Jet Origin Identification at Electron-Positron Higgs Factory

PhysRevLett.132.221802

https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification

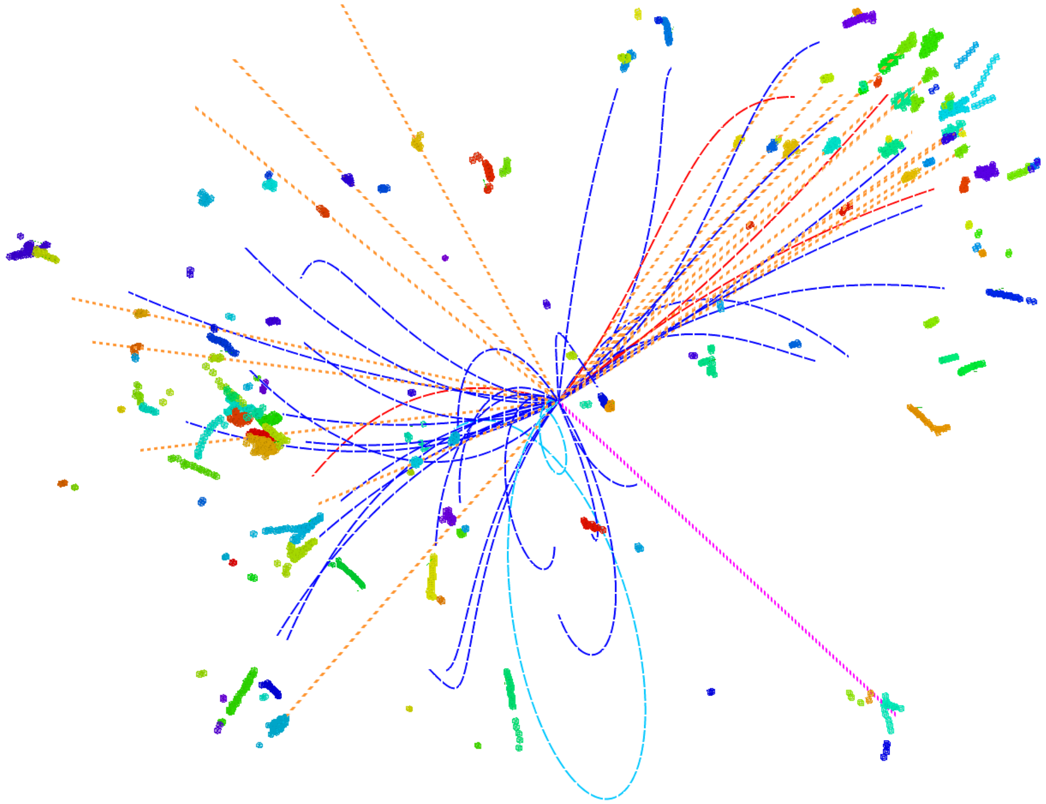
朱永峰

2024.11.14

CLHCP2024 @ Qingdao

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 $\sqrt{s} = 240\ GeV$

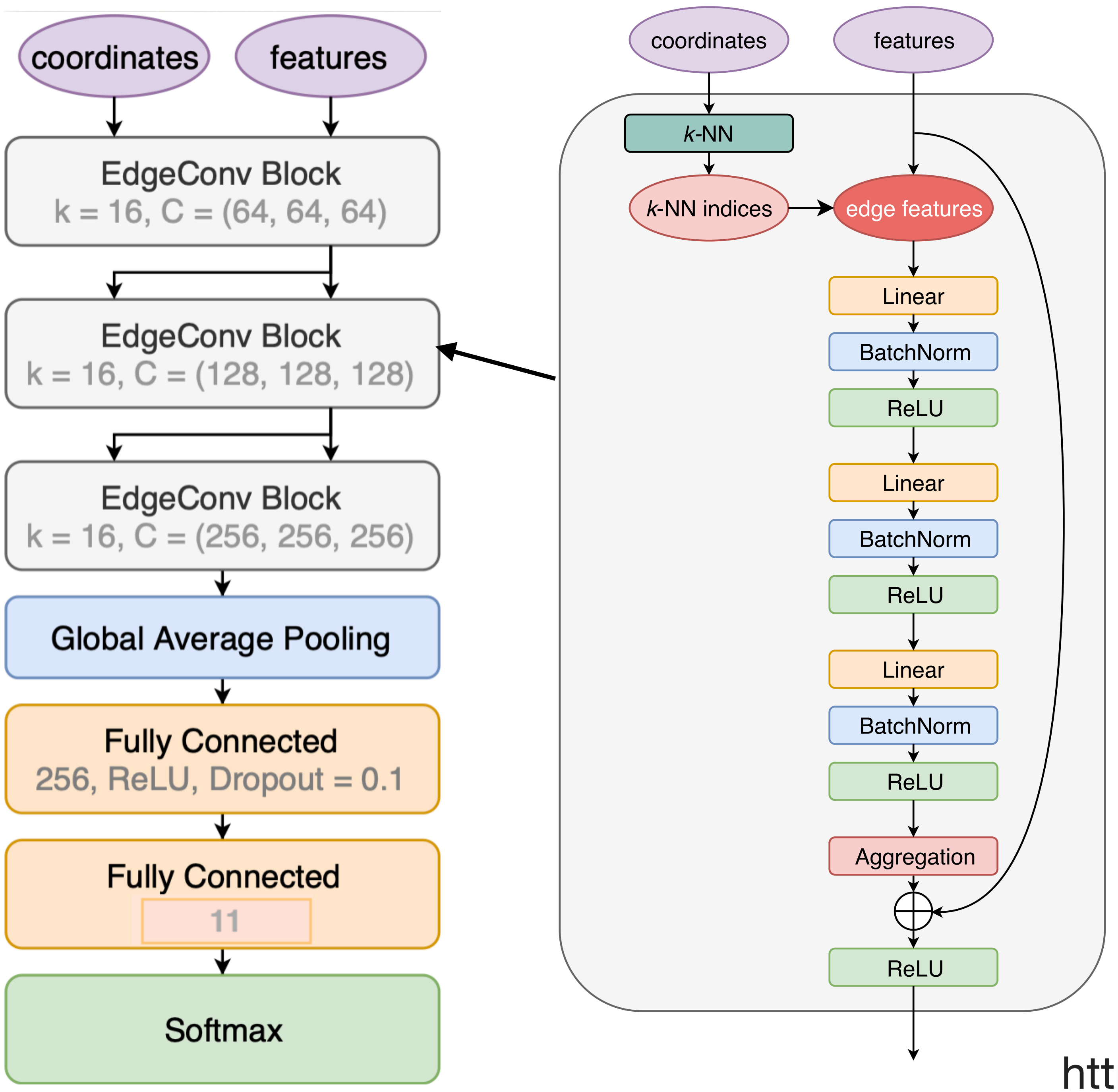
Jet Origin Identification: categorizes jets into 5 quarks (b, c, s, u, d), 5 anti-quark ($\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d}$), and gluon.
It comprises the concepts of jet flavor tagging, jet charge measurement, s-quark tagging, and gluon finding.

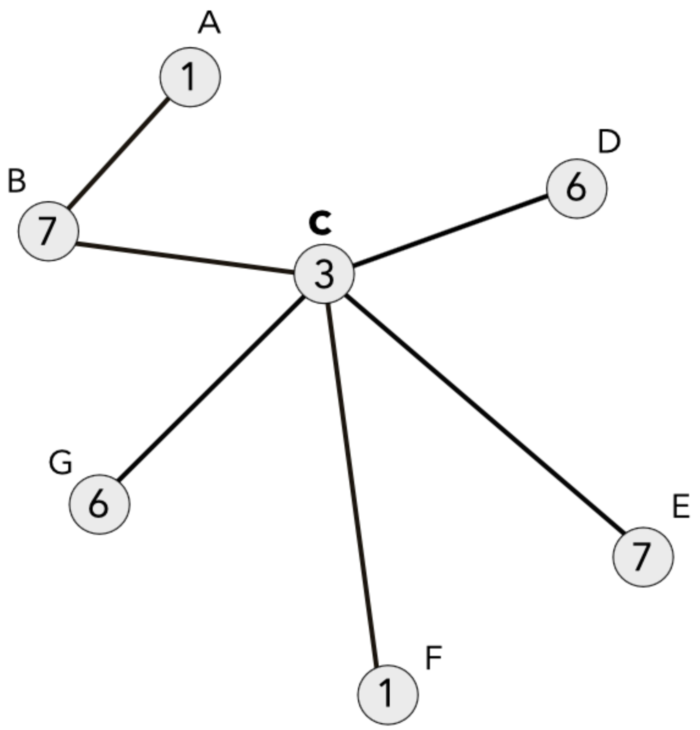
Contents

- samples and ParticleNet
- the stability of ParticleNet at the CEPC
- the performance of JOI
- Benchmark physics analyses
- the dependent factors of JOI performance
- Performance comparison of ParticleNet, Particle Transformer, and LLM

Samples and ParticleNet

samples: Based on the CEPC baseline detector, full simulation of $e^+e^- \to \nu\bar{\nu}H, H \to b\bar{b}/c\bar{c}/s\bar{s}/u\bar{u}/d\bar{d}/gg$ at $\sqrt{s} = 240\ GeV$ generated with Whizard1.95 and Pythia-6.4.





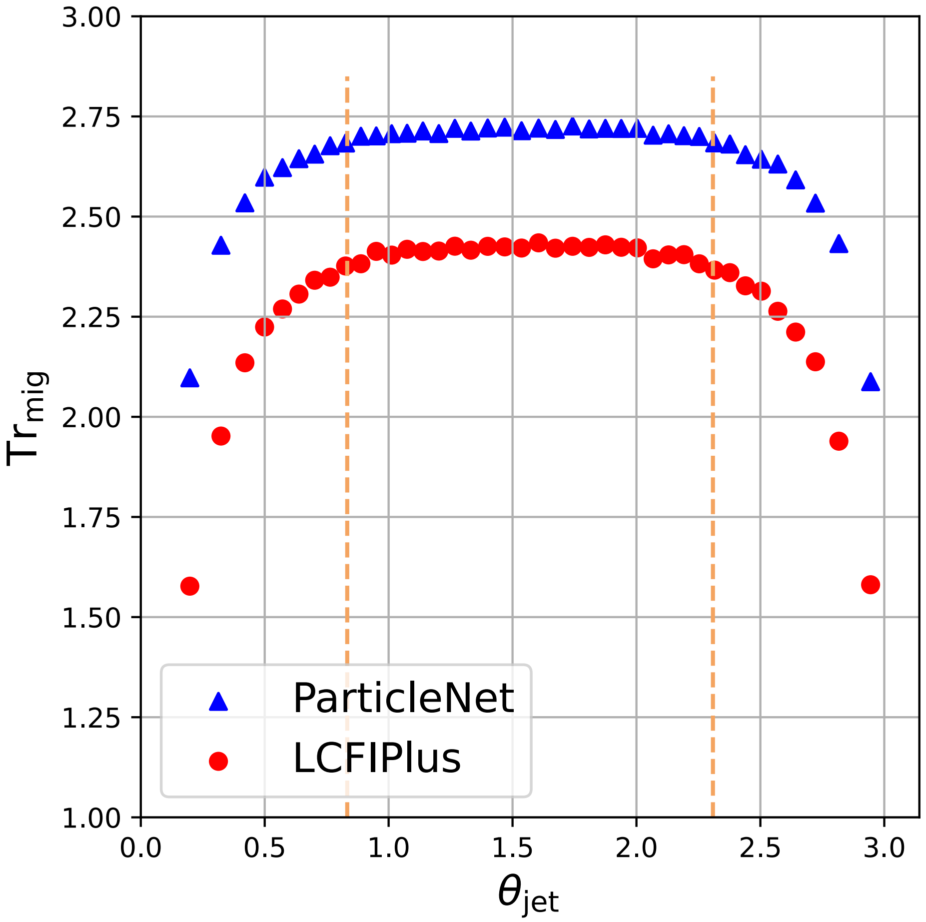
input features:

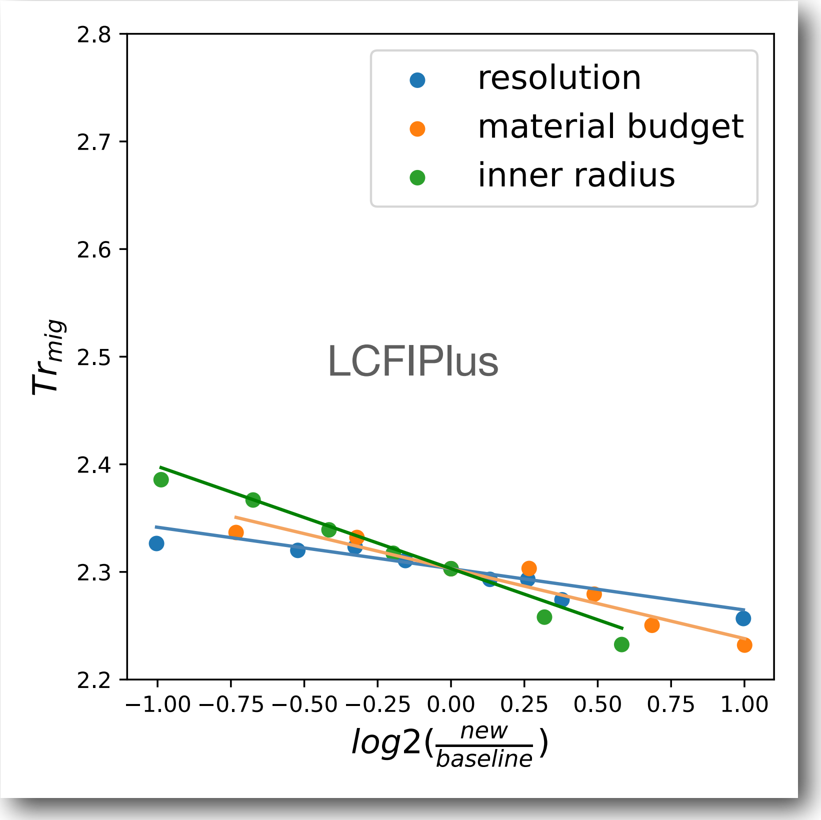
impact parameters, kinematic variables, and PID information of particles in a jet

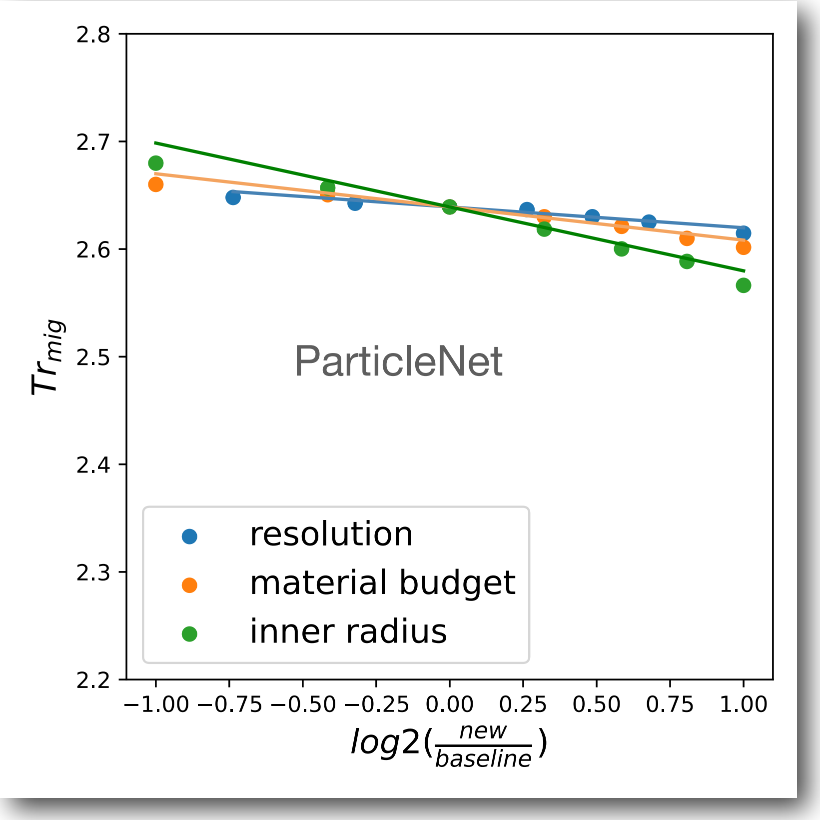
https://doi.org/10.1103/PhysRevD.101.056019

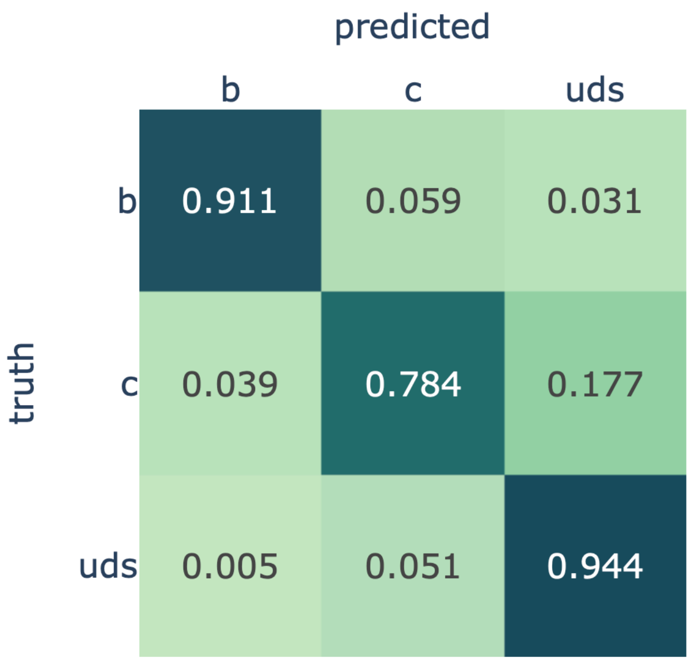
validate the stability of ParticleNet, b/c/light tagging

(dependence of performance on jet polar angle and vertex detector configuration)









predicted

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

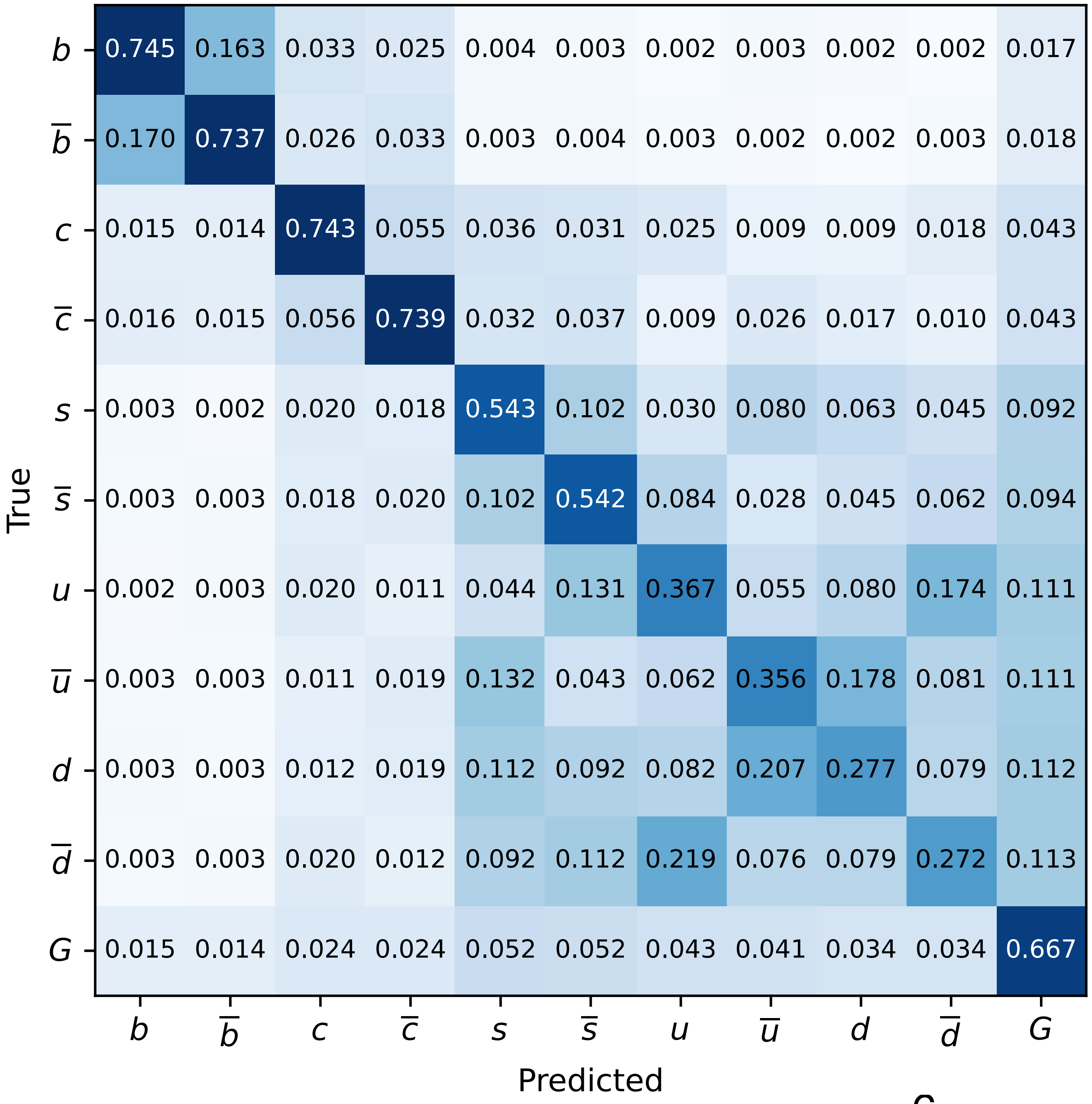
$$\mathrm{PN}: Tr_{mig} = 2.639 + 0.031 \cdot log_2\frac{R^0_{material}}{R_{material}} + 0.019 \cdot log_2\frac{R^0_{resolution}}{R_{resolution}} + 0.059 \cdot log_2\frac{R^0_{radius}}{R_{radius}}$$

$$\mathrm{LCFIPlus}: Tr_{mig} = 2.303 + 0.065 \cdot log_2\frac{R^0_{material}}{R_{material}} + 0.038 \cdot log_2\frac{R^0_{resolution}}{R_{resolution}} + 0.095 \cdot log_2\frac{R^0_{radius}}{R_{radius}}$$

Eur. Phys. J. C (2024) 84:152

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.



	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

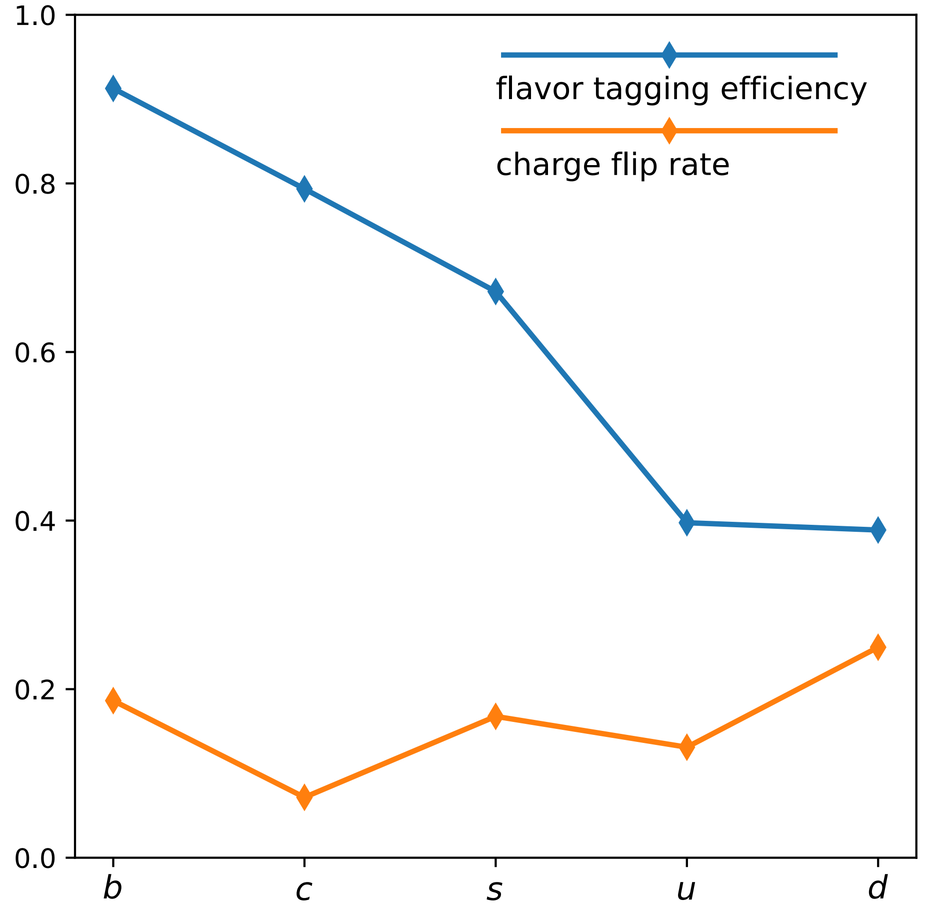
jet flavor tagging efficiency

is defined as

$$max(b+\bar{b},\ c+\bar{c},\ s+\bar{s},\ u+\bar{u},\ d+\bar{d},\ g)$$

jet charge flip rate

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



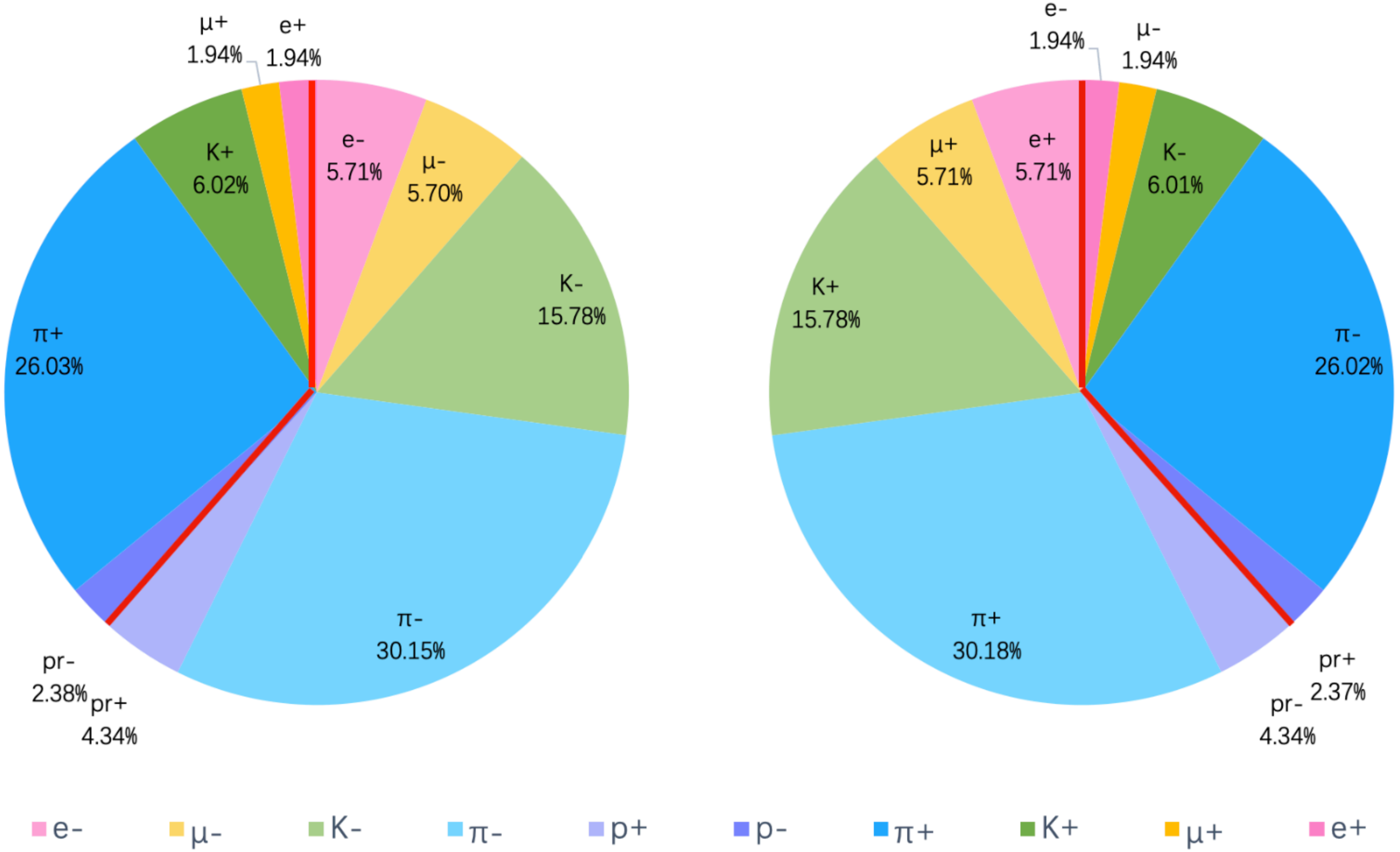
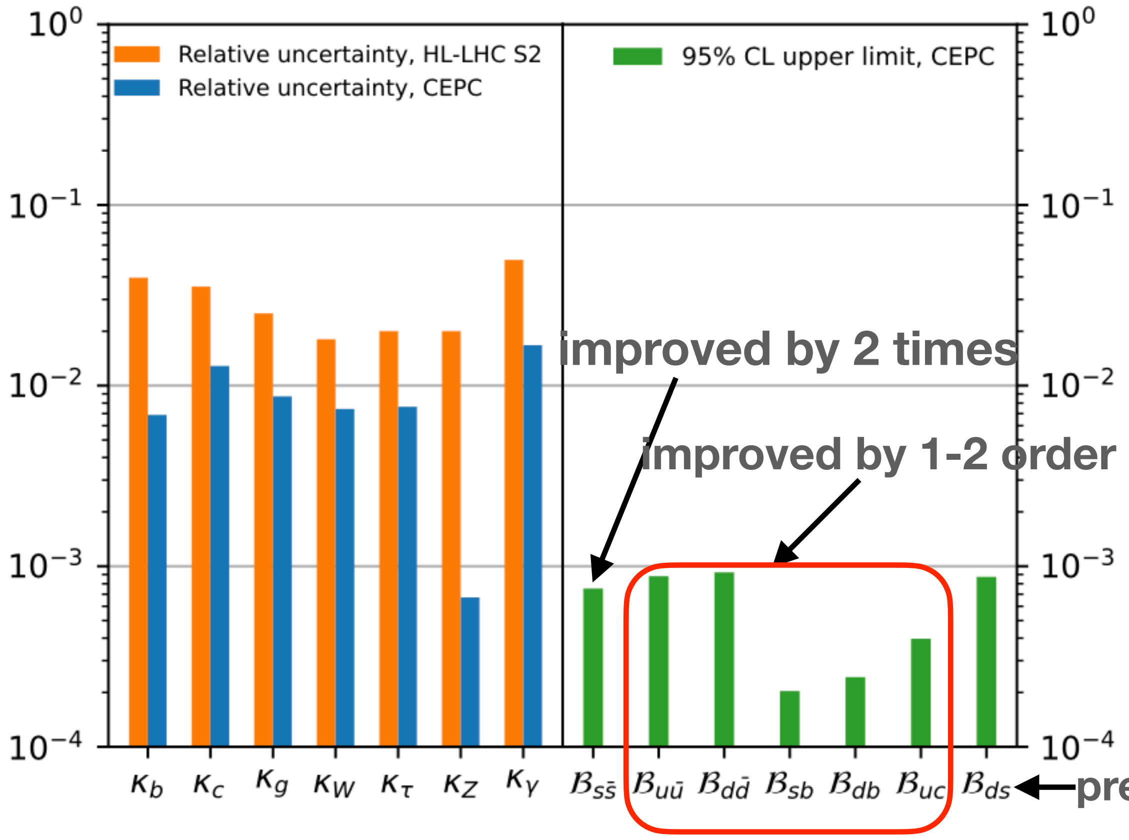


Figure 5. The percentages of species of final state leading charged particles within the b jet (left) and the $\bar{b}$ jet (right) by WHIZARD 1.95.

arXiv:2306.14089

Benchmark physics analyses

Begin with the existing analyses of $\nu\bar{\nu}H, H \to b\bar{b}/c\bar{c}/gg$, (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 1-2 order of magitude

presumably first quantified

FIG. 5. Expected upper limits on the branching ratios of rare Higgs boson decays from this Letter (green) and the relative uncertainties of Higgs couplings anticipated at CEPC [19] (blue) and HL-LHC [43] (orange) under the kappa-0 fit scenario [54] and scenario S2 of systematics [55], as cited in Ref. [19]. The limit on $B_{s\bar{s}}$ corresponds to an upper limit of 1.7 on the Higgs-strange coupling modifier κ_s (not shown).

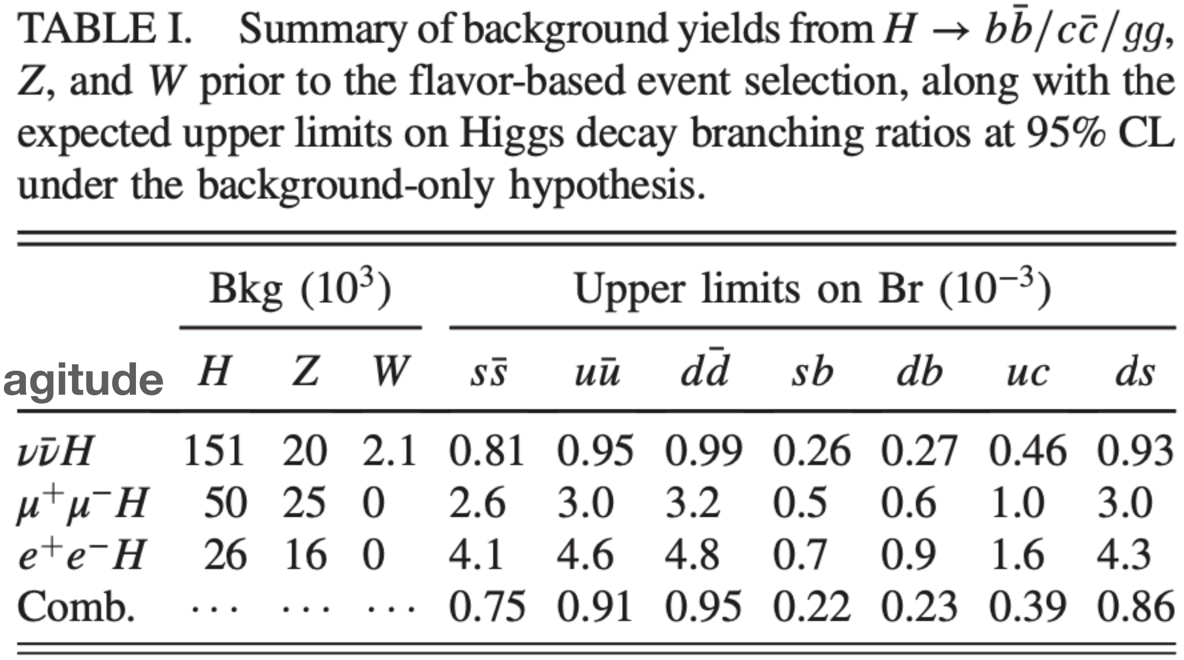


TABLE I. Summary of background yields from $H \to b\bar{b}/c\bar{c}/gg$, Z, and W prior to the flavor-based event selection, along with the expected upper limits on Higgs decay branching ratios at 95% CL under the background-only hypothesis.

	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].

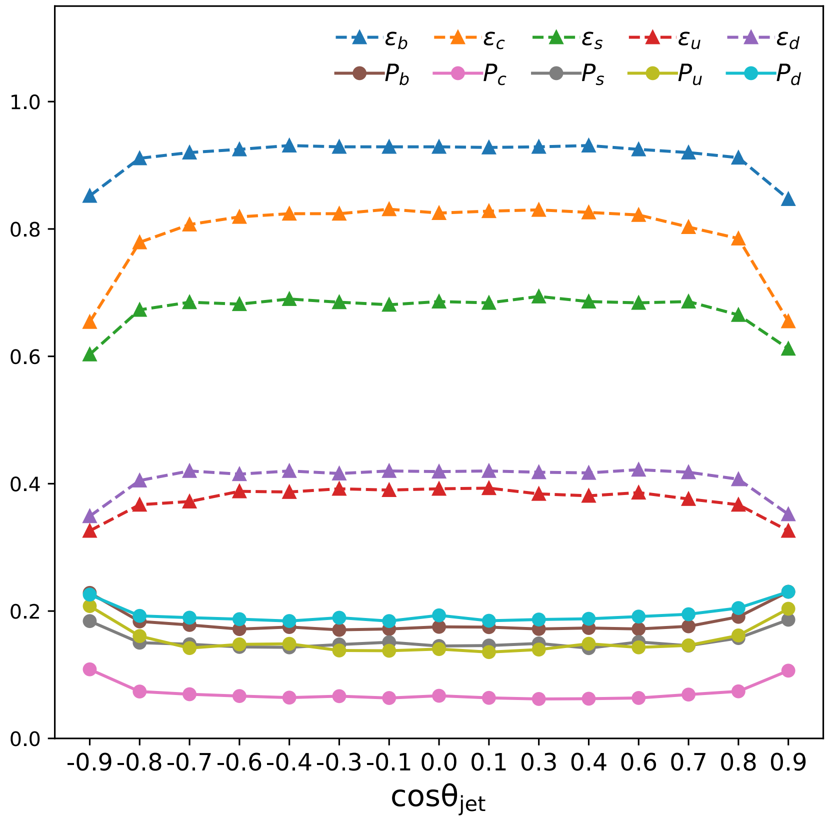
[44] A. Albert *et al.*, "Strange quark as a probe for new physics in the higgs sector," (2022), arXiv:2203.07535 [hep-ex].

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

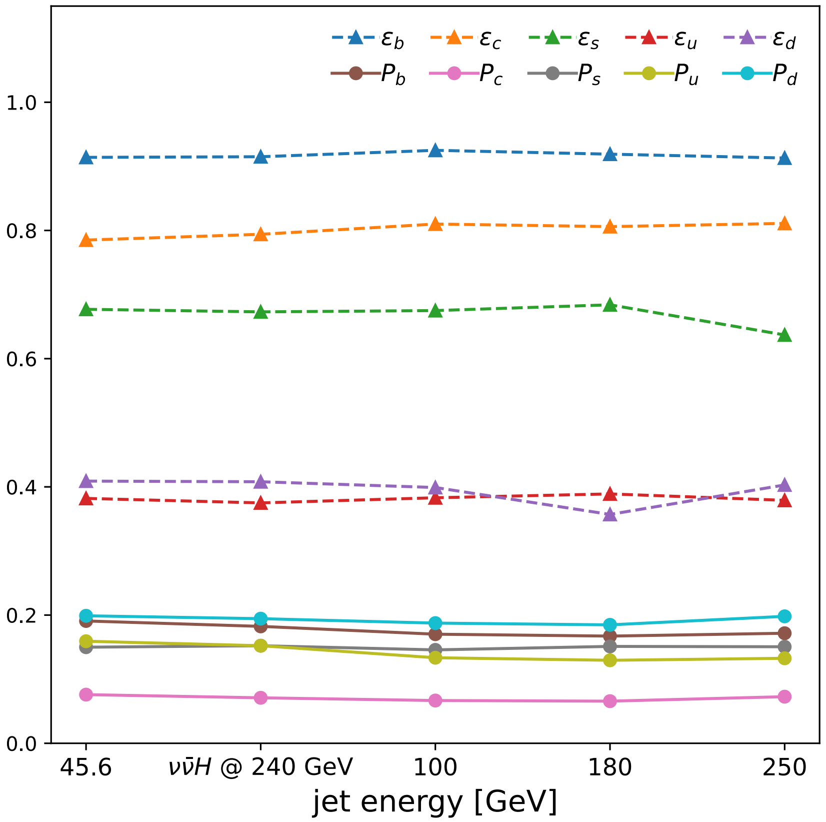
[54] J. De Blas, G. Durieux, C. Grojean, J. Gu, and A. Paul, JHEP **12**, 117 (2019), arXiv:1907.04311 [hep-ph].

dependence on

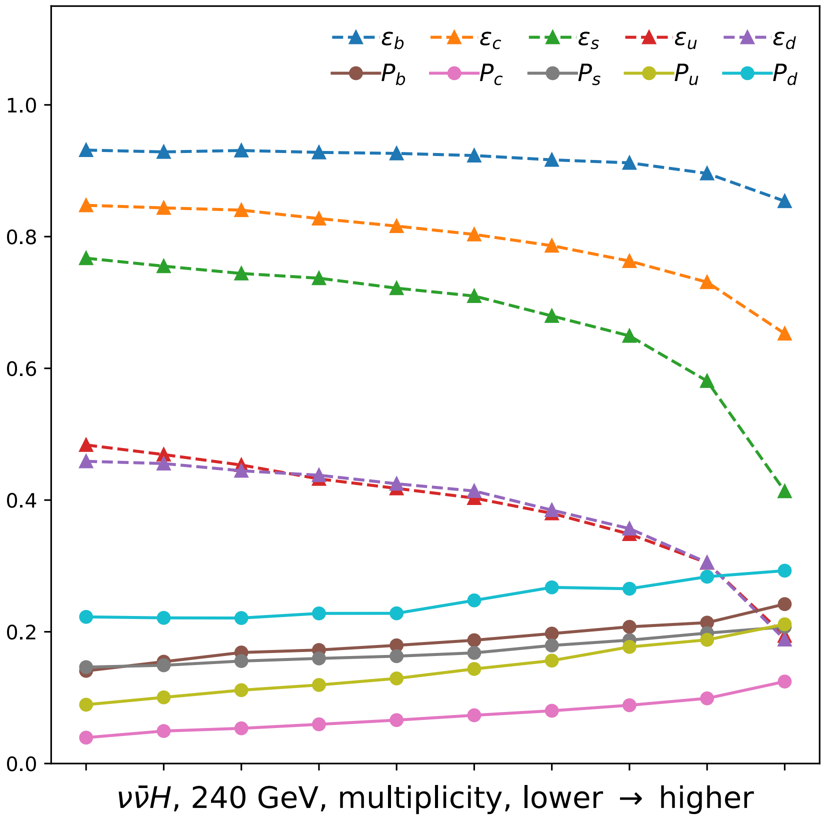
jet polar angle,



jet energy,



multiplicity

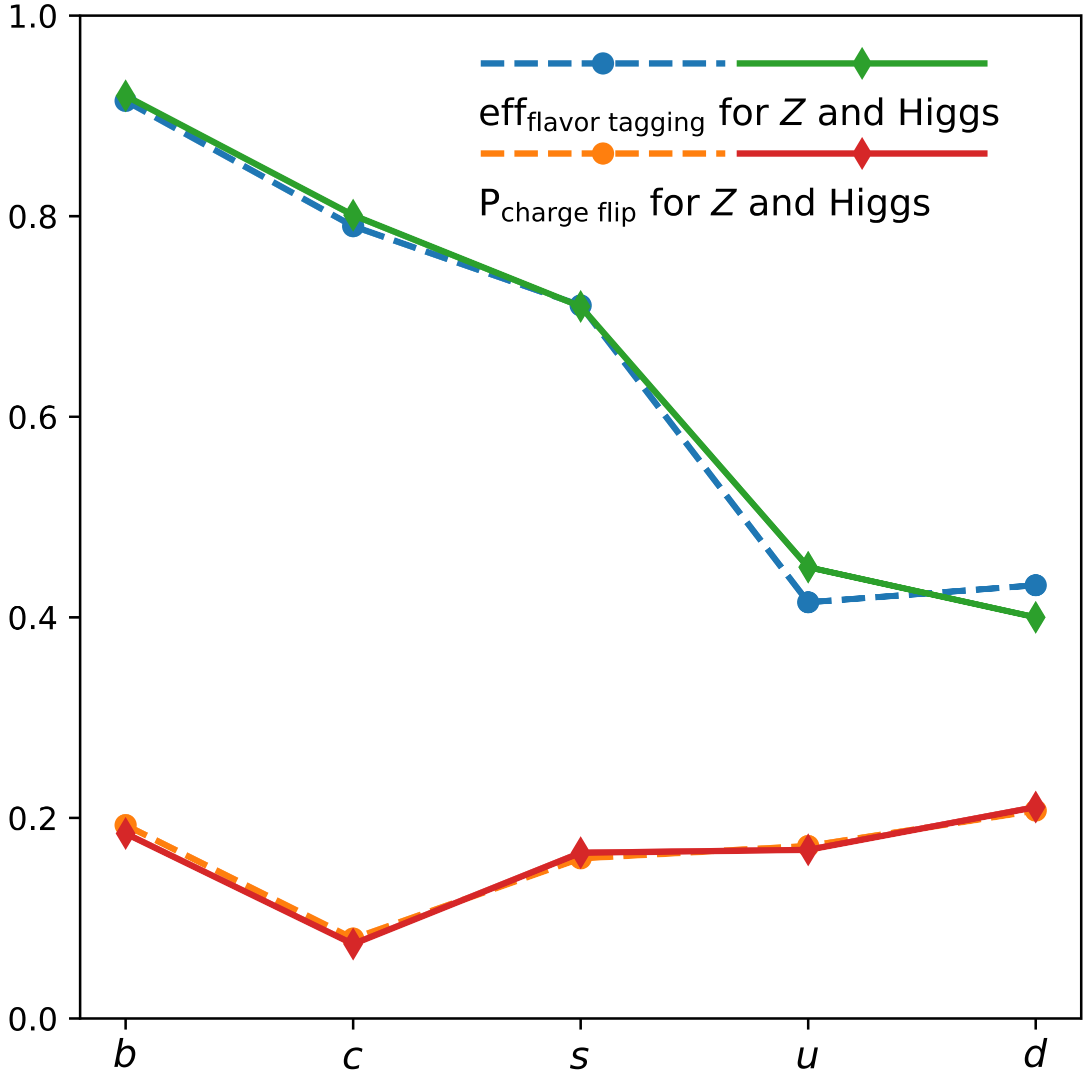


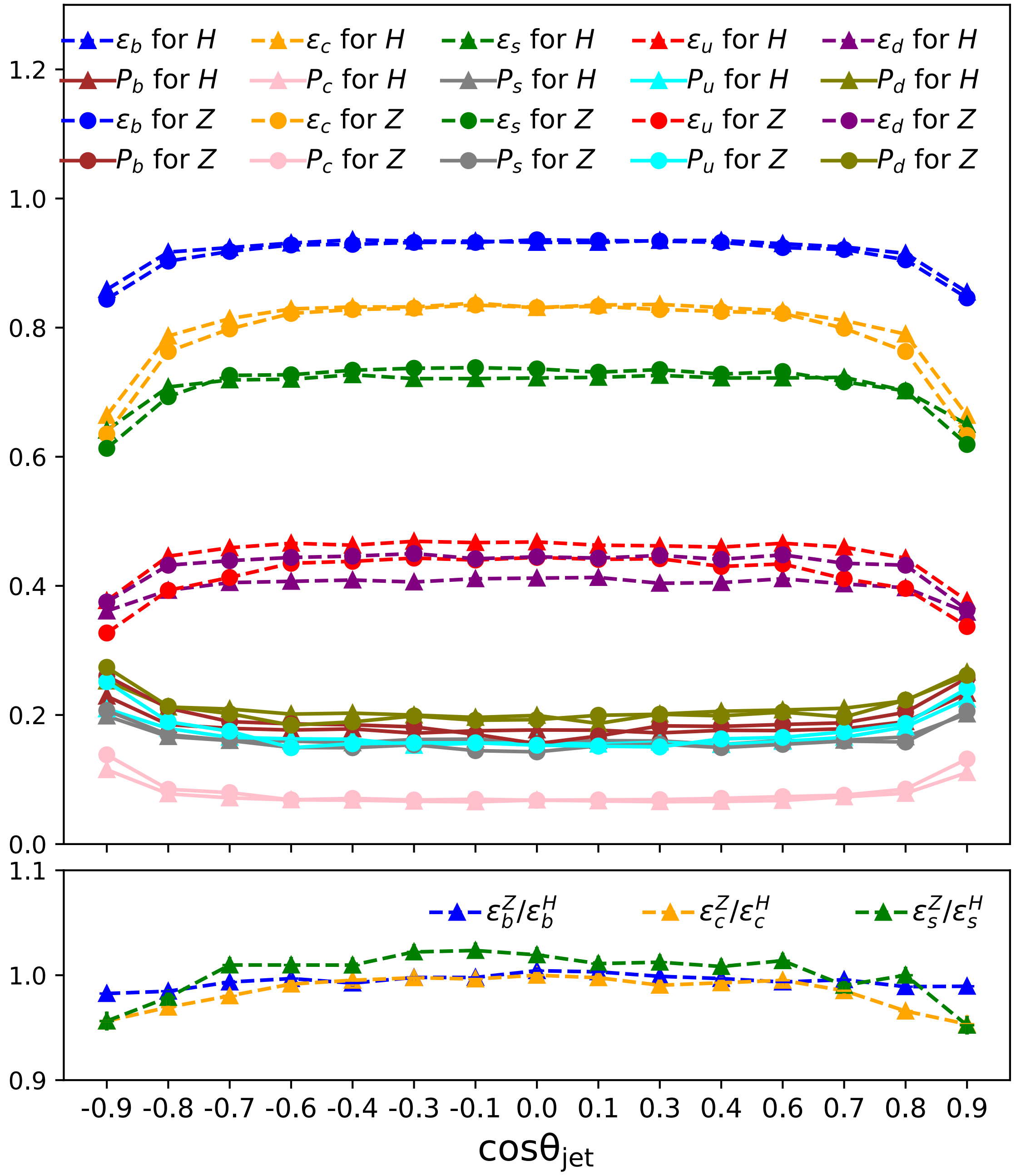
- With $\nu\bar{\nu}H, H \rightarrow jj$ at $\sqrt{s} = 240\ GeV$, the jet tagging efficiencies and charge flip rates are flat in the barrel region of the detector and exhibit slight degrading in the endcap region.
- The jet tagging efficiencies and charge flip rates are rather stable with various jet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

dependence on different physics processes

$Z \to q\bar{q}$ at $\sqrt{s} = 91.2\ GeV$

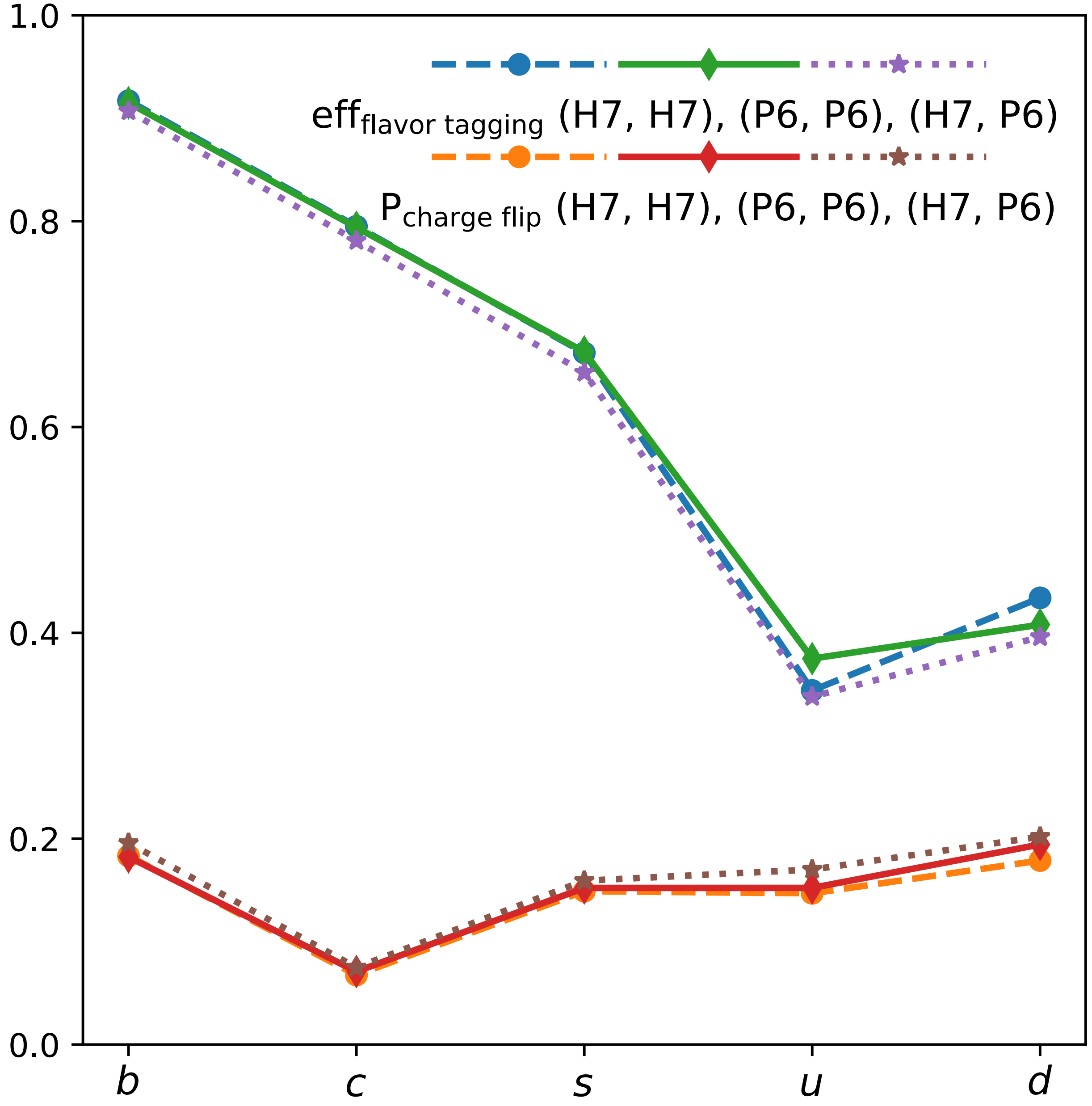
$\nu\bar{\nu}H, H \to q\bar{q}$ at $\sqrt{s} = 240\ GeV$





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.

dependence on 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.

- 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.

$|V_{cb}|$ measurement from W boson decays

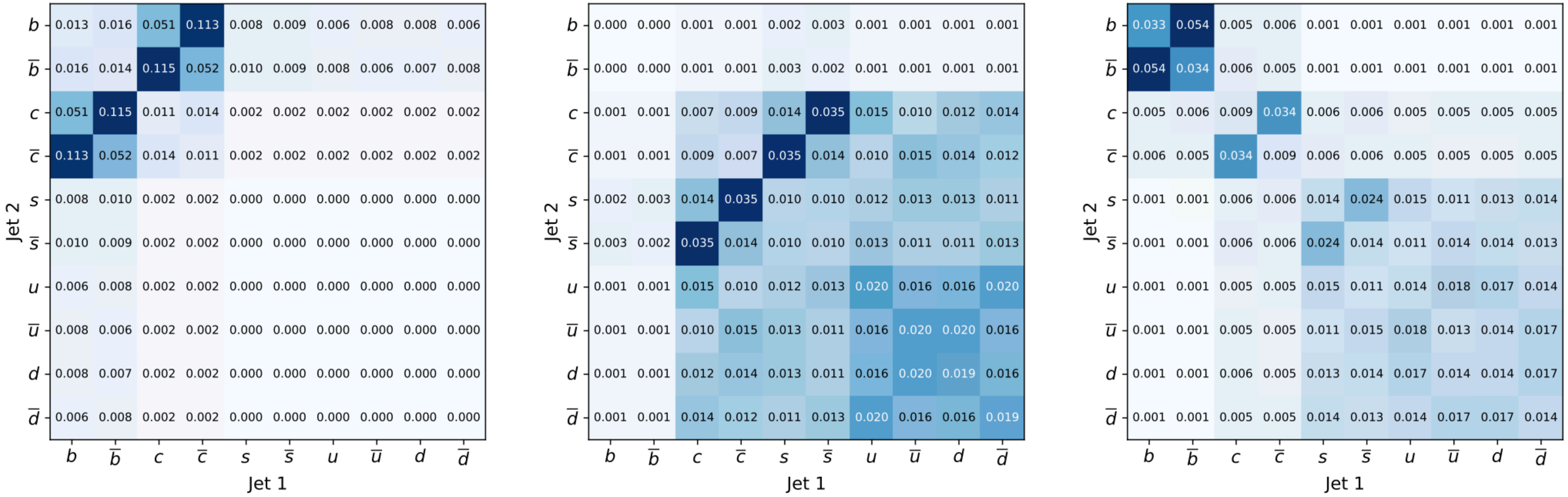
https://arxiv.org/pdf/2406.01675

Advantages:

- Measuring the decay width in W → cb transitions provides an alternative approach to determine $|V_{cb}|$ at an energy scale $\gg \Lambda_{QCD}$, minimizing non-perturbative corrections.
- Furthermore, by extracting $|V_{cb}|$ values from the ratio of $\Gamma(W \to cb)/\Gamma(W \to qq)$, many theoretical and experimental uncertainties will cancel out, such as the leading perturbative QCD corrections.
- Such a measurement also provides independent consistency checks since its systematic uncertainties differ from those in b-hadron decays.

Status:

- The only CKM element measurements via W boson decay are the $|V_{cs}|$ measurements at LEP, while the $|V_{cb}|$ measurement at LEP is impractical due to the small Br(W → cb) value (6×10^{-4} in the SM) and the limited statistics of W.
- Recent studies have explored $|V_{cb}|$ measurement through on-shell W bosons at the LHC [25–27]. However, the projected relative sensitivity of $|V_{cb}|$ at the LHC is only O(10%), which is too weak to resolve the current tension.



	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$
b	0.013	0.016	0.051	0.113	0.008	0.009	0.006	0.008	0.008	0.006
$\bar{b}$	0.016	0.014	0.115	0.052	0.010	0.009	0.008	0.006	0.007	0.008
c	0.051	0.115	0.011	0.014	0.002	0.002	0.002	0.002	0.002	0.002
$\bar{c}$	0.113	0.052	0.014	0.011	0.002	0.002	0.002	0.002	0.002	0.002
s	0.008	0.010	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
$\bar{s}$	0.010	0.009	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
u	0.006	0.008	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
$\bar{u}$	0.008	0.006	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
d	0.008	0.007	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
$\bar{d}$	0.006	0.008	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000

	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$
b	0.000	0.000	0.001	0.001	0.002	0.003	0.001	0.001	0.001	0.001
$\bar{b}$	0.000	0.000	0.001	0.001	0.003	0.002	0.001	0.001	0.001	0.001
c	0.001	0.001	0.007	0.009	0.014	0.035	0.015	0.010	0.012	0.014
$\bar{c}$	0.001	0.001	0.009	0.007	0.035	0.014	0.010	0.015	0.014	0.012
s	0.002	0.003	0.014	0.035	0.010	0.010	0.012	0.013	0.013	0.011
$\bar{s}$	0.003	0.002	0.035	0.014	0.010	0.010	0.013	0.011	0.011	0.013
u	0.001	0.001	0.015	0.010	0.012	0.013	0.020	0.016	0.016	0.020
$\bar{u}$	0.001	0.001	0.010	0.015	0.013	0.011	0.016	0.020	0.020	0.016
d	0.001	0.001	0.012	0.014	0.013	0.011	0.016	0.020	0.019	0.016
$\bar{d}$	0.001	0.001	0.014	0.012	0.011	0.013	0.020	0.016	0.016	0.019

	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$
b	0.033	0.054	0.005	0.006	0.001	0.001	0.001	0.001	0.001	0.001
$\bar{b}$	0.054	0.034	0.006	0.005	0.001	0.001	0.001	0.001	0.001	0.001
c	0.005	0.006	0.009	0.034	0.006	0.006	0.005	0.005	0.005	0.005
$\bar{c}$	0.006	0.005	0.034	0.009	0.006	0.006	0.005	0.005	0.005	0.005
s	0.001	0.001	0.006	0.006	0.014	0.024	0.015	0.011	0.013	0.014
$\bar{s}$	0.001	0.001	0.006	0.006	0.024	0.014	0.011	0.014	0.014	0.013
u	0.001	0.001	0.005	0.005	0.015	0.011	0.014	0.018	0.017	0.014
$\bar{u}$	0.001	0.001	0.005	0.005	0.011	0.015	0.018	0.013	0.014	0.017
d	0.001	0.001	0.006	0.005	0.013	0.014	0.017	0.014	0.014	0.017
$\bar{d}$	0.001	0.001	0.005	0.005	0.014	0.013	0.014	0.017	0.017	0.014

Figure 3: The average tensor products of the two jets' 10 dimensional scores for $WW \to \mu\nu cb$ (left), other WW background events (middle), and the rest of backgrounds (right).

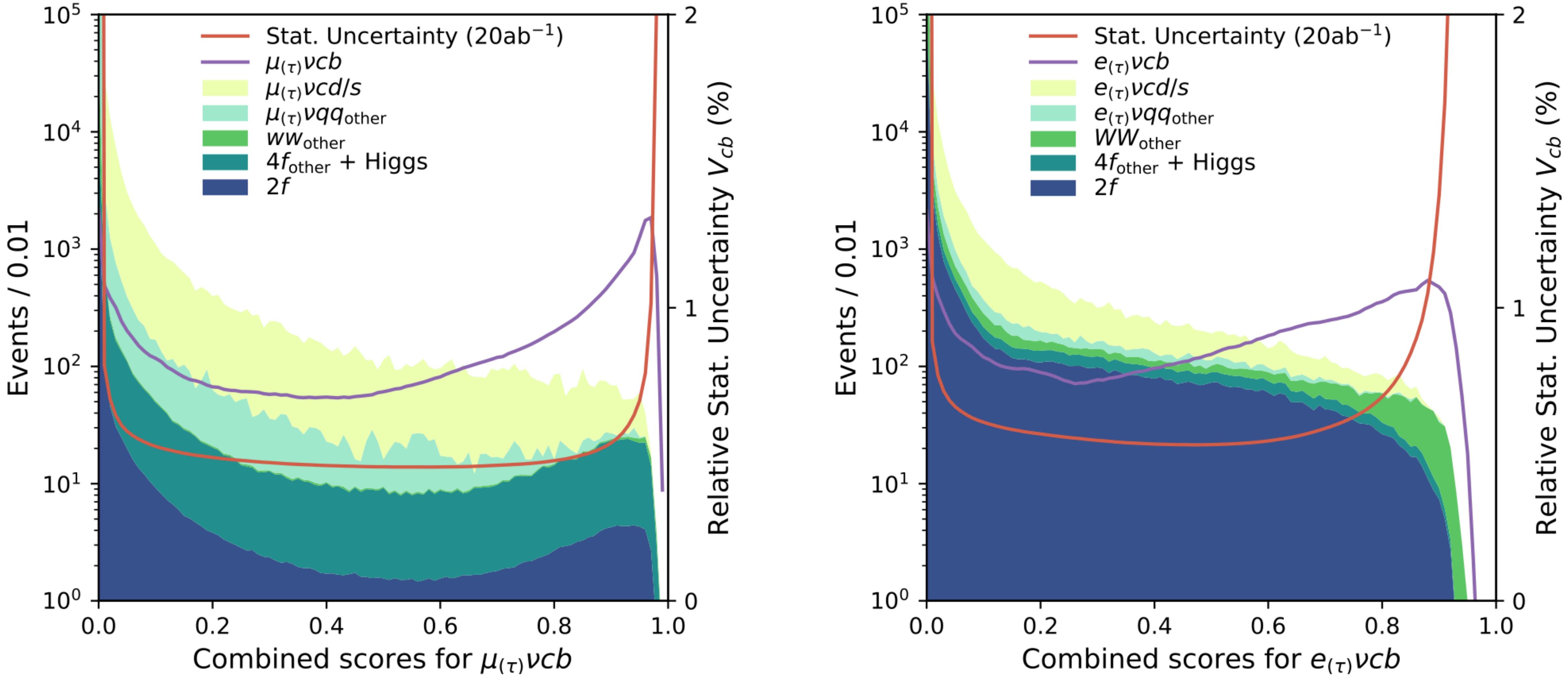
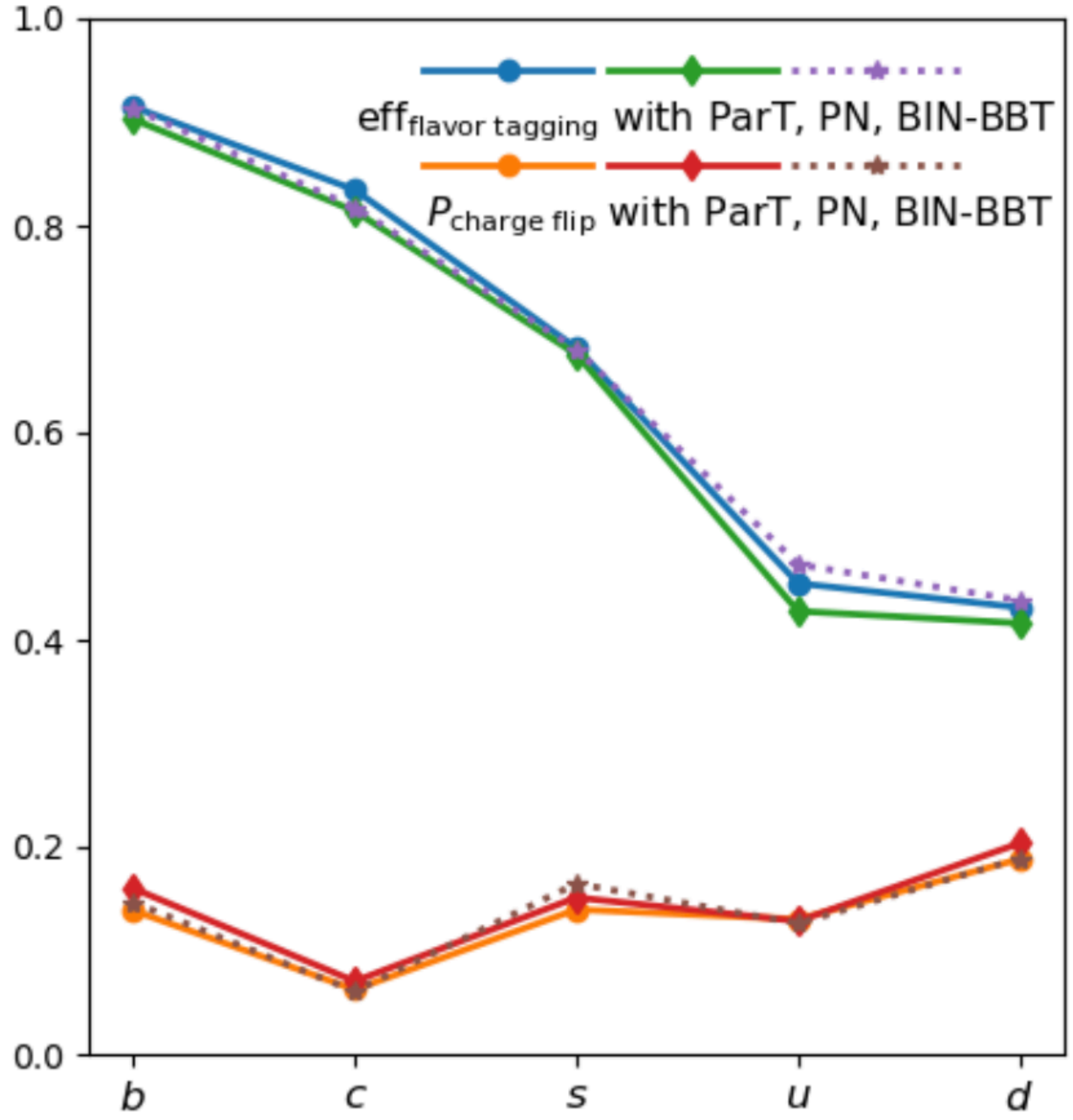


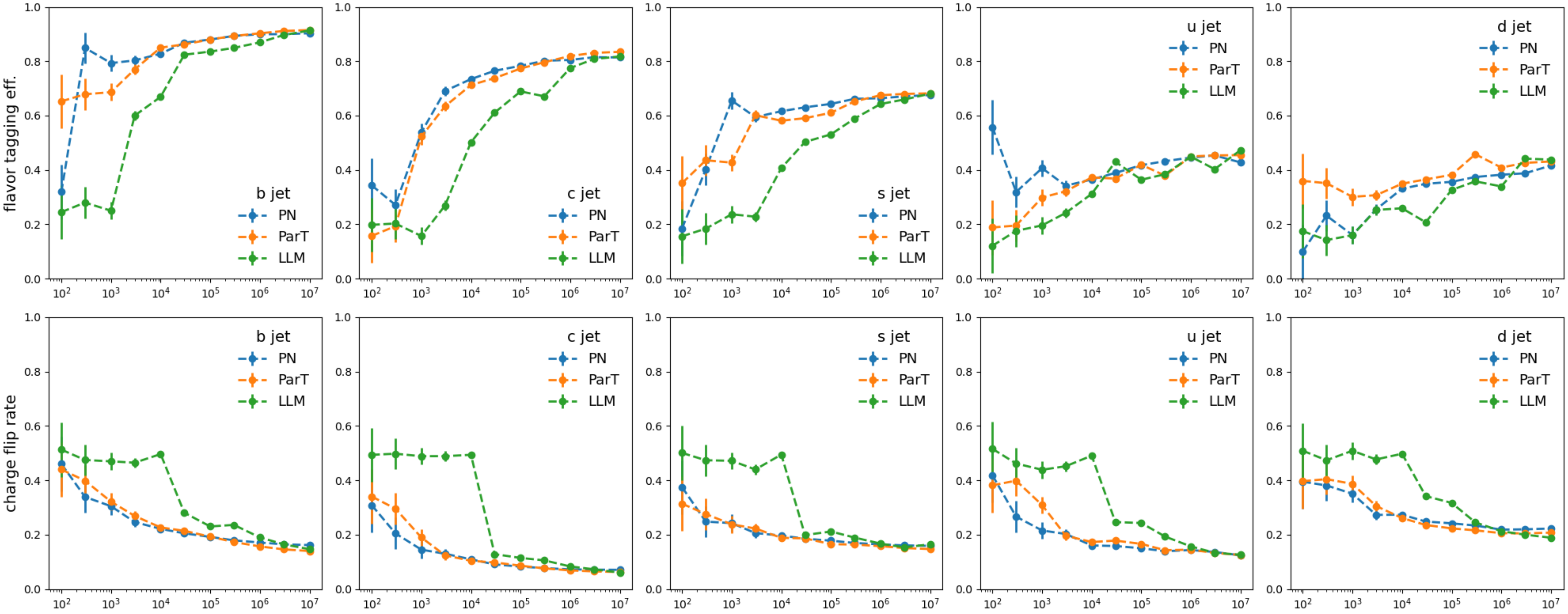
Figure 4: The BDT score distribution of signal and backgrounds in: the muon channel (left) and electron channel (right). The red curve indicates the projected statistical relative sensitivity estimated from Eq. 4.1 assuming a luminosity of 20 ab^{-1}.

compare ParticleNet, Particle Transformer, and LLM



the details will be shown in the upcoming manuscript in Arxiv

Scaling law on the volume of the dataset

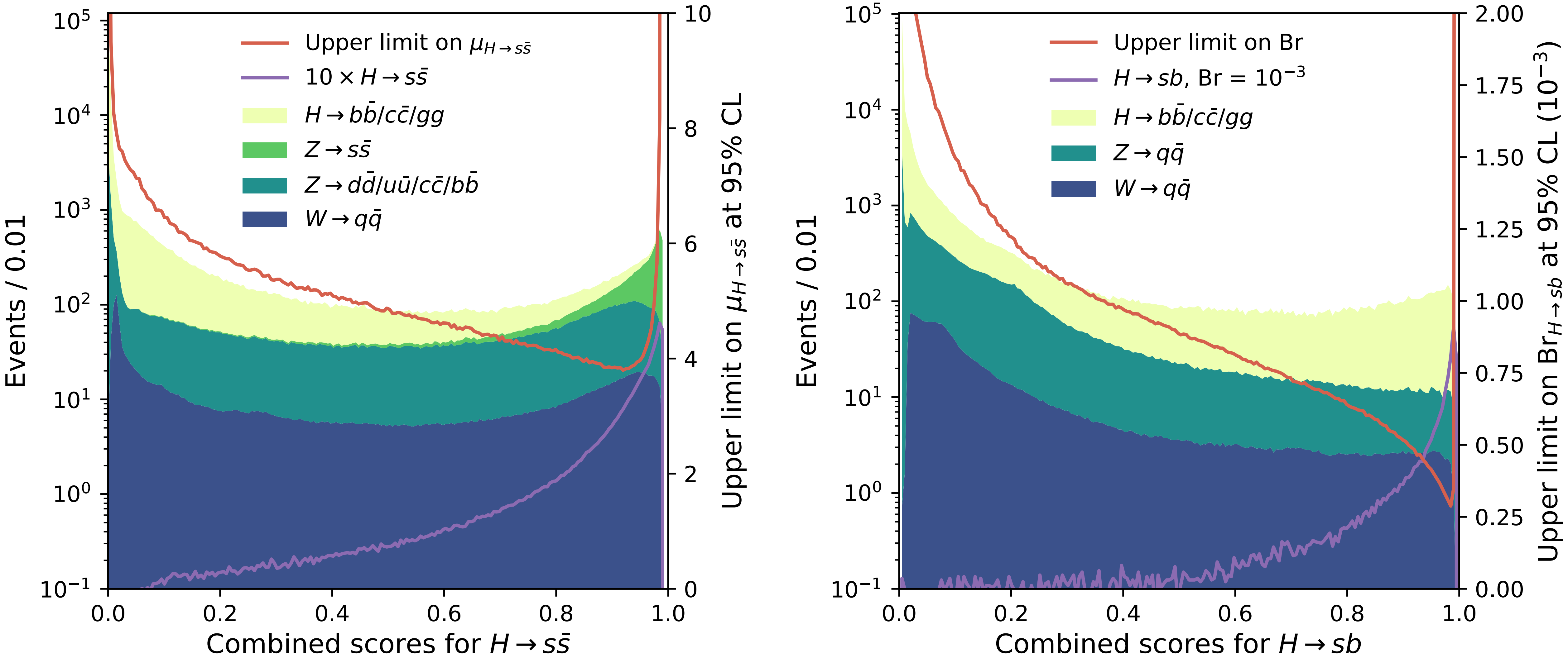


Summary

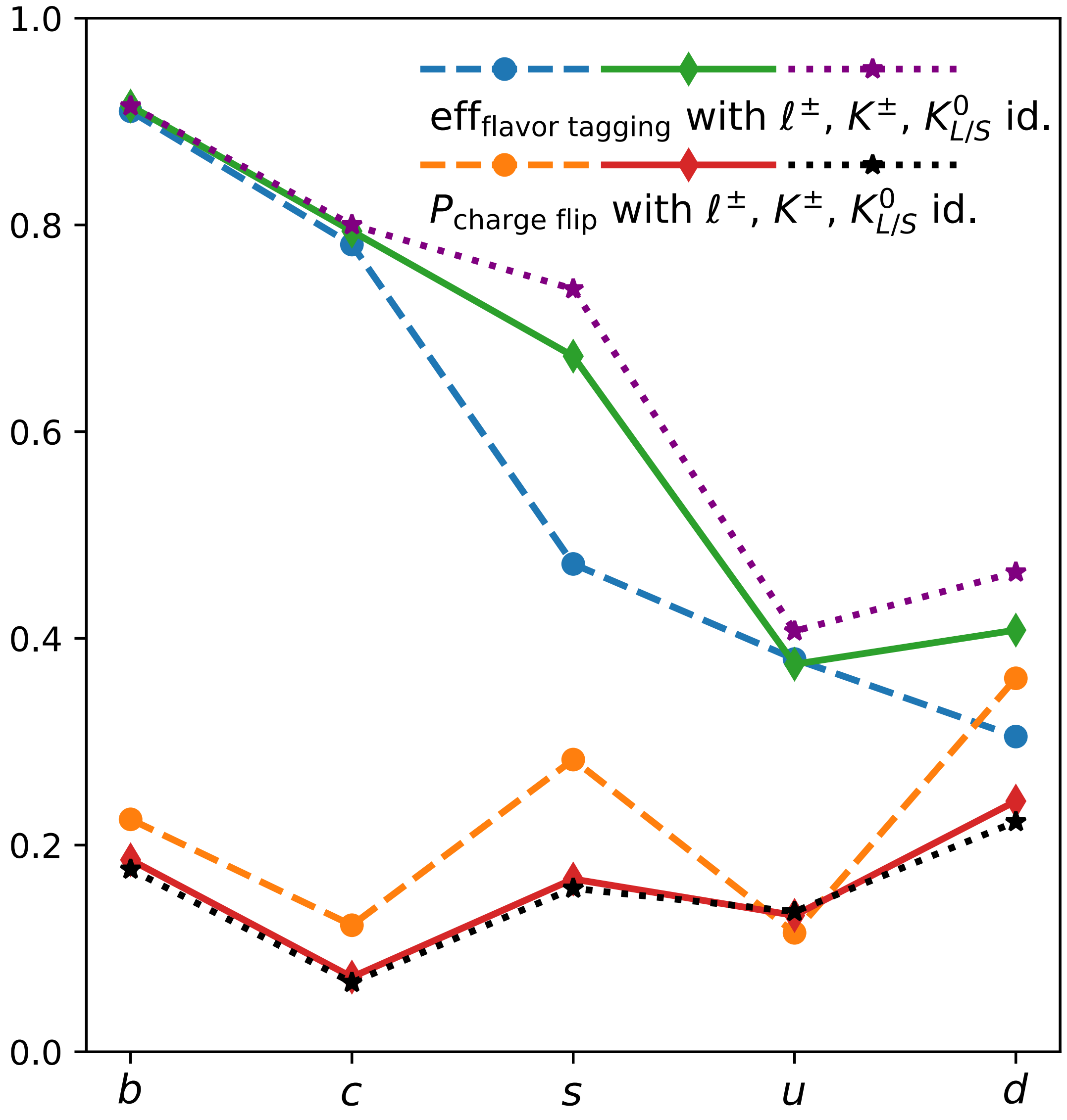
- We proposed and realized the JOI and achieved significant improvement in the measurement of Higgs rare and FCNC decay.
- The performance of JOI is stable versus various kinematic variables, physics processes, and hadronization models.
- The performance of JOI obtained with ParticleNet, Particle Transformer, and LLM is comparable to each other.

Many thanks !

application of JOI on Higgs rare and exotic decay



the performance dependence of JOI on input PID information



To understand the impact of PID, three scenarios are compared.

$\ell^{\pm}$: perfect identification of charged leptons ($\ell^{\pm}$)

$K^{\pm}$: perfect identification of the charged hadrons ($K^{\pm}$)

$K^0_{L/S}$: perfect identification of K_L and K_S.