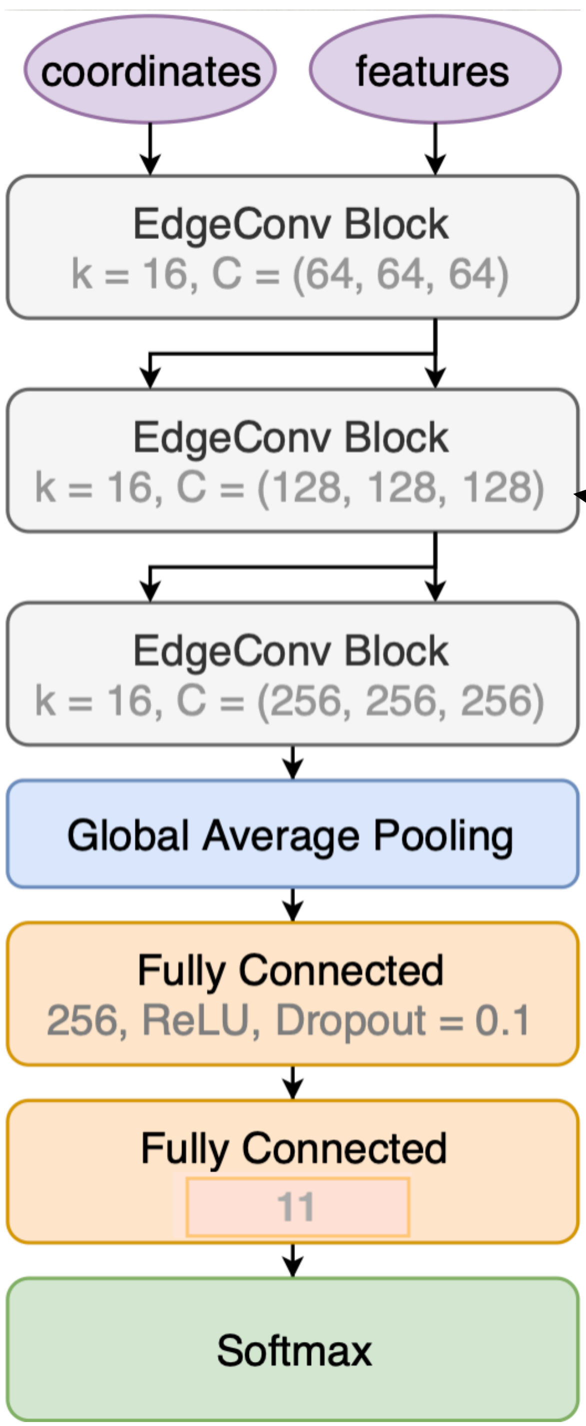
- A successful jet origin identification is critical for experimental particle physics at the energy frontier.
- The jet flavor tagging is essential for the Higgs property measurements.
- The determination of jet charge is essential for weak mixing angle measurement, is critical for the time-dependent CP measurements, and could have a significant impact on Higgs property measurements.

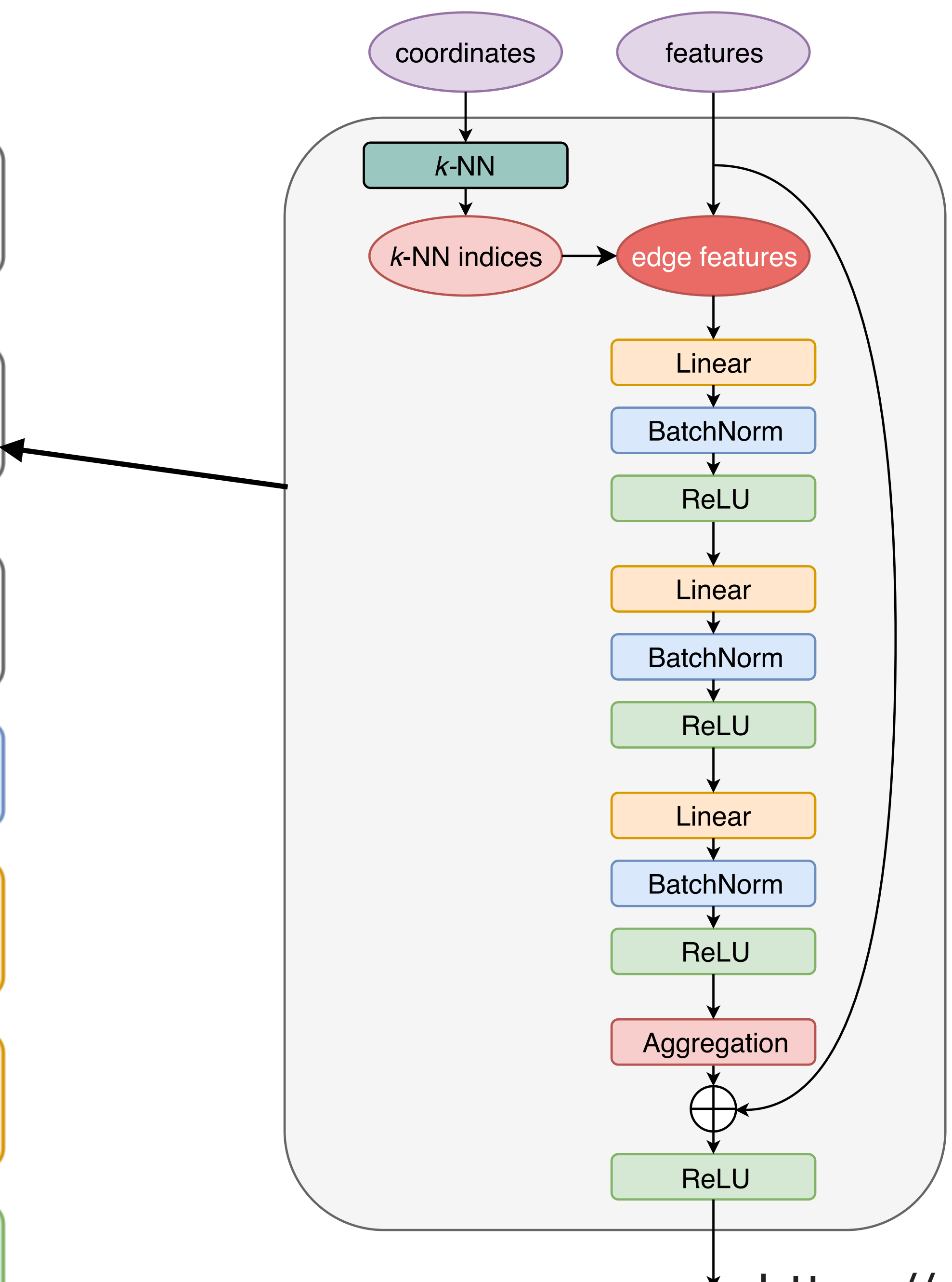
Contents

- samples and ParticleNet
- the performance of jet origin identification
- Benchmark physics analyses
- the dependence of jet origin ID on the jet energy and jet polar angle
- comparison between different physics processes
- comparison between different hadronization models

Samples and ParticleNet

samples: Based on the CEPC baseline detector, full simulation of $e^+e^- \rightarrow \nu\bar{\nu}H, H \rightarrow b\bar{b}/c\bar{c}/s\bar{s}/u\bar{u}/d\bar{d}/gg$ at $\sqrt{s} = 240\ GeV$ generated with Pythia-6.4





input features:
impact parameters, kinematic variables, and PID information of particles in a jet

https://doi.org/10.1103/PhysRevD.101.056019

the performance of jet origin identification

ParticleNet algorithm attaches each jet with 11 likelihoods corresponding to 11 types of jets. Then the jet type is determined according to the maximum likelihood.

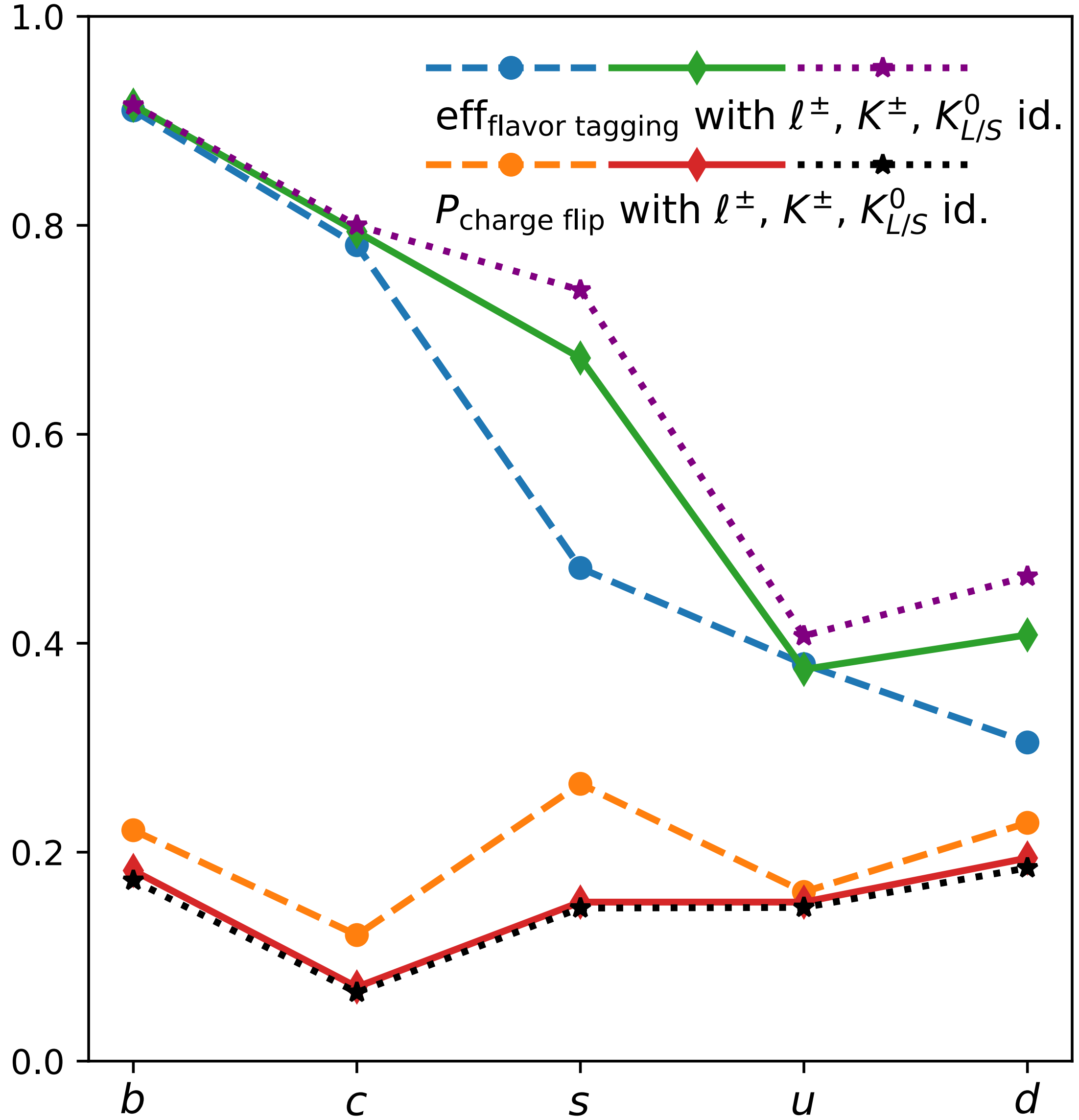
True \ Predicted	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.745	0.163	0.033	0.025	0.004	0.003	0.002	0.003	0.002	0.002	0.017
$\bar{b}$	0.170	0.737	0.026	0.033	0.003	0.004	0.003	0.002	0.002	0.003	0.018
c	0.015	0.014	0.743	0.055	0.036	0.031	0.025	0.009	0.009	0.018	0.043
$\bar{c}$	0.016	0.015	0.056	0.739	0.032	0.037	0.009	0.026	0.017	0.010	0.043
s	0.003	0.002	0.020	0.018	0.543	0.102	0.030	0.080	0.063	0.045	0.092
$\bar{s}$	0.003	0.003	0.018	0.020	0.102	0.542	0.084	0.028	0.045	0.062	0.094
u	0.002	0.003	0.020	0.011	0.044	0.131	0.367	0.055	0.080	0.174	0.111
$\bar{u}$	0.003	0.003	0.011	0.019	0.132	0.043	0.062	0.356	0.178	0.081	0.111
d	0.003	0.003	0.012	0.019	0.112	0.092	0.082	0.207	0.277	0.079	0.112
$\bar{d}$	0.003	0.003	0.020	0.012	0.092	0.112	0.219	0.076	0.079	0.272	0.113
G	0.015	0.014	0.024	0.024	0.052	0.052	0.043	0.041	0.034	0.034	0.667

M11 matrix

the performance of jet origin identification

jet flavor is defined as $max(b+\bar{b},\ c+\bar{c},\ s+\bar{s},\ u+\bar{u},\ d+\bar{d},\ g)$

jet charge is assigned by comparing the quark and anti-quark likelihoods of the corresponding flavor



To understand the impact of particle identification, three scenarios are compared.

1, assumes perfect identification of charged leptons ($\ell^\pm$)

2, further assumes perfect identification of the charged hadrons ($K^\pm$)

3, on top of the second scenario, assumes perfect identification of K_L and K_S.

default scenario: 2 scenario, based on:

Eur. Phys. J. C 80, 7 (2020)

Journal of Instrumentation 16, P06013 (2021)

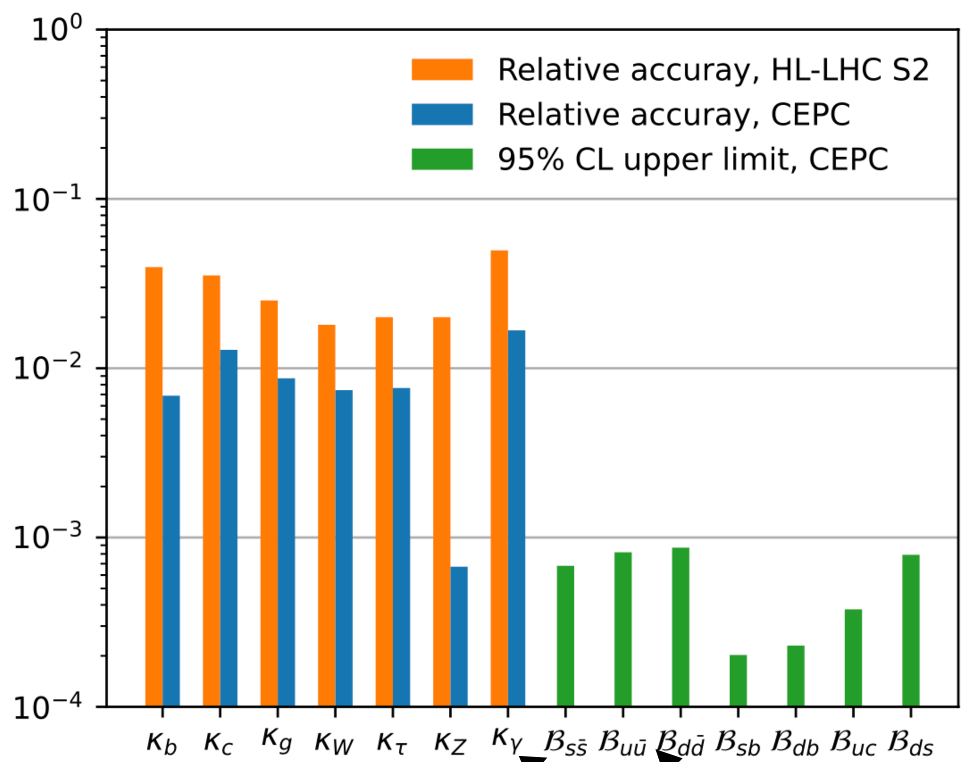
Eur. Phys. J. C 78, 464 (2018)

Eur. Phys. J. C 83, 93 (2023)

Nucl. Instrum. Meth. A 1047, 167835 (2023)

Benchmark physics analyses

Begin with the existing analyses of $\nu\bar{\nu}H, H \rightarrow b\bar{b}/c\bar{c}/gg$, (arXiv:2203.01469) and combining the jet origin identification, we obtain the upper limits on branching ratios of seven Higgs rare and FCNC hadronic decay modes.



improved by 2 times **improved by one order of magitude**

FIG. 5: The upper limits on the branching ratios of Higgs rare decays from this work are shown in green. The relative uncertainties of Higgs couplings anticipated at the CEPC [19] are shown in blue, and those at HL-LHC [42] are shown in orange, both with the kappa-0 fit scenario [53] and scenario S2 of systematics [54], as cited in Ref. [19].

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

	Bkg. (10^3)			Upper limits on Br. (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

[23] J. Duarte-Campderros, G. Perez, M. Schlaffer, and A. Soffer, Phys. Rev. D **101**, 115005 (2020), arXiv:1811.09636 [hep-ph].

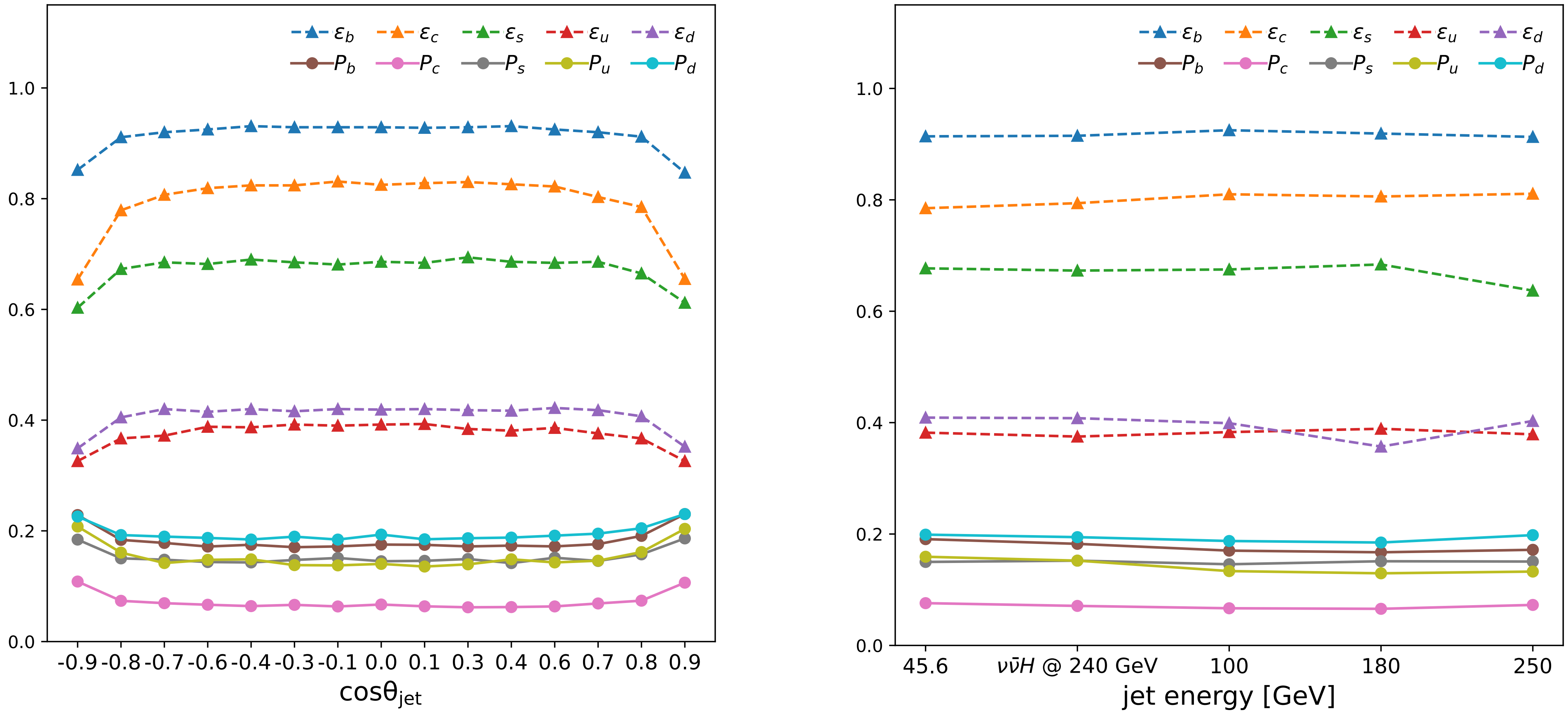
[44] A. Albert *et al.*, "Strange quark as a probe for new physics in the higgs sector," (2022), arXiv:2203.07535 [hep-ex].

[53] J. de Blas *et al.*, JHEP **01**, 139 (2020), arXiv:1905.03764 [hep-ph].

[54] J. De Blas, G. Durieux, C. Grojean, J. Gu, and A. Paul, JHEP **12**, 117 (2019), arXiv:1907.04311 [hep-ph].

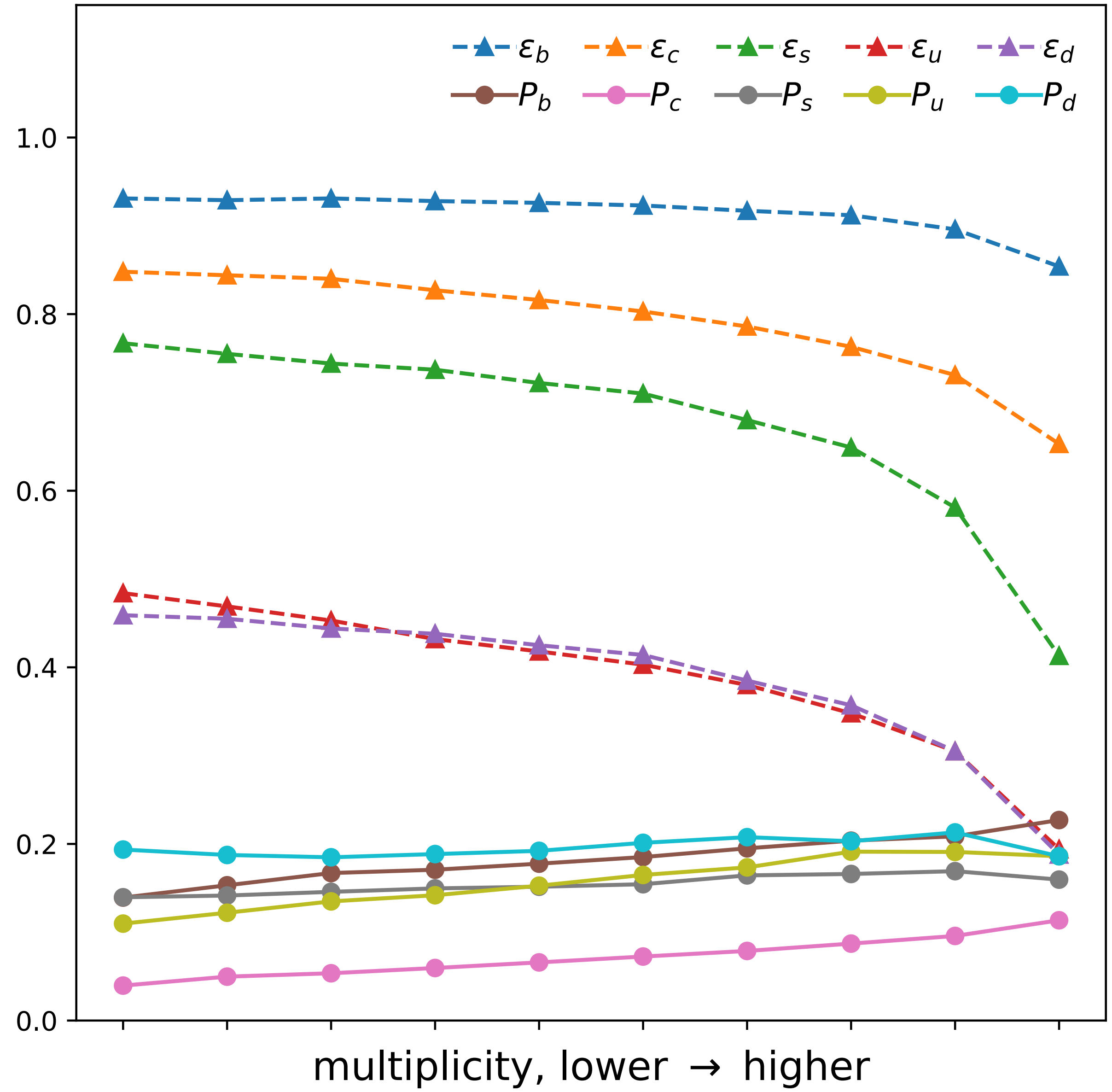
Many uncertainties are relevant to the performance of jet origin identification, including the detector performance, <u>the jet kinematics</u>, the jet clustering algorithms, <u>the physics processes</u>, and <u>the hadronization models</u>.

Dependence on the jet energy, and jet polar angle



- With $\nu\bar{\nu}H, H \to jj$ at $\sqrt{s} = 240\ GeV$, the jet tagging efficiencies and charge flip rates are flat in the barrel region of the detector and exhibit slight degrading in the endcap region.
- The jet tagging efficiencies and charge flip rates are rather stable with various jet energy.

Dependence on the particle multiplicity of the jet

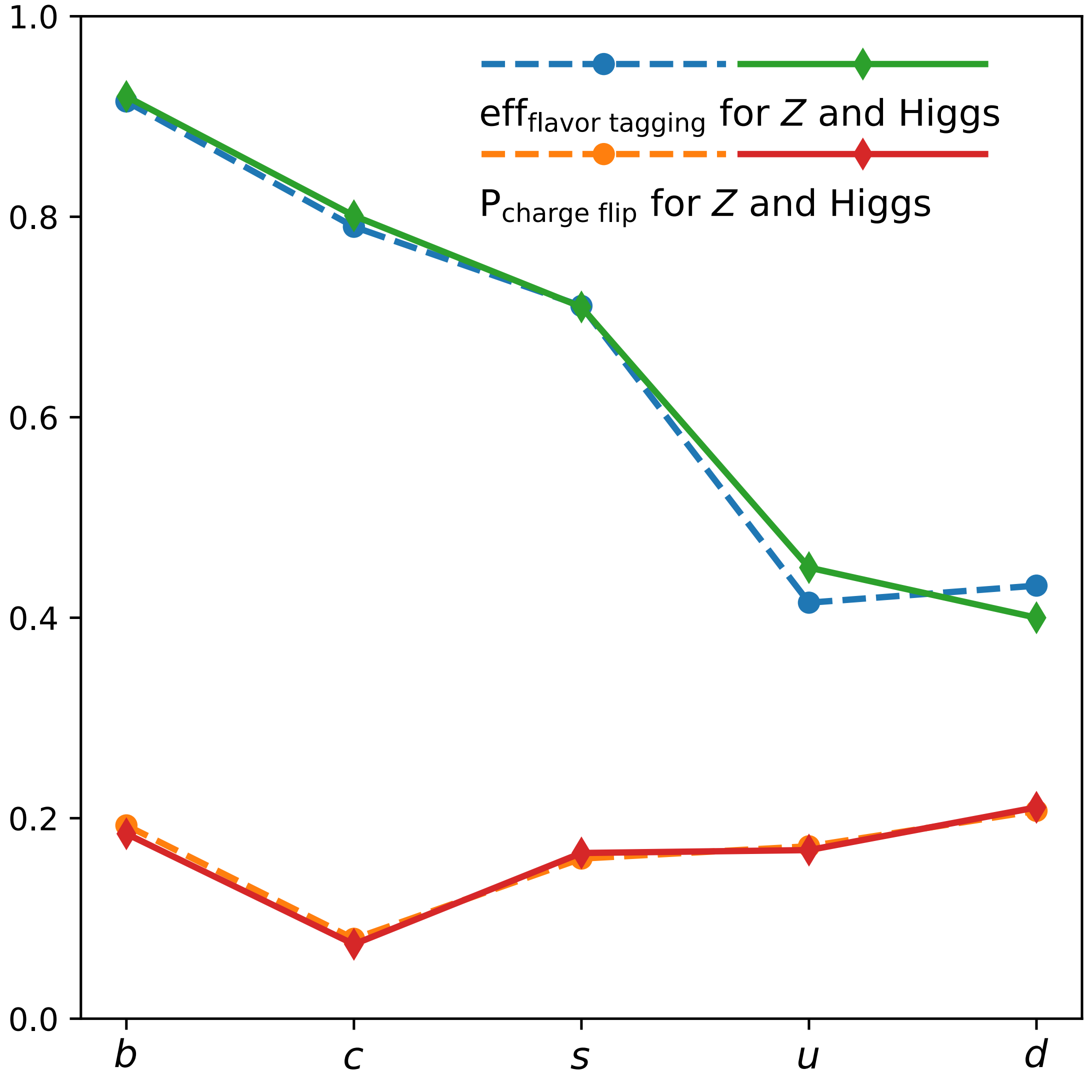


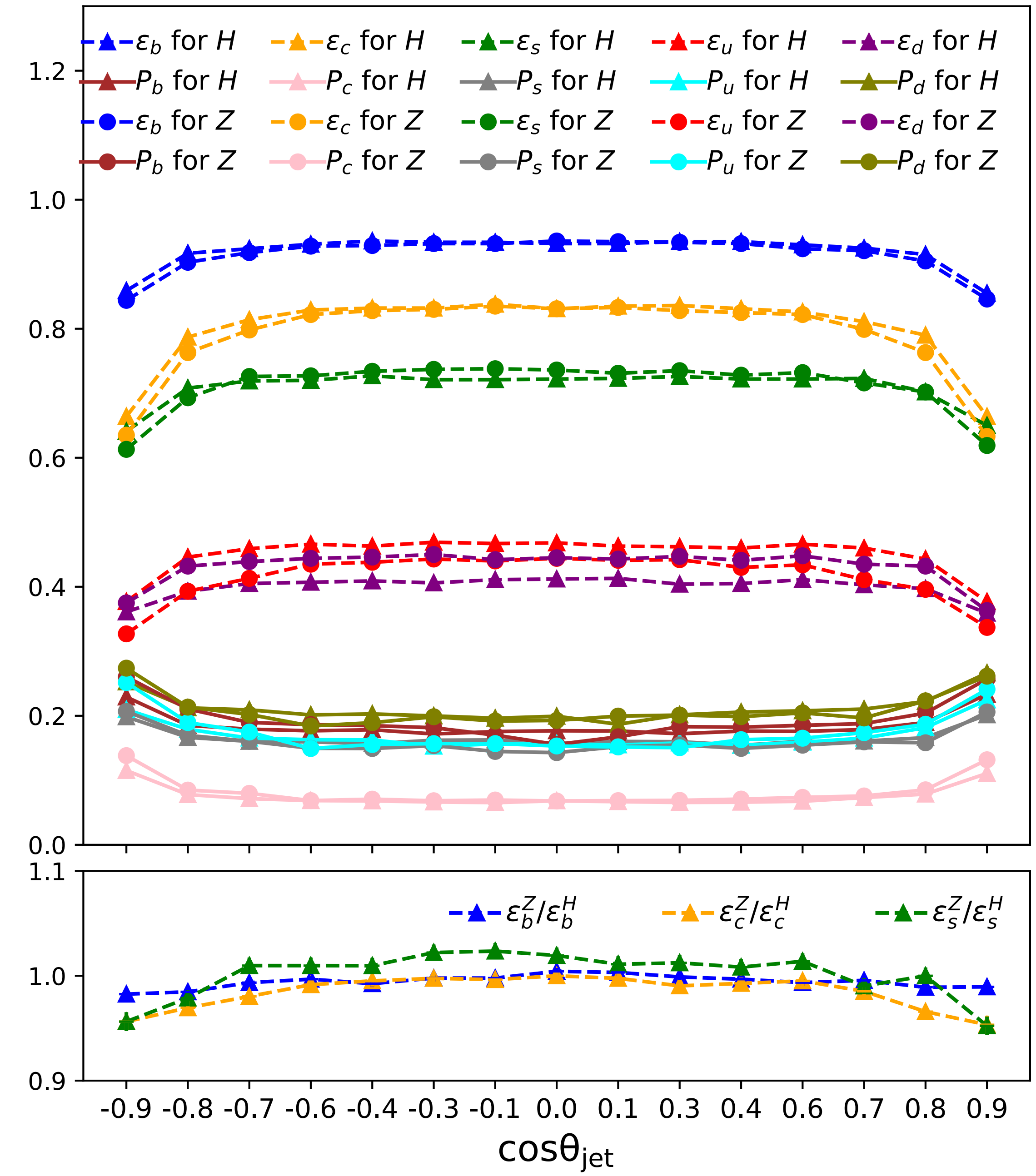
$\nu\bar{\nu}H, H \to jj$ at $\sqrt{s} = 240\ GeV$

The jet origin identification performance will degrade with increasing particle multiplicity of the jet.

Comparison between different physics processes

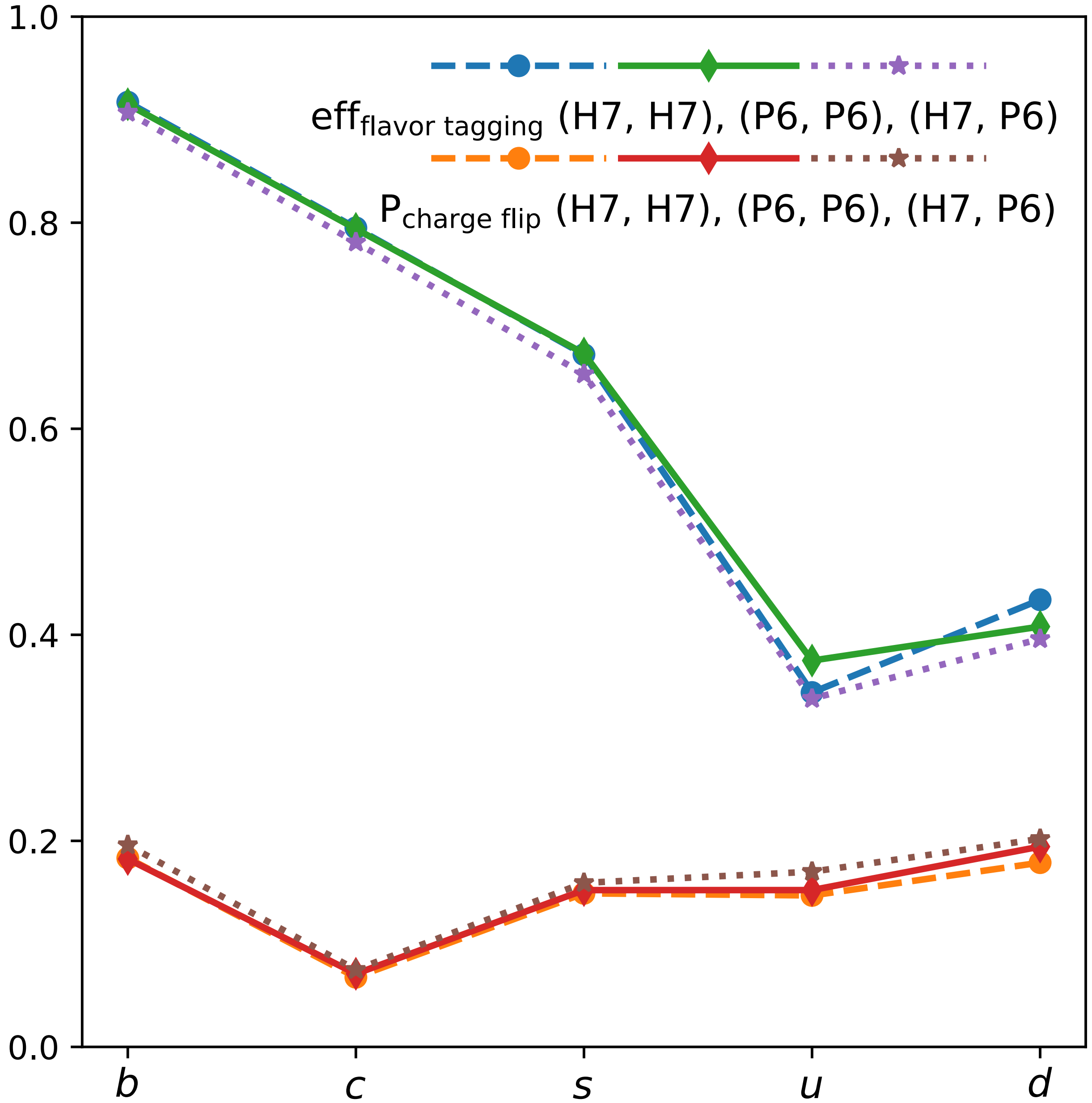
$$Z \to q\bar{q} \text{ at } \sqrt{s} = 91.2\ GeV$$
$$\nu\bar{\nu}H, H \to q\bar{q} \text{ at } \sqrt{s} = 240\ GeV$$





The jet origin identification performance agrees with each other, especially in the fiducial barrel region of the detector for the flavor tagging performance of b, c, and s.

Comparison between different hadronization models



Pythia-6.4
Herwig-7.2.2

$\nu\bar{\nu}H, H \rightarrow jj$ at $\sqrt{s} = 240\ GeV$

(A, B) means: training on A, test on B

The jet origin identificaion performanc agrees with each other, especially for b, c, and s jets, while exhibits small but visible differences for u and d jets.

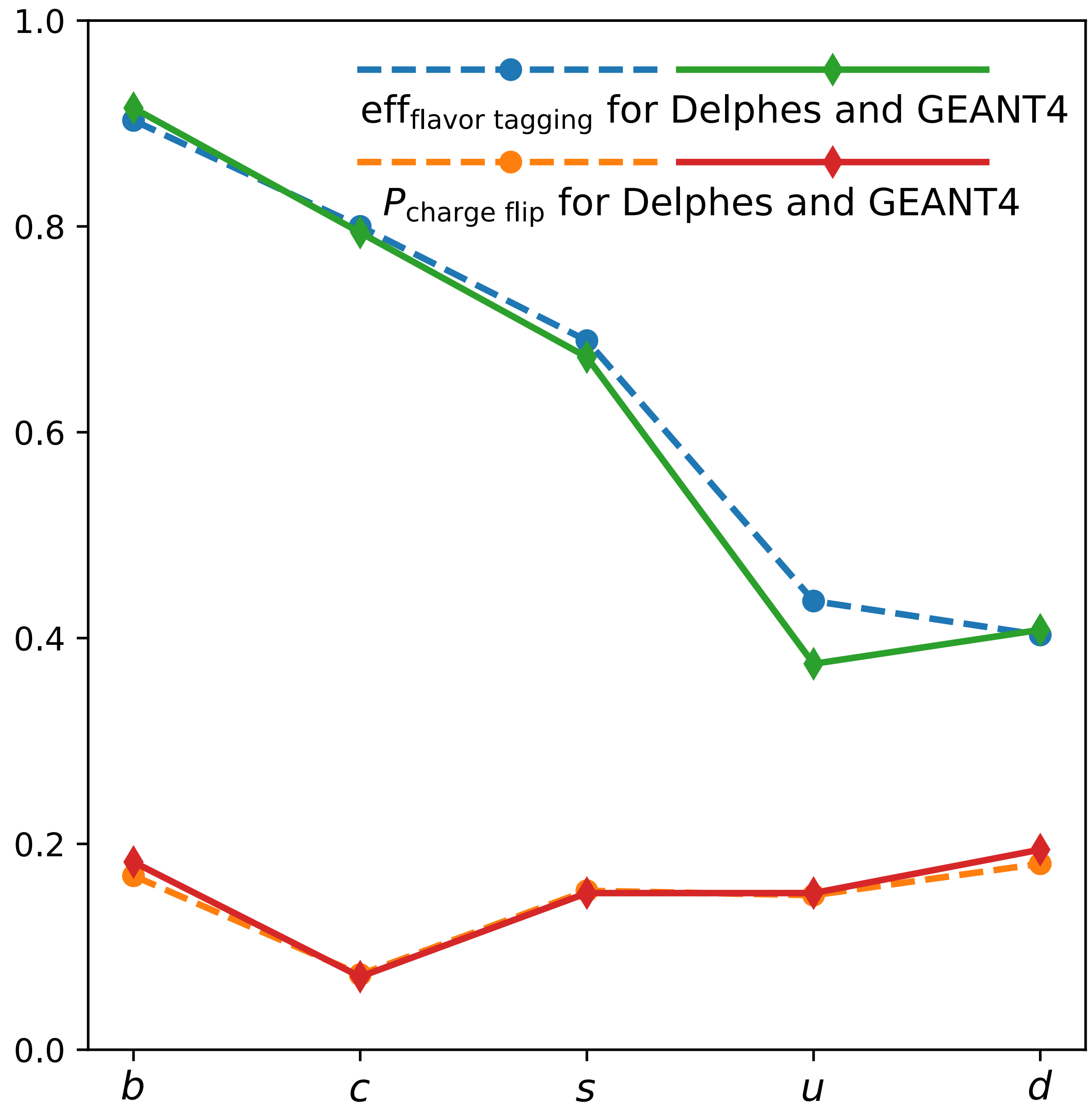
Summary

- We propose and realize the concept of jet origin identification that distinguishes jets generated from b, c, s, u, d, $\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d}$, and gluon.
- Based on the CEPC baseline detector and the full simulation $\nu\bar{\nu}H, H \rightarrow jj$ samples $\sqrt{s} = 240\ GeV$, the jet flavor tagging efficiencies range from 67% to 92% for heavy and strange quarks, and a jet charge flip rate of 7% to 18% for all species of quarks is achieved.
- Using jet origin identification, we estimate the upper limits on the branching ratios of seven Higgs rare and FCNC decay modes. The upper limit for H->ss decay can be improved by more than a factor of two compared to previous studies. The upper limits for H->uu/dd can be improved by roughly one order of magnitude compared to existing analyses.
- We found that the jet origin identification performance, especially those concerning the heavy and strange quarks, is rather stable versus the jet energy, jet polar angle, different physics processes, and even different hadronization models. The observed stability is vital for applying jet origin identification in real experiments.

Many thanks !

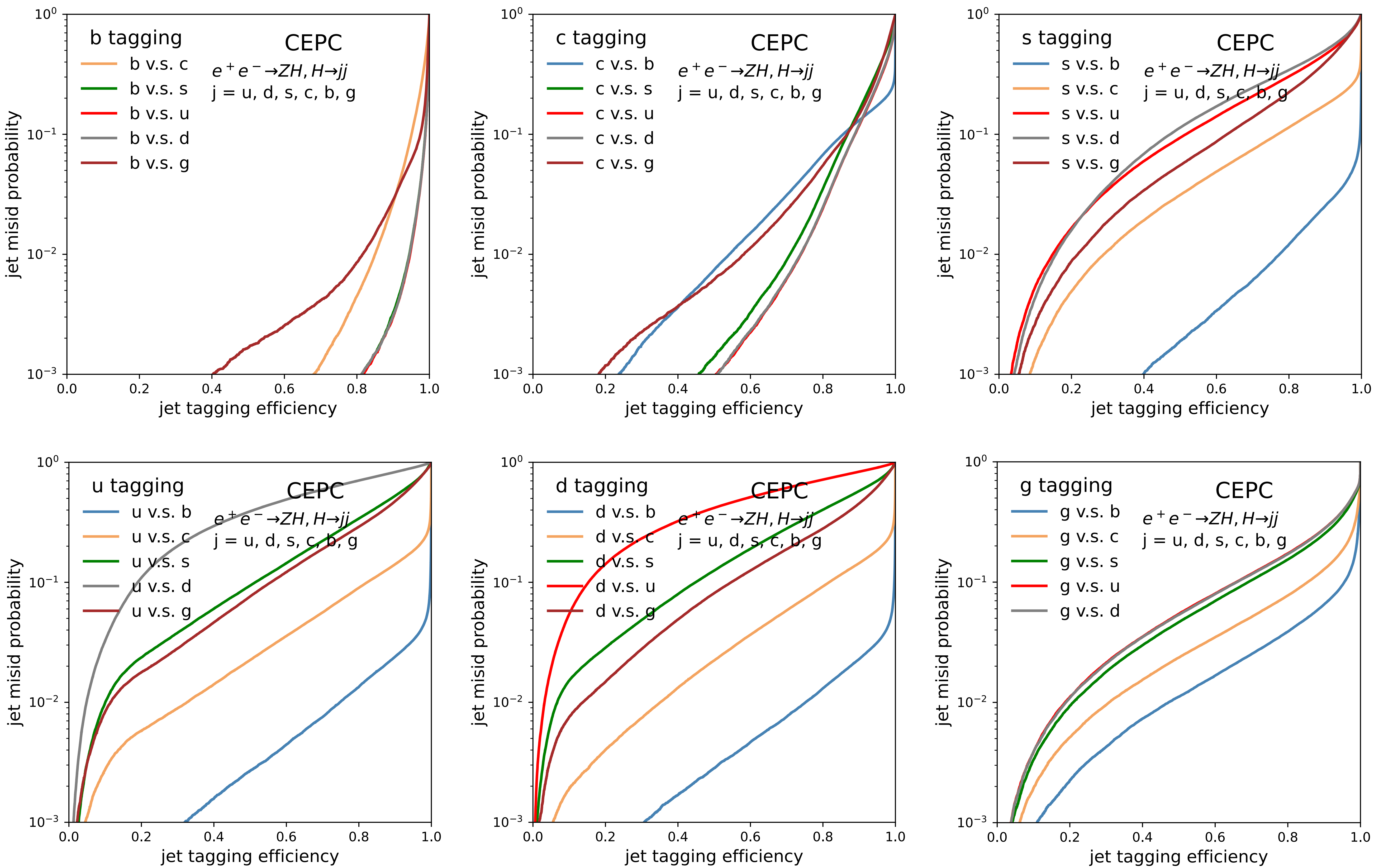
Backup

Comparison between full simulation and Delphes

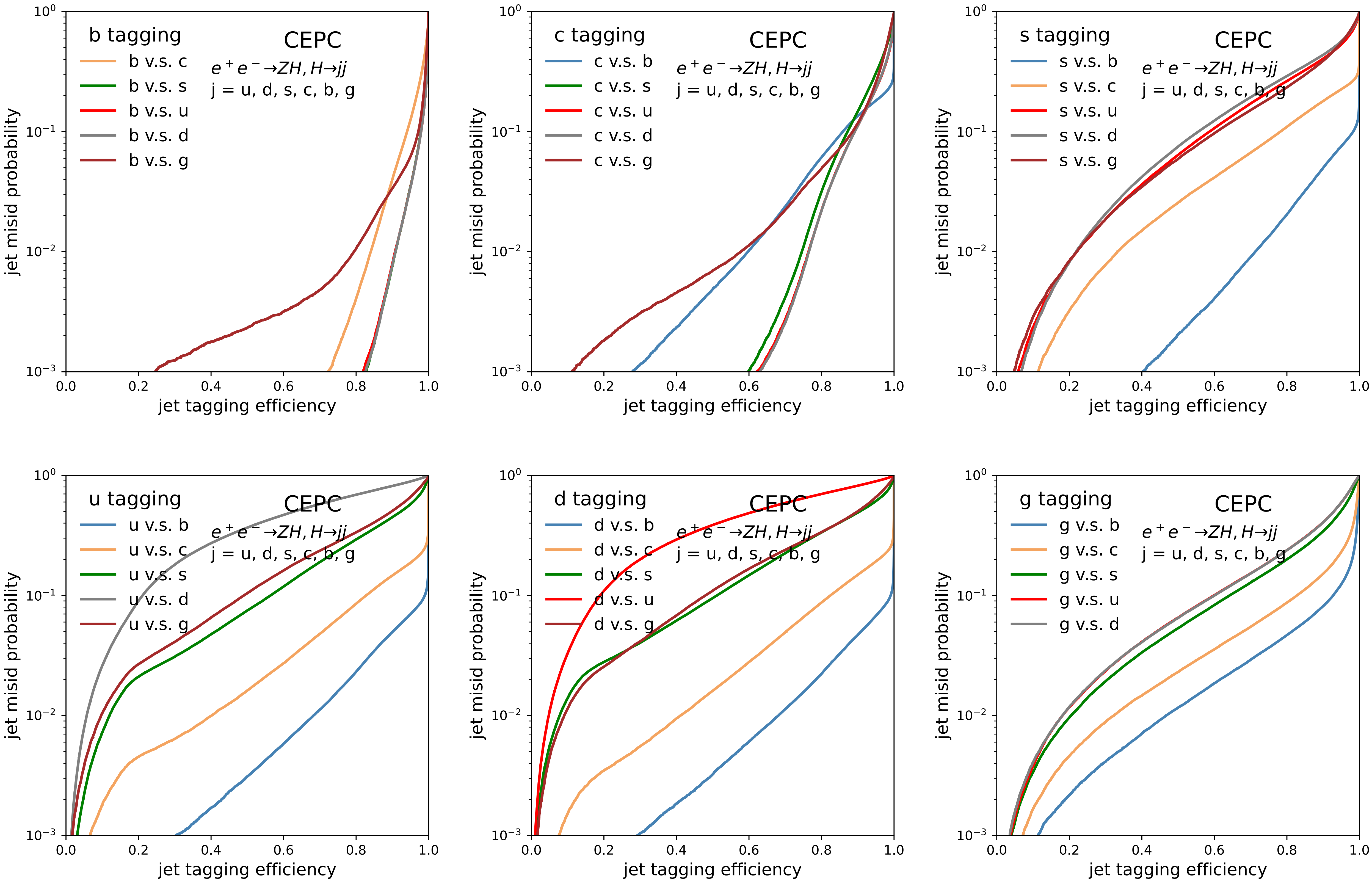


The Delphes fast simulation performance agrees with the full simulation one.

full simulation



Delphes


b tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
b v.s. c
b v.s. s
b v.s. u
b v.s. d
b v.s. g
c tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
c v.s. b
c v.s. s
c v.s. u
c v.s. d
c v.s. g
s tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
s v.s. b
s v.s. c
s v.s. u
s v.s. d
s v.s. g
u tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
u v.s. b
u v.s. c
u v.s. s
u v.s. d
u v.s. g
d tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
d v.s. b
d v.s. c
d v.s. s
d v.s. u
d v.s. g
g tagging
CEPC
$e^+e^- \to ZH, H \to jj$
j = u, d, s, c, b, g
g v.s. b
g v.s. c
g v.s. s
g v.s. u
g v.s. d
jet misid probability
jet tagging efficiency
10^{0}
10^{-1}
10^{-2}
10^{-3}
0.0
0.2
0.4
0.6
0.8
1.0