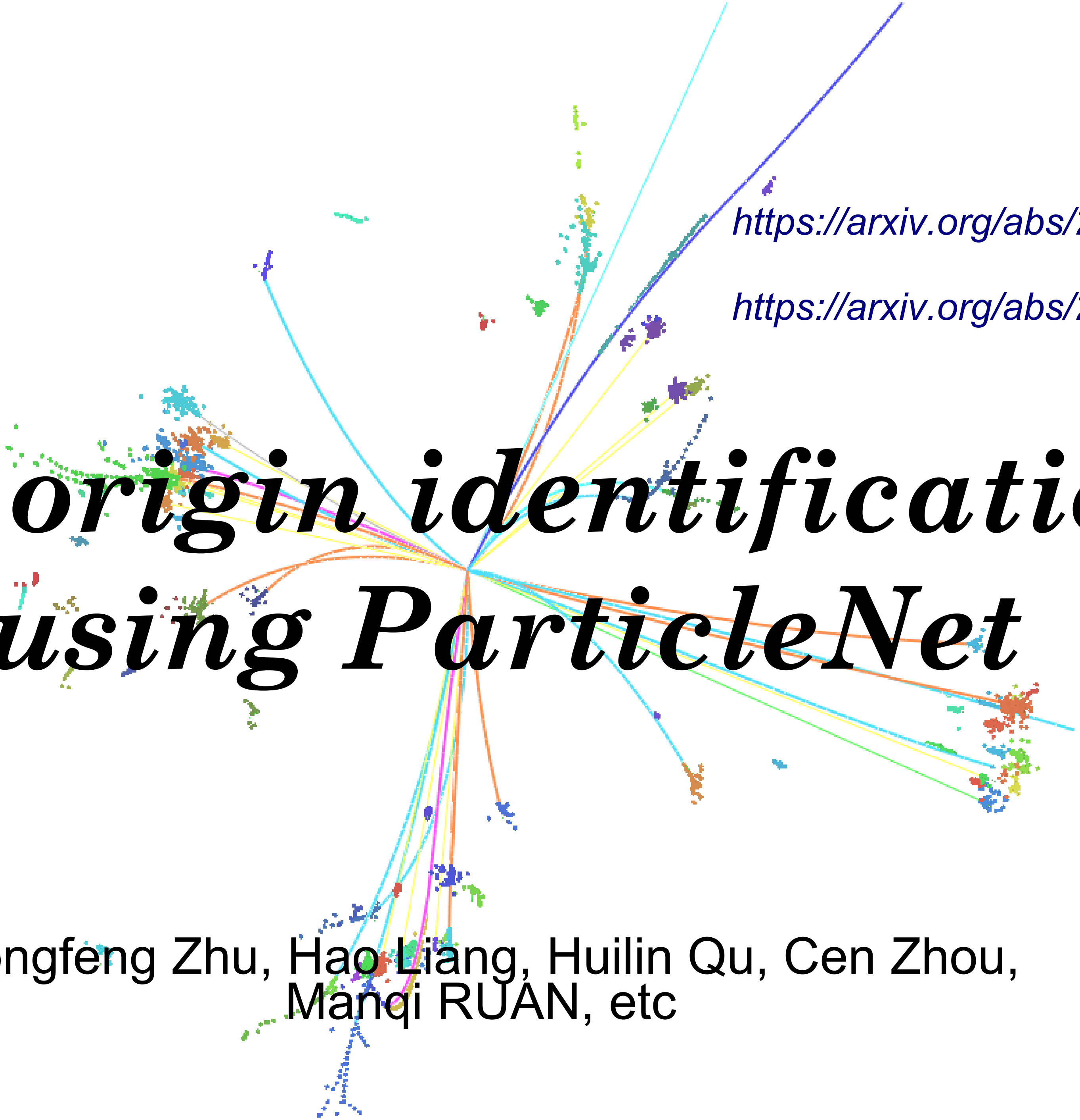


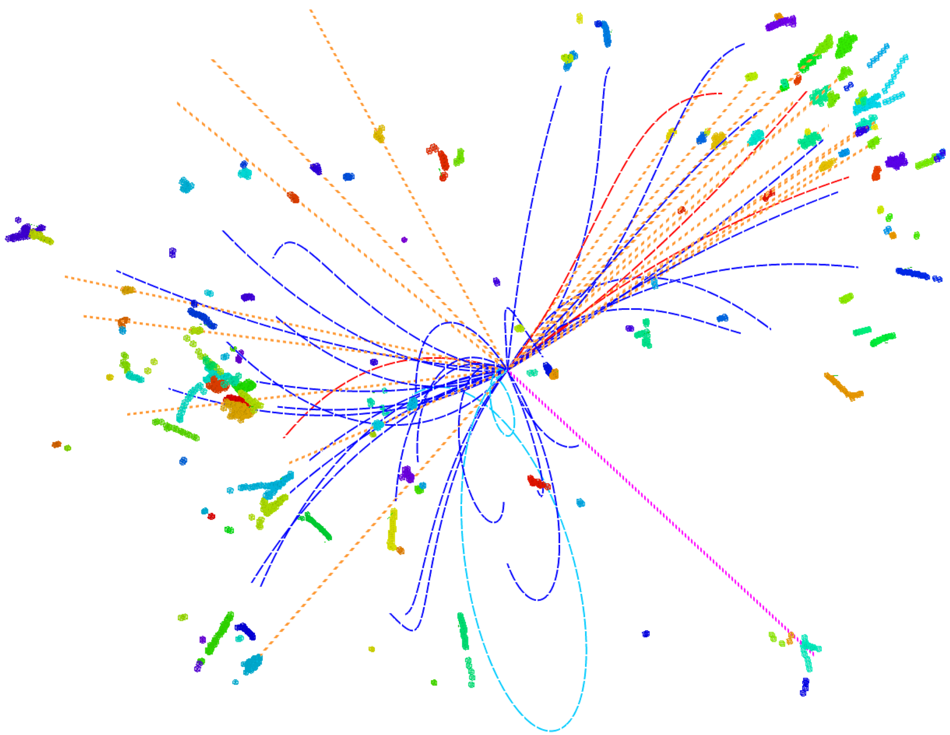
https://arxiv.org/abs/2310.03440

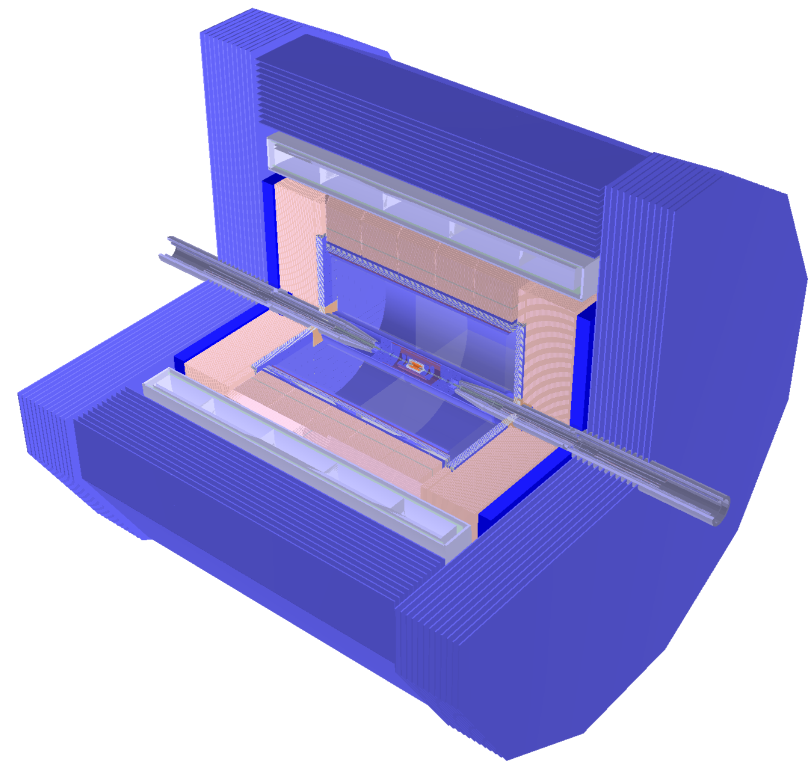
https://arxiv.org/abs/2309.13231

Jet origin identification using ParticleNet

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

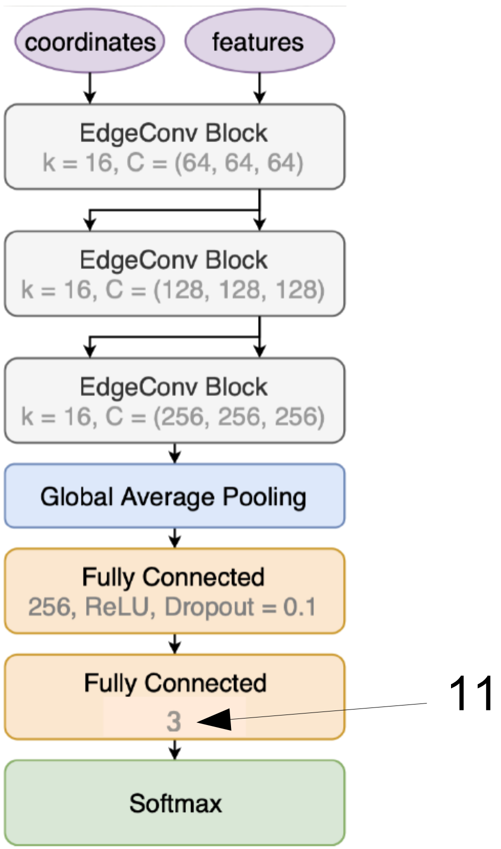
Geo. & Tools

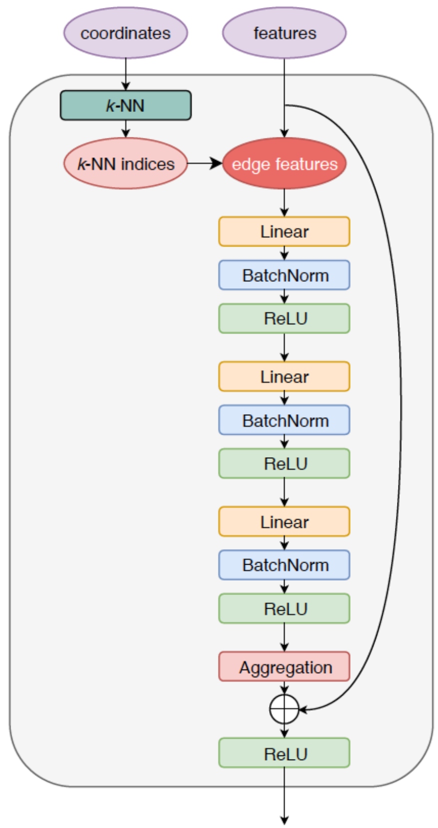




- **Jet origin identification: 11 categories (5 quarks + 5 anti quarks + gluon)**
  - Jet Flavor Tagging + Jet Charge measurements + s-tagging + gluon tagging...
- Full Simulated vvH, Higgs to two jets sample at CEPC baseline configuration: CEPC-v4 detector, reconstructed with Arbor.

Particle Net: IO





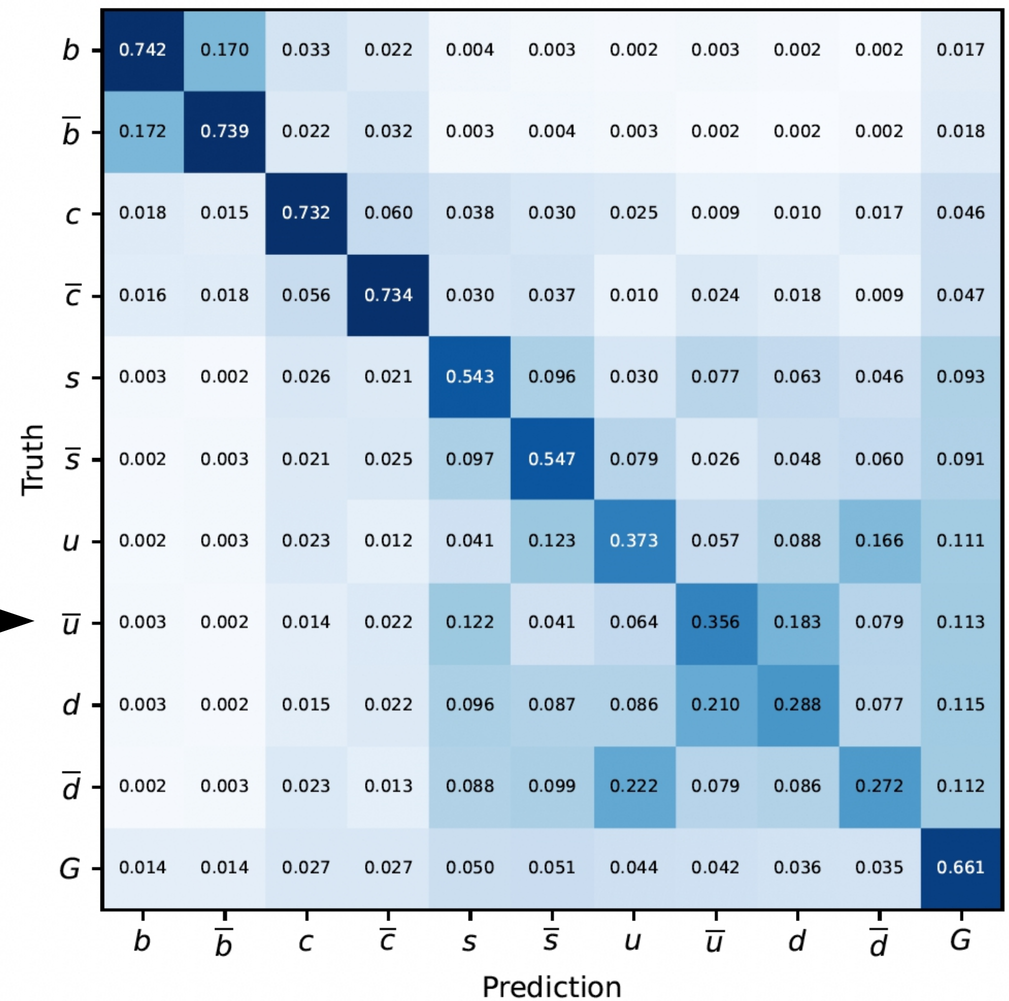
Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$logp_T$	logarithm of the particle's p_T
$logE$	logarithm of the particle's energy
$log\frac{p_T}{p_T(jet)}$	logarithm of 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.

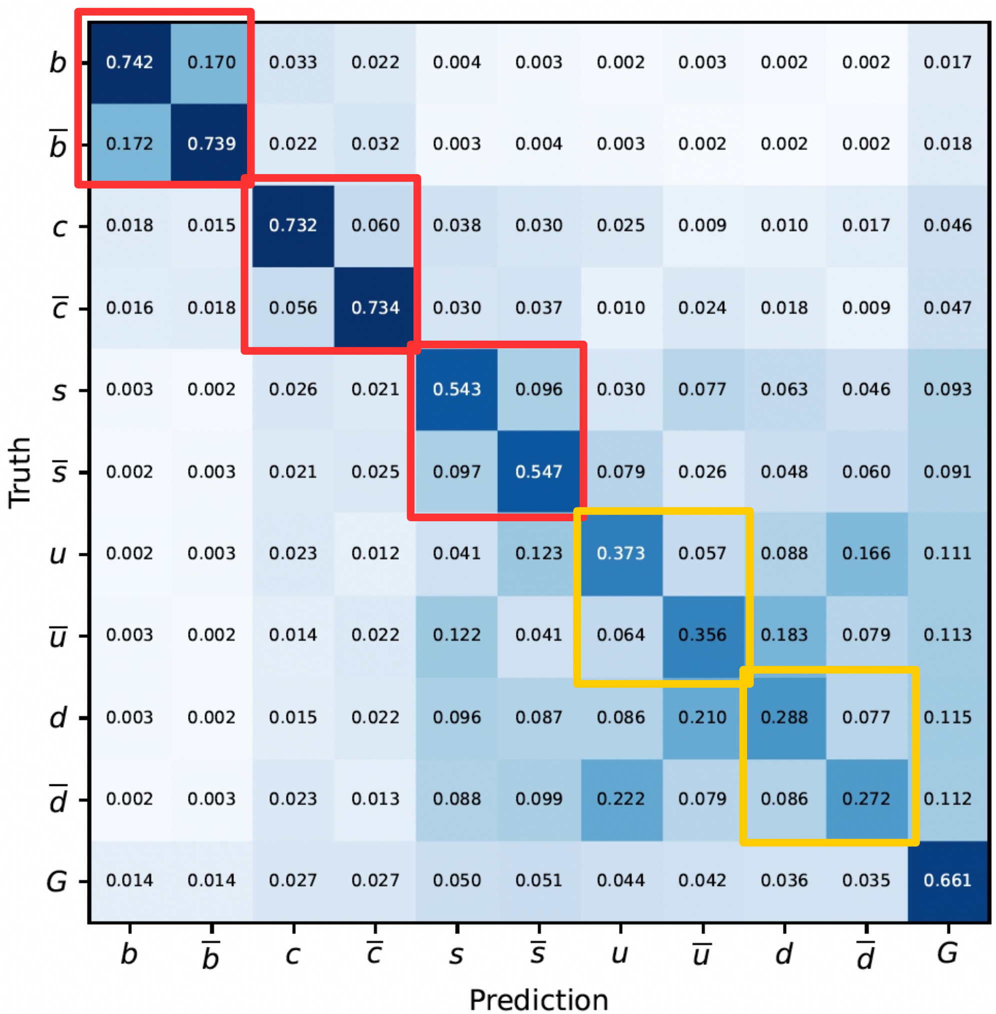
- Input: reco particles corresponding to 1 jet...
- Output: likelihoods to 11 different categories (sum =1)

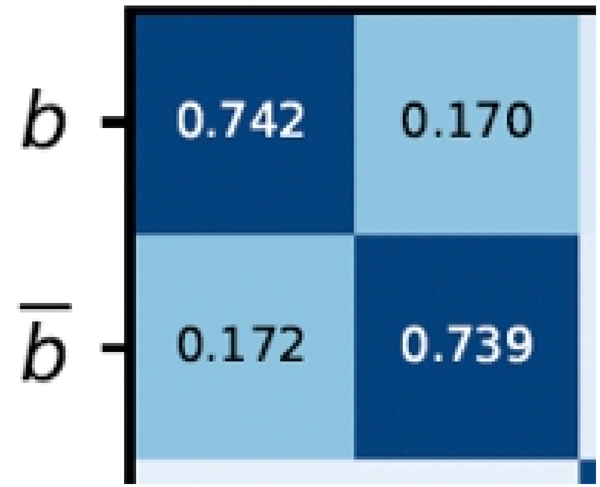
Jet origin id: 11 categories

- 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...



Jet origin id: 11 categories

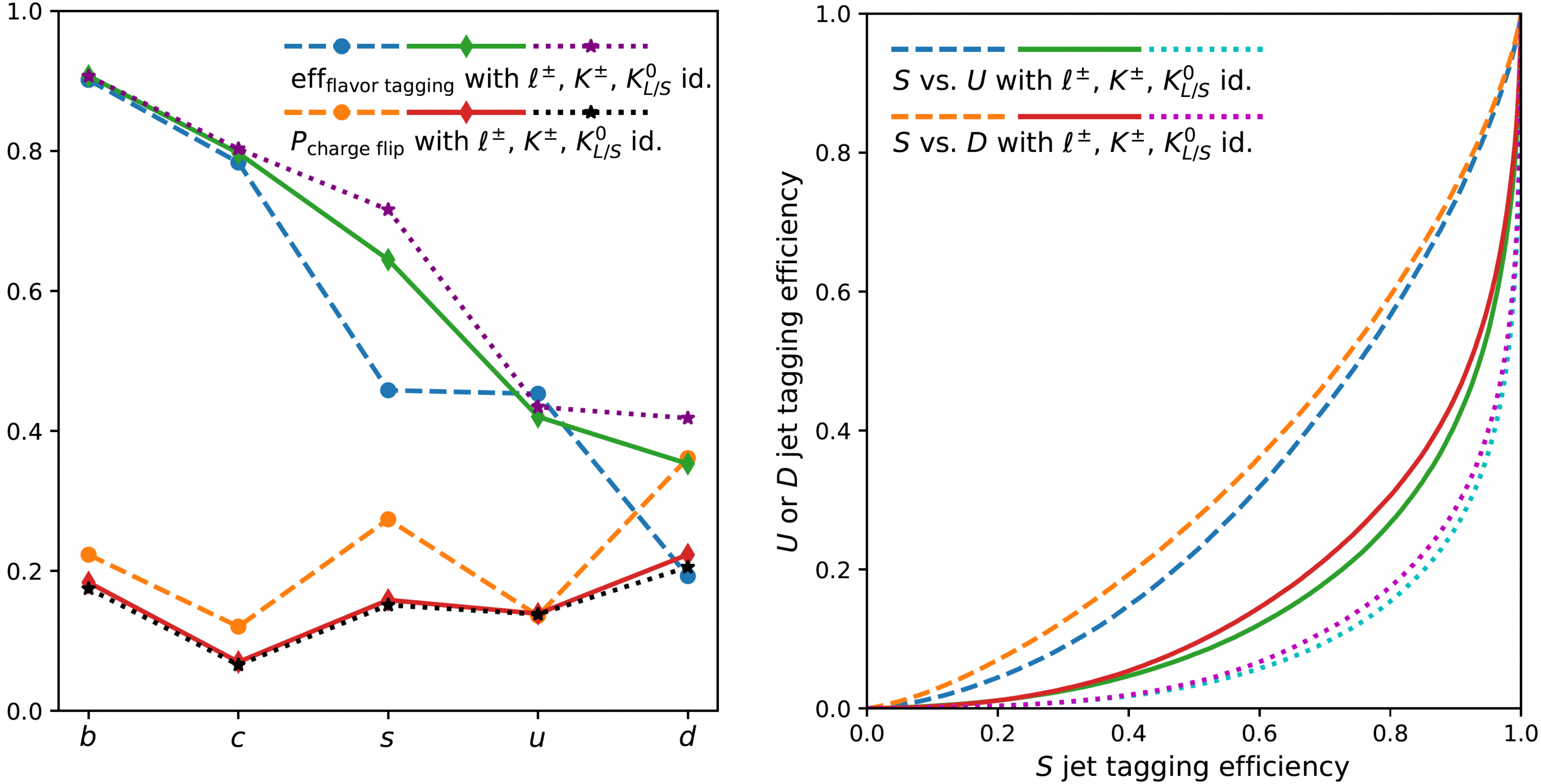




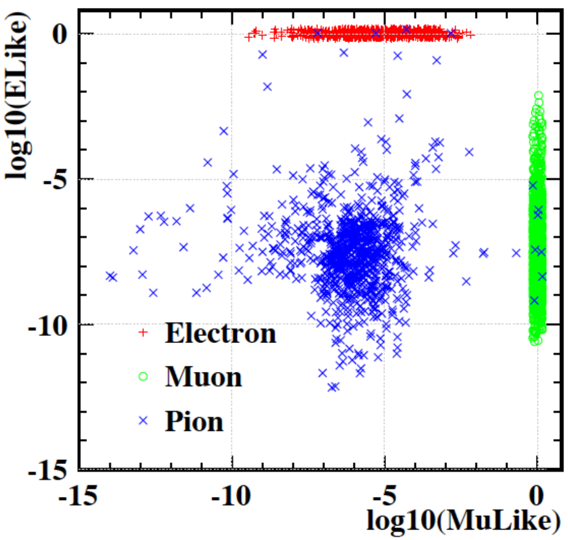
Eff = (0.74 + 0.17 + 0.74 + 0.17)/2 = 0.91

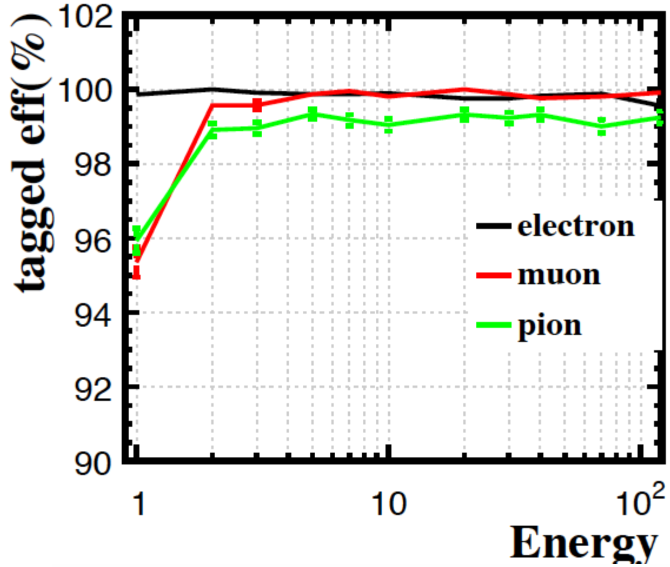
Charge flip rate = 0.17/0.91 = 0.19

Performance with different PID scenarios



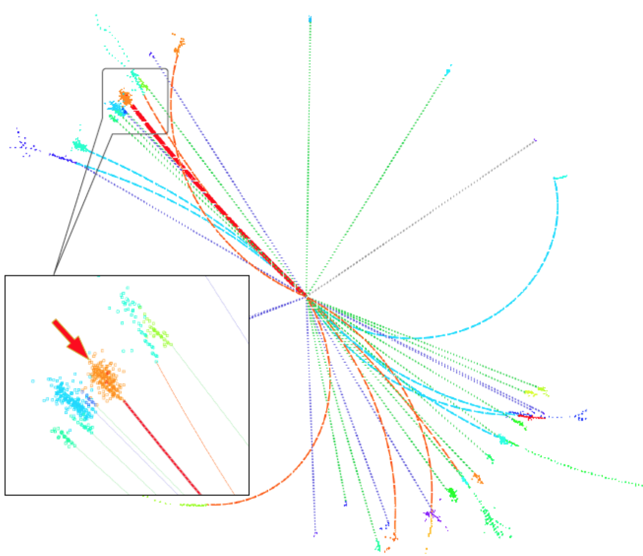
Lepton: isolated & Inside jet

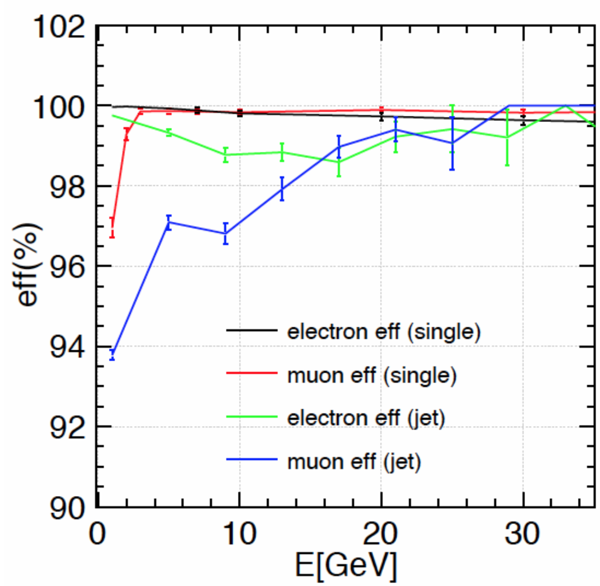


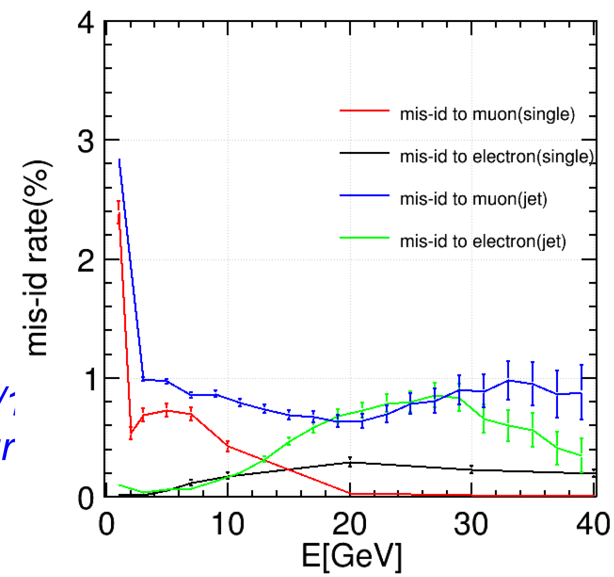


Isolated: for E > 2 GeV: lepton efficiency > 99.5% && Pion mis id rate ~ 1%

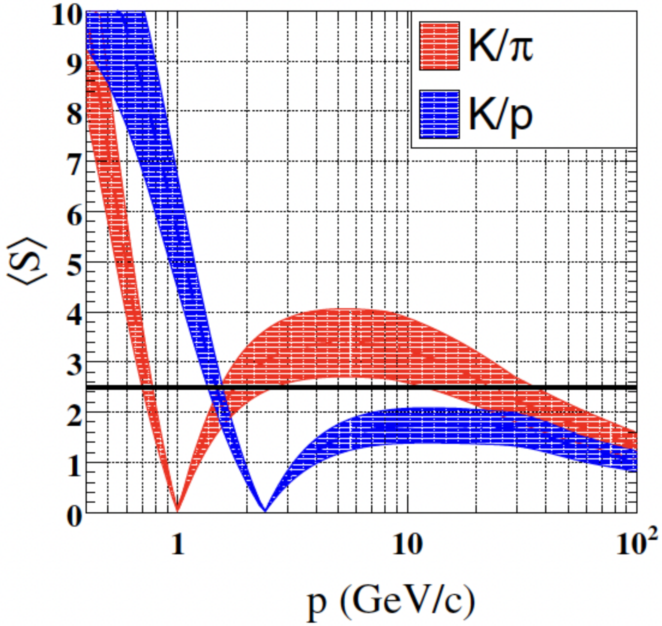
Inside Jet: percentage level degrading

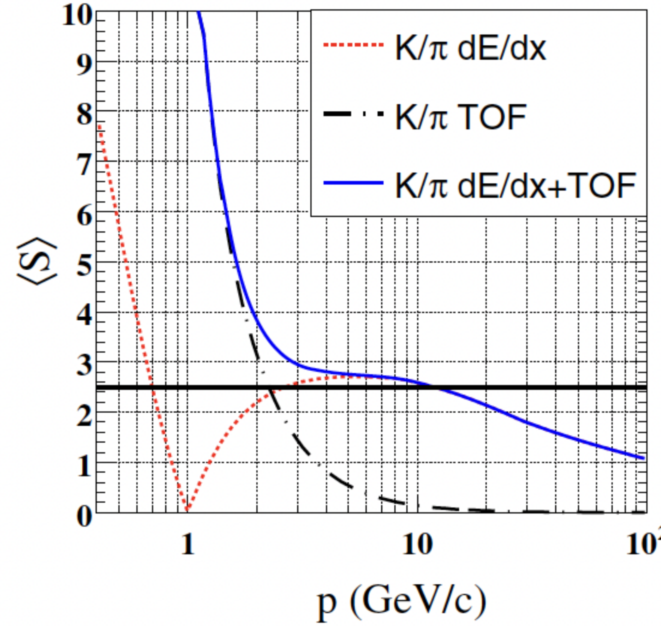


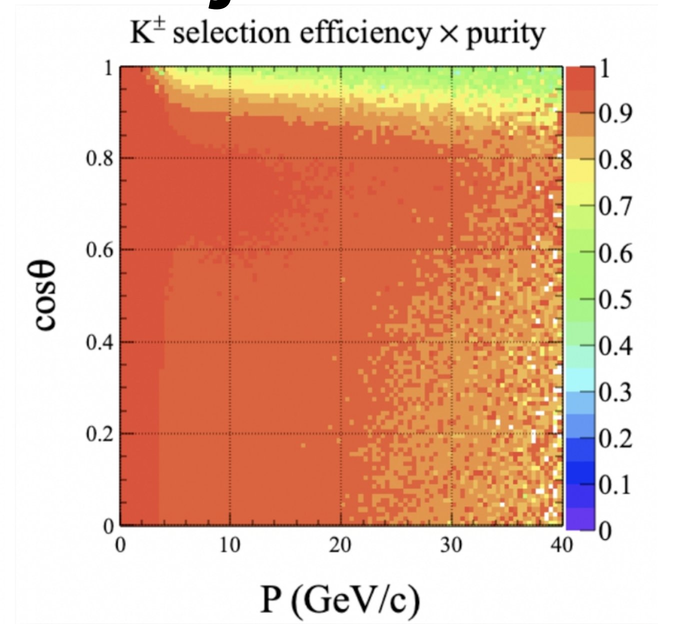




Charged Hadron ID: inside jet...







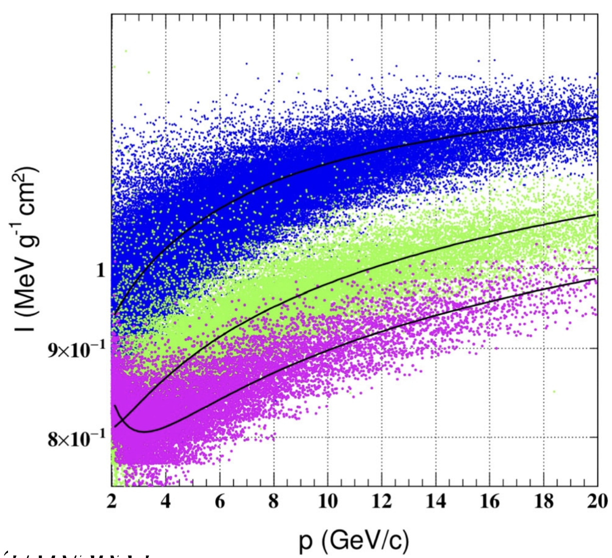


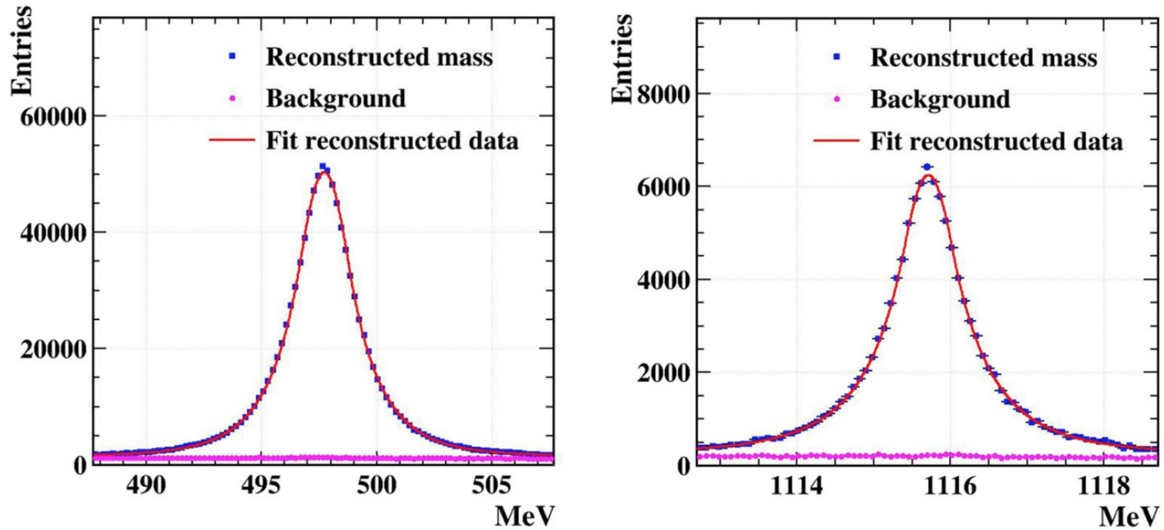
Table 3
The $K^{\pm}$ identification performance with different factors, $\sigma_{actual} = factor \cdot \sigma_{intrinsic}$, with/without combination of TOF information at the Z-pole.

	Factor	1.	1.2	1.5	2.
dE/dx	ε_K (%)	95.97	94.09	91.19	87.09
	$purity_K$ (%)	81.56	78.17	71.85	61.28
dE/dx & TOF	ε_K (%)	98.43	97.41	95.52	92.3
	$purity_K$ (%)	97.89	96.31	93.25	87.33

- Pid via dEdx or dNdx: < 3% in barrel region for GeV hadron
- Pid at Drift Chamber using dN/dx: even better performance

Kshort & Lambda

Eur. Phys. J. Plus (2020) 135:274 Page 9 of 14 274

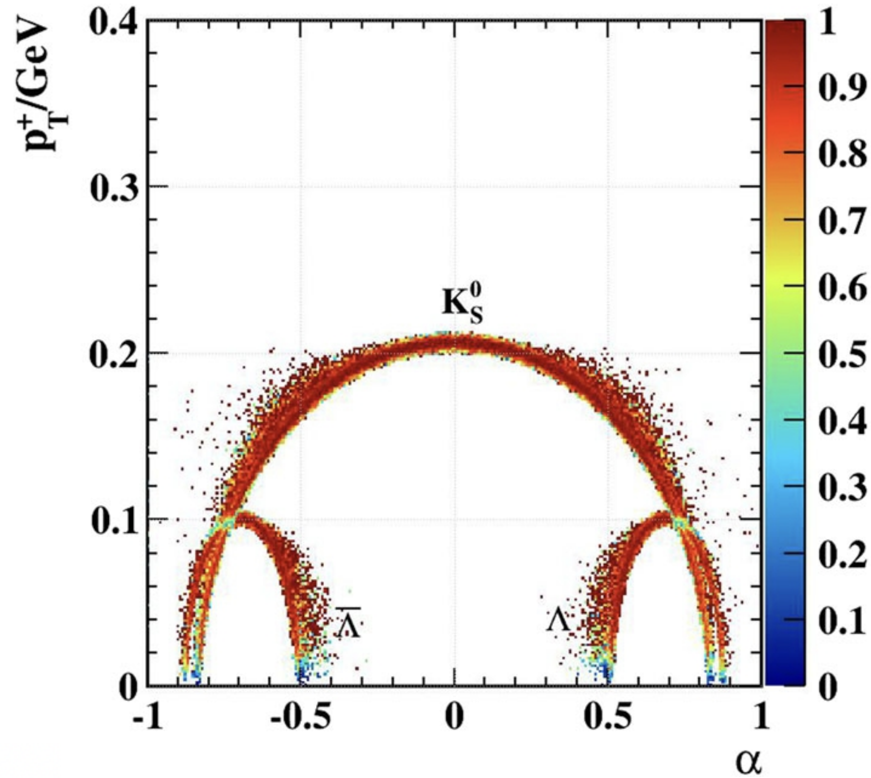


(a) K_S^0 (b) Λ

Fig. 7 All reconstructed mass distributions of K_S^0 and Λ. They are fitted with double-sided crystal ball functions

Table 3 K_S^0 and Λ reconstruction performance

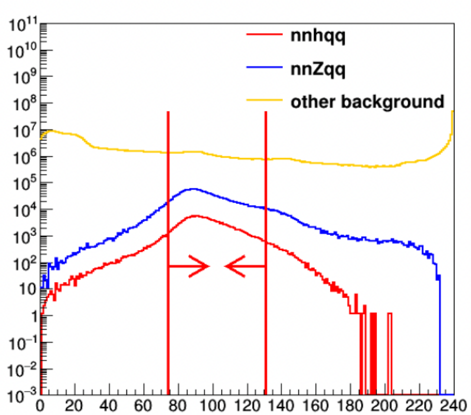
Particle	K_S^0 (%)	Λ (%)
ϵ_R	81.3	70.1
ϵ_T	40.6	27.3
P	92.4%	86.4%
$\epsilon_R \cdot P$	0.751	0.606
$\epsilon_T \cdot P$	0.375	0.236



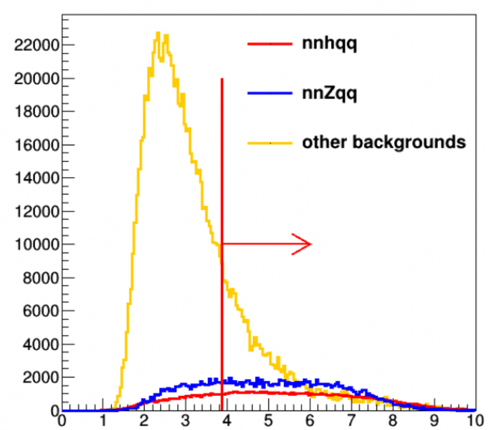
High eff/purity reco. of charged Final states at least...

Impact on benchmark: ννH, H→jets

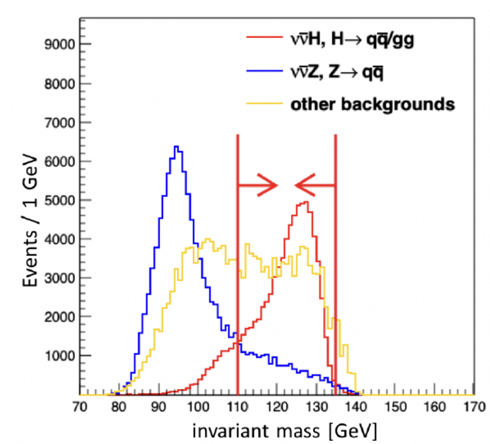
	$\nu\nu Hq\bar{q}/gg$	2f	SW	SZ	WW	ZZ	Mixed	ZH	$\frac{\sqrt{S+B}}{S}$ (%)
total	178890	8.01E8	1.95E7	9.07E6	5.08E7	6.39E6	2.18E7	961606	16.86
recoilMass (GeV) $\in (74, 131)$	157822	5.11E7	2.17E6	1.38E6	4.78E6	1.30E6	1.08E6	74991	4.99
visEn (GeV) $\in (109, 143)$	142918	2.37E7	1.35E6	8.81E5	3.60E6	1.03E6	6.29E5	50989	3.92
leadLepEn (GeV) $\in (0, 42)$	141926	2.08E7	3.65E5	7.24E5	2.81E6	9.72E5	1.34E5	46963	3.59
multiplicity $\in (40, 130)$	139545	1.66E7	2.36E5	5.24E5	2.62E6	9.07E5	4977	42751	3.29
leadNeuEn (GeV) $\in (0, 41)$	138653	1.46E7	2.24E5	4.72E5	2.49E6	8.69E5	4552	42303	3.12
Pt (GeV) $\in (20, 60)$	121212	248715	1.56E5	2.48E5	1.51E6	4.31E5	999	35453	1.37
Pl (GeV) $\in (0, 50)$	118109	52784	1.05E5	74936	7.30E5	1.13E5	847	34279	0.94
-log10(Y23) $\in (3.375, +\infty)$	96156	40861	26088	60349	2.25E5	82560	640	10691	0.76
InvMass (GeV) $\in (116, 134)$	71758	22200	11059	6308	77912	13680	248	6915	0.64
BDT $\in (-0.02, 1)$	60887	9140	266	2521	3761	3916	58	1897	0.47



74 < recoil mass < 131

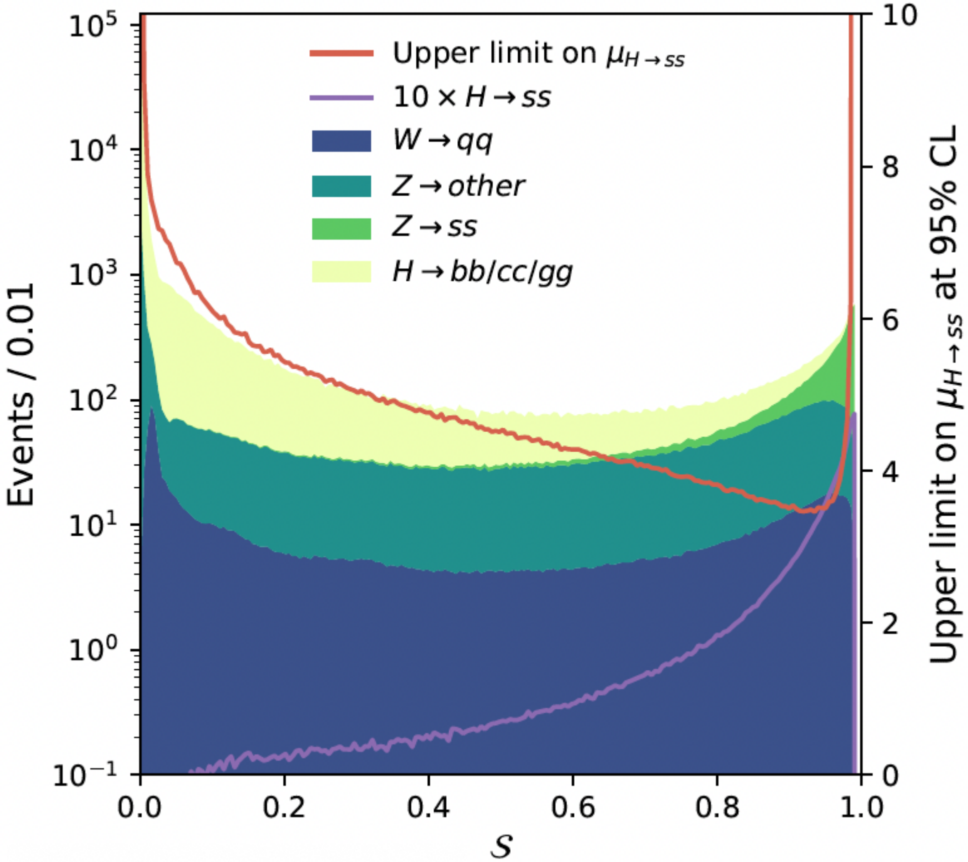


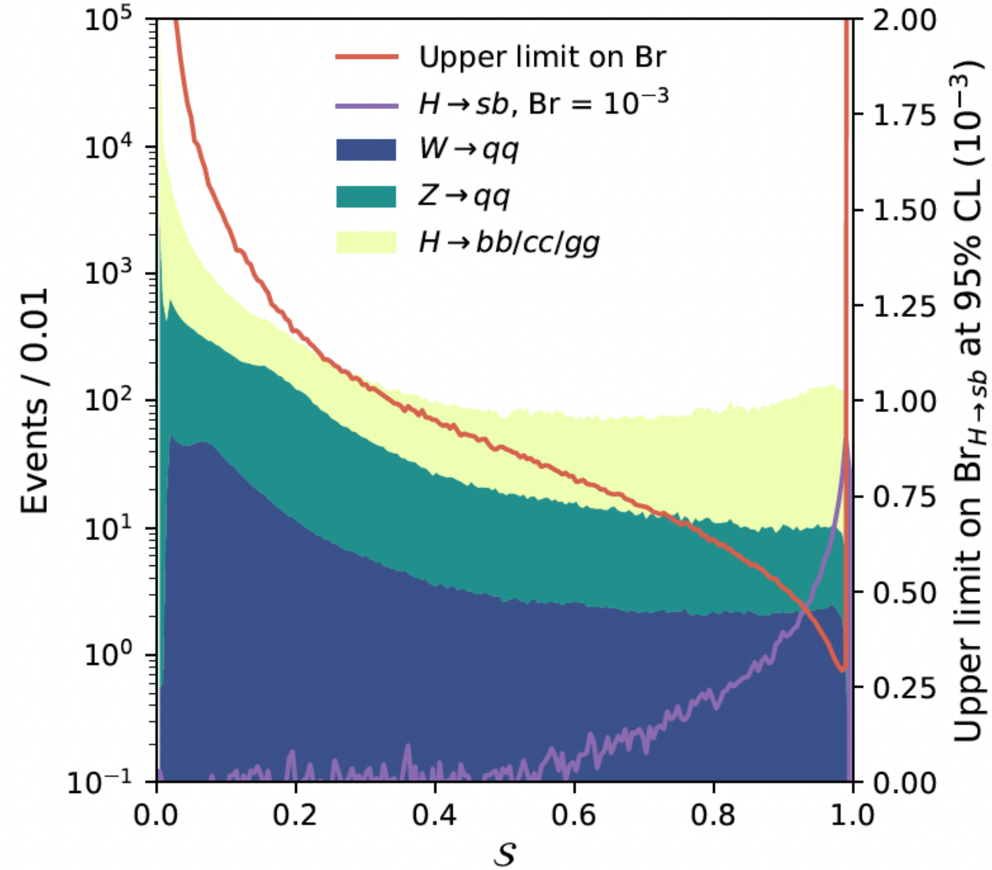
-log(Y23) > 3.875

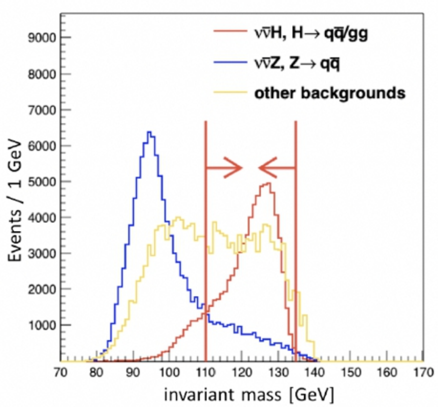


Thanks to BMR ~ 3.8%!

Benchmark analyses using Jet origin ID

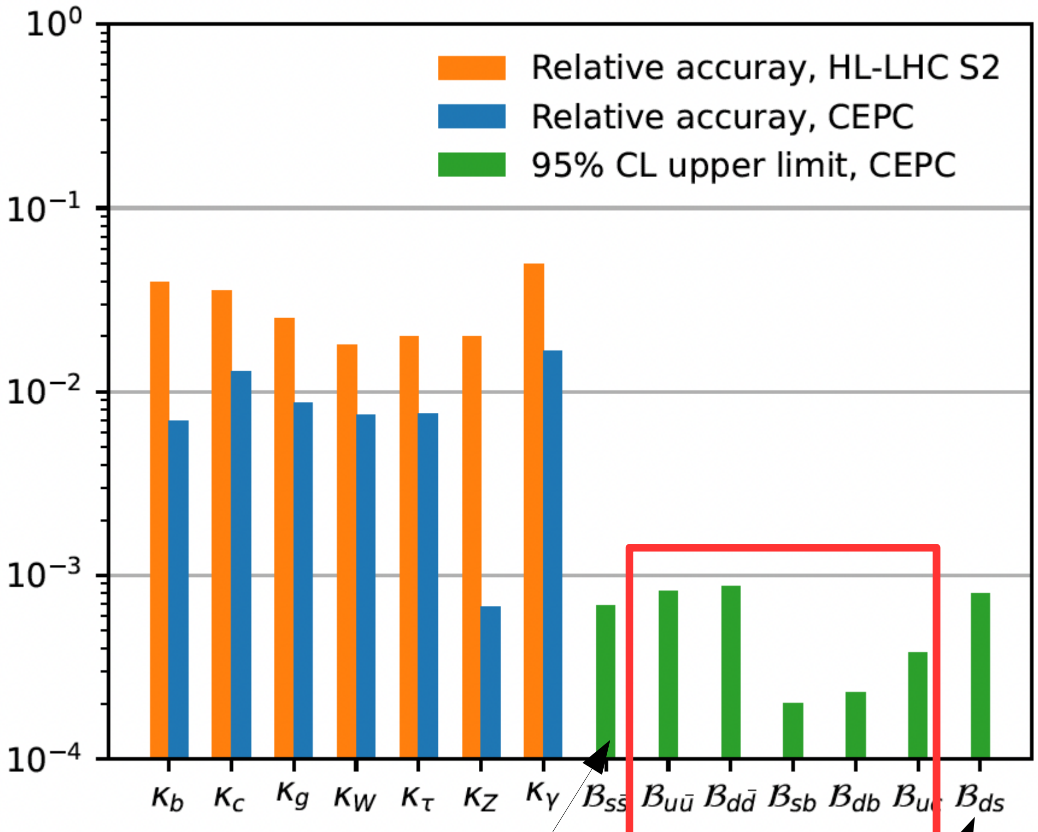






Applied to quasi-data of vvH;
H → ss: be limited to 3*SM using vvH + llH at 20 iab
H → sb: up limit of 2E-4 at 95% C.L.

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.

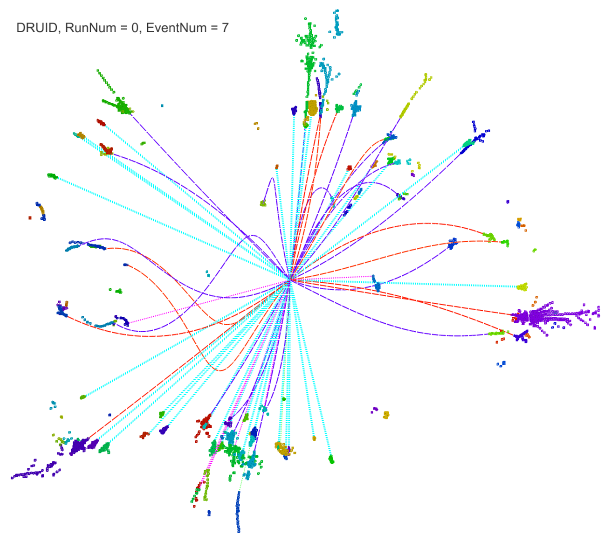
[50] Alexander Albert et al. Strange quark as a probe for new physics in the Higgs sector. In *Snowmass 2021*, 3 2022.

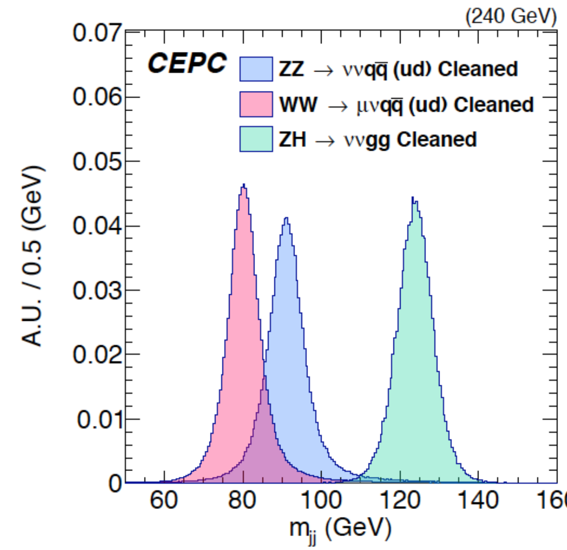
[59] J. de Blas et al. Higgs Boson Studies at Future Particle Colliders. *JHEP*, 01:139, 2020.

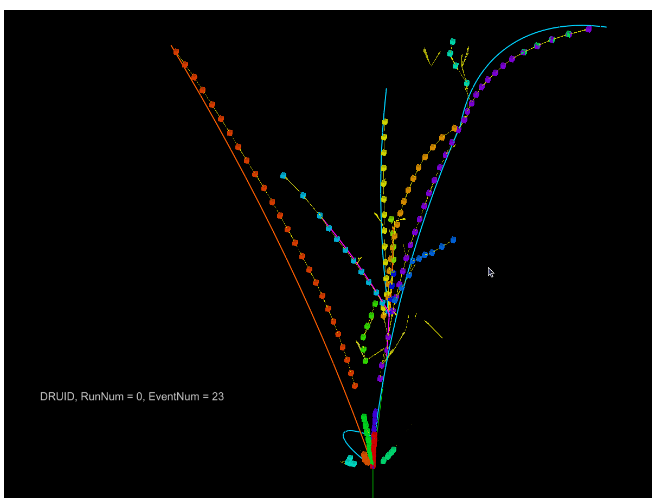
[60] Jorge De Blas, Gauthier Durieux, Christophe Grojean, Jiayin Gu, and Ayan Paul. On the future of Higgs, electroweak and diboson measurements at lepton colliders. *JHEP*, 12:117, 2019.

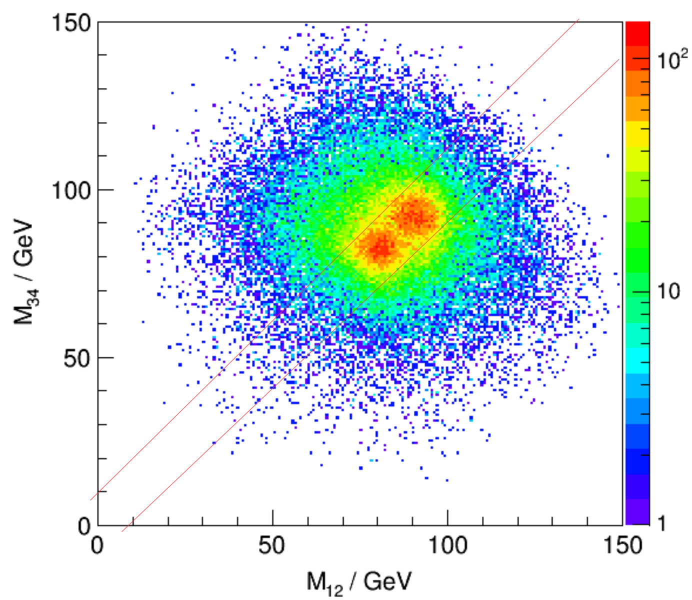
Critical Problems to be addressed

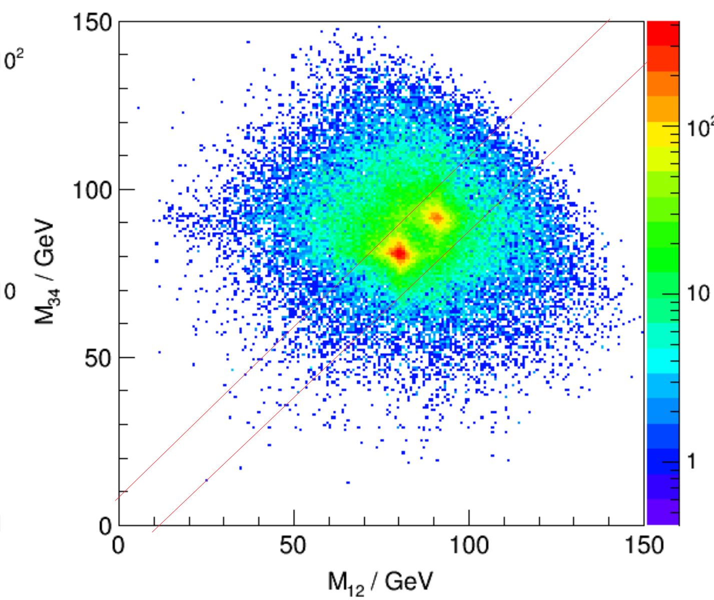
- Particle Flow
- Color Singlet Identification
- QCD – hadronization,
- Beyond jets...











Summary

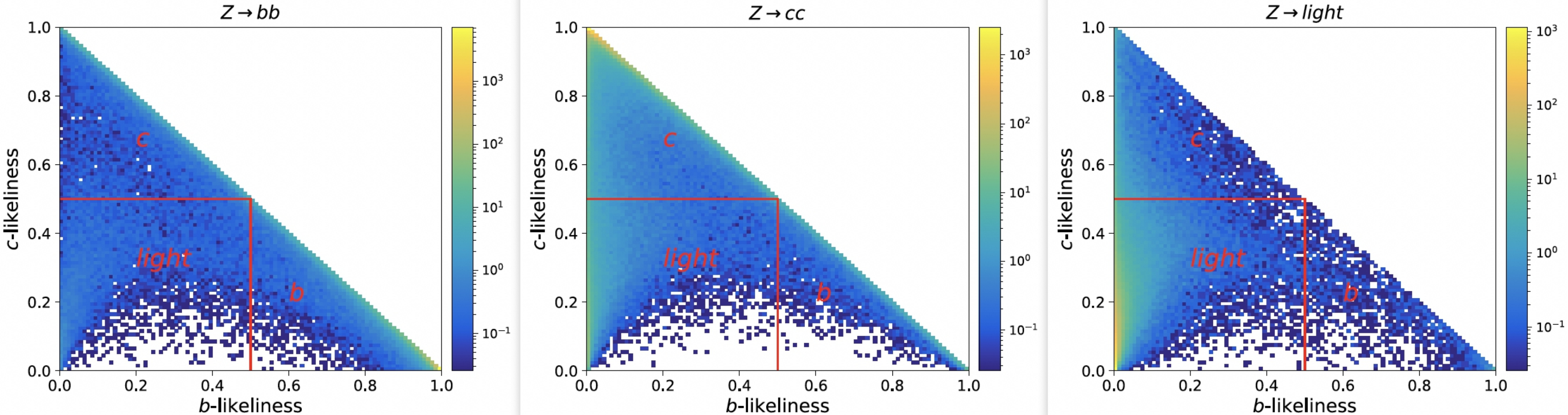
- PFA oriented detector design ~ CALICE laid solid foundation for the excellent reco/measurement at high energy frontier, especially with hadronic events at electron positron Higgs factories.
  - Better BMR shall always be pursues,
  - To be in cope with beam background & event rates,
  - Provide Pid: 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 (ννH + 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

Summary

- 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...

Backup

Three categories: b, c, & light



Hadronic Z pole sample

1 M Z→bb, cc, (uds) each

60/20/20% for training/validating/testing.

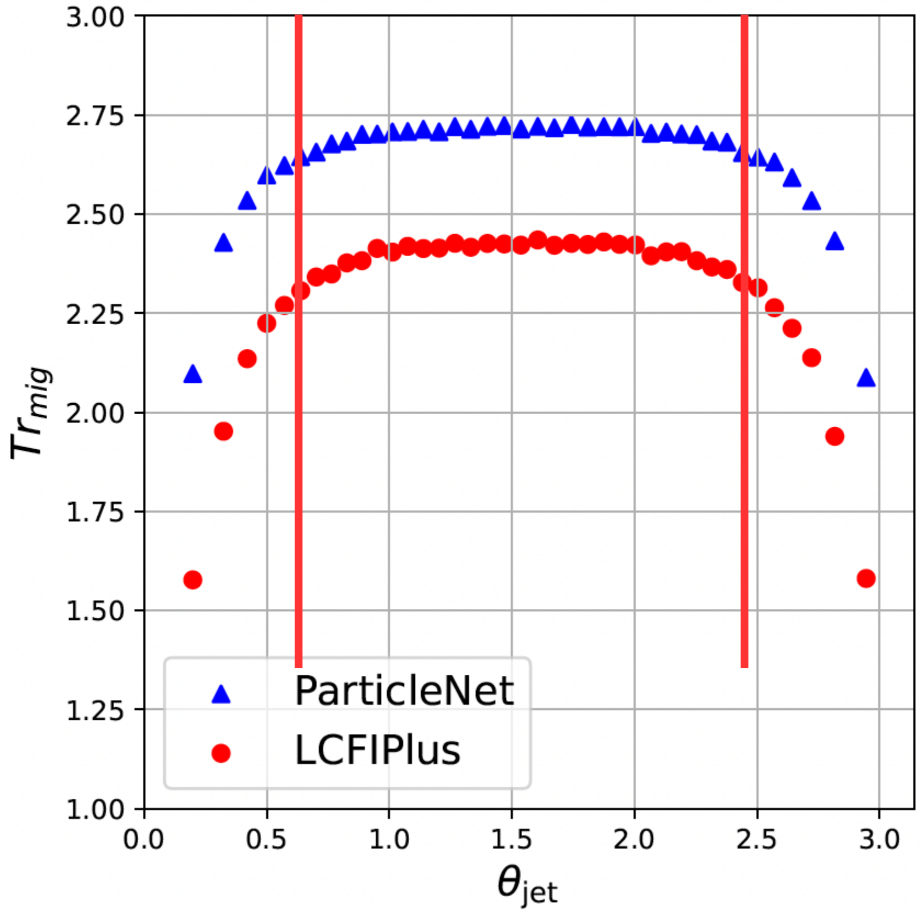
Result on Testing sample

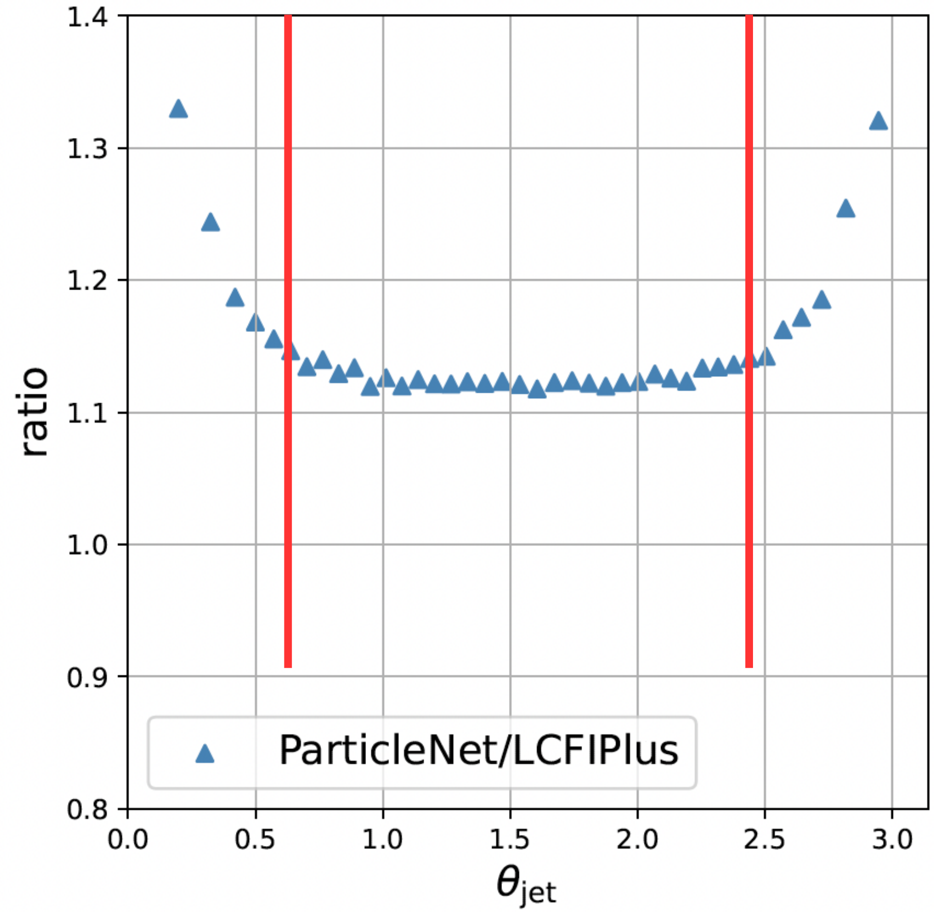
truth \ predicted	b	c	uds
b	0.911	0.059	0.031
c	0.039	0.784	0.177
uds	0.005	0.051	0.944

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

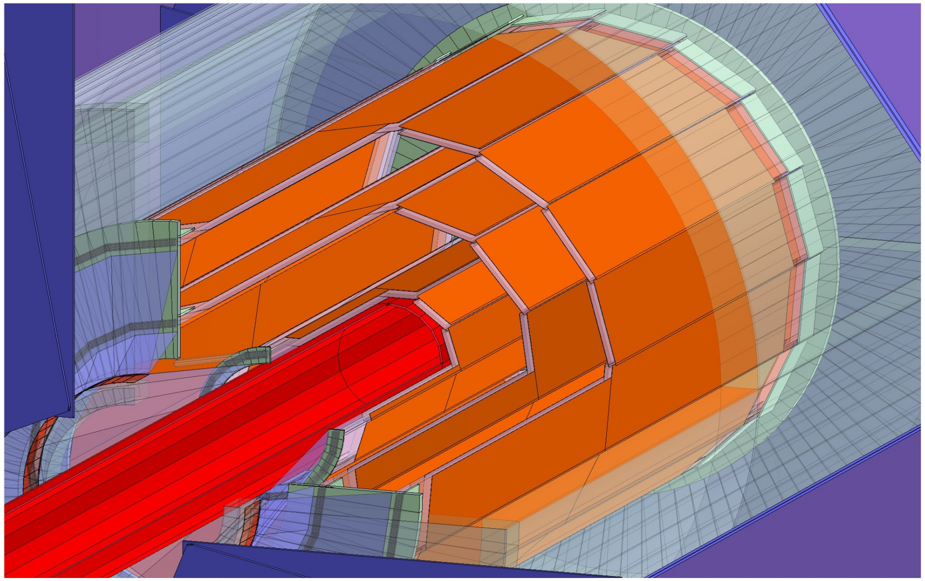
Figure 7. The migration matrix of ParticleNet (left) and LCFIPlus (right) at the CEPC.

Dependence on polar angle



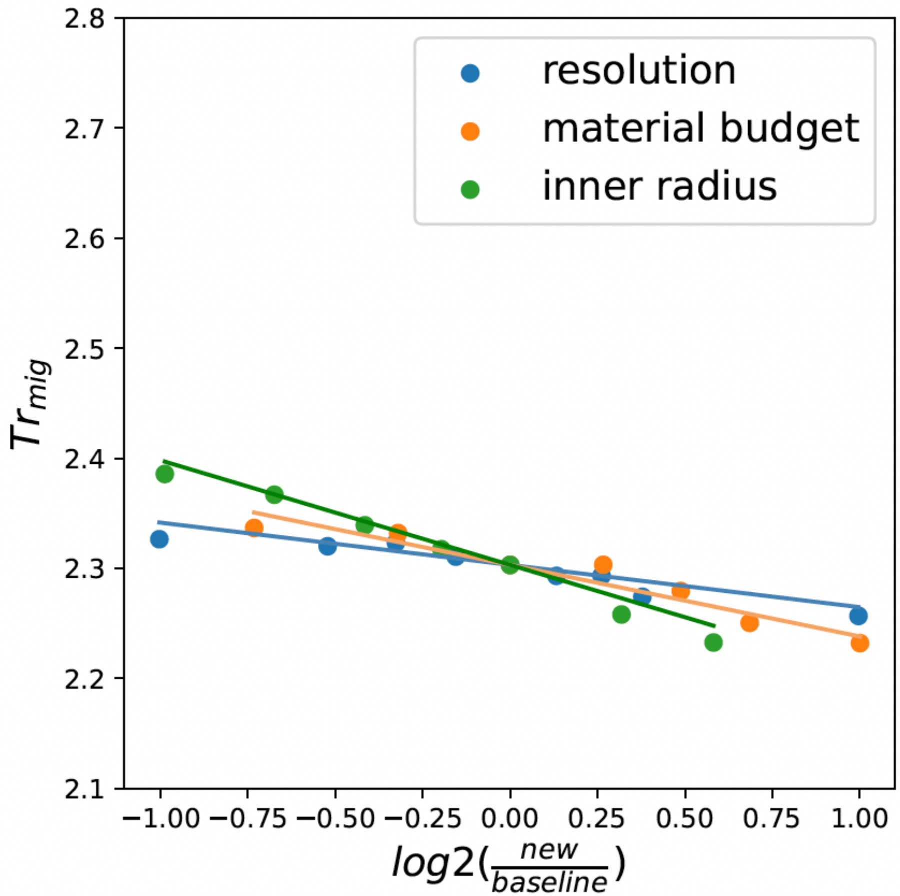


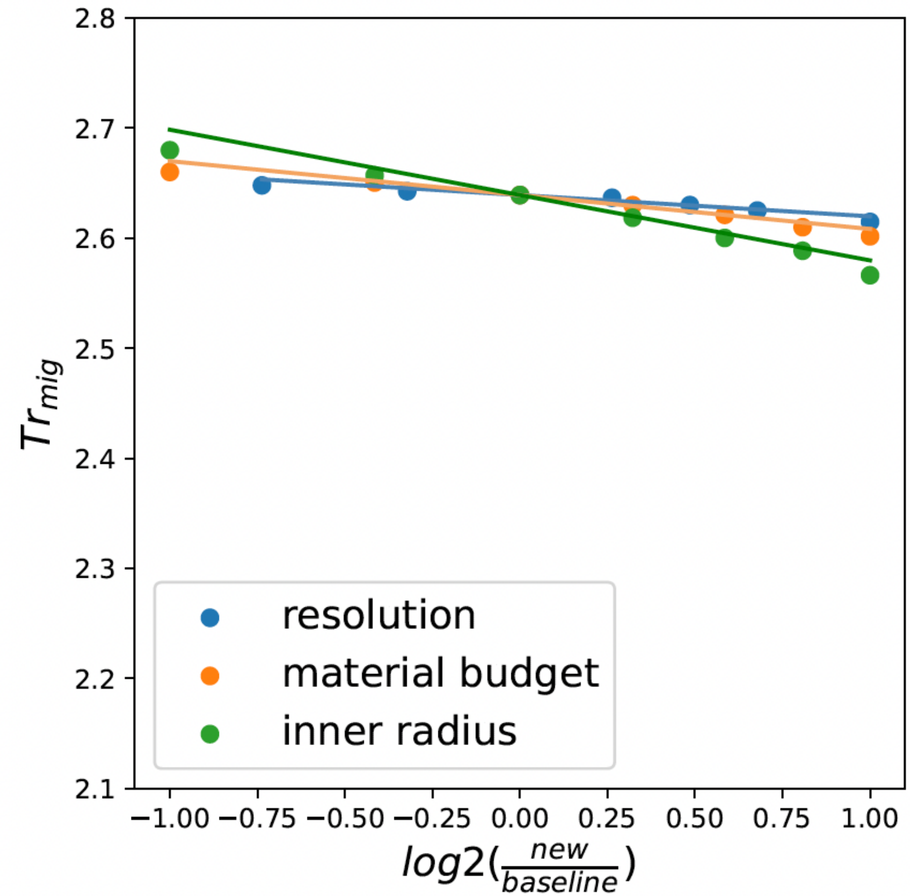
Comparison on Det. Optimization



	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

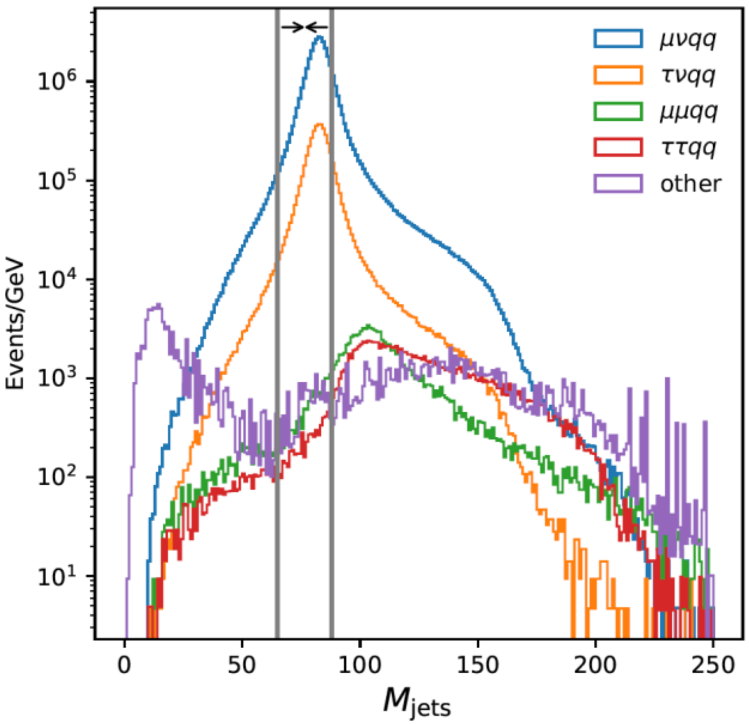




$$Tr_{mig} = 2.30 + 0.06 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2 \frac{R^0_{radius}}{R_{radius}} \quad (4.1)$$

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

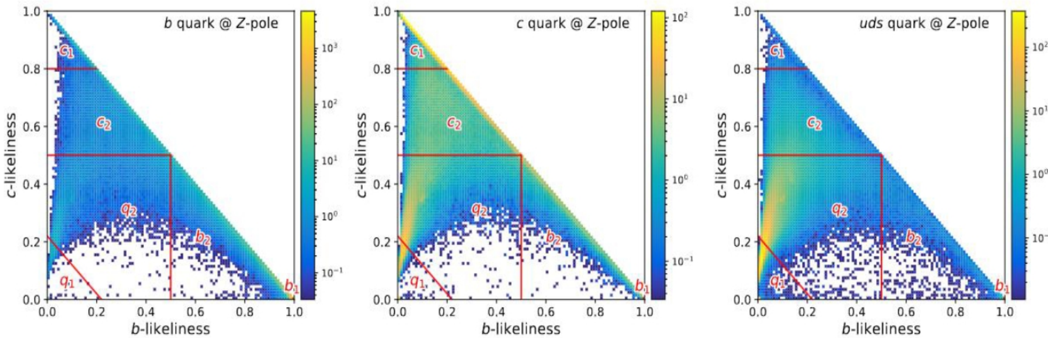
Vcb from W decay



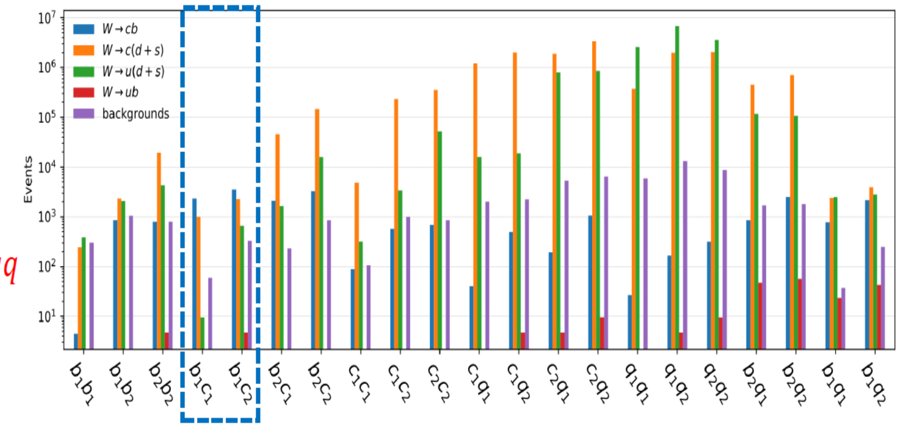
	$\mu\nu W, W \to$				$\tau(\mu\nu)\nu_\tau W, W \to$				$\tau\nu_\tau qq, \tau \to$					
	cb	ub	$c(d/s)$	$u(d/s)$	cb	ub	$c(d/s)$	$u(d/s)$	$e2\nu$	had.ν_τ	$\tau\tau qq$	$\mu\mu qq$	Higgs	others
w/o slections	40.3K	363	24.2M	24.2M	7.73K	74	4.2M	4.2M	8.66M	31.4M	2.18M	4.47M	4.07M	2.06G
$E_{\mathrm{L}\mu} > 12\mathrm{GeV}$	37.9K	330	22.6M	22.6M	5.59K	56	2.98M	2.97M	133K	687K	422K	2.82M	645K	186.3M
$R_{\mathrm{L}\mu} > 0.85$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$\cos(\theta_{\mathrm{L}\mu})$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$q_{\mathrm{L}\mu}\cos(\theta_{\mathrm{L}\mu}) < 0.20$	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
$M_{\mathrm{L}\mu\mathrm{S}\mu} < 75$ 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	50.6	53.5	53.7	34.9	25.0	36.7	36.9	0.0	0.1	0.1	0.1	0.0	0.0
	(0.7)	(8.1)	(0.0)	(0.0)	(1.5)	(12.5)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$b_1c_{1,2}$	5.14K	4	2.79K	571	632	0	407	65	0	14	67	228	0	0
$\epsilon_{b_1c_{1,2}}$ (%)	12.8	1.3	0.0	0.0	8.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(0.4)	(1.3)	(0.0)	(0.0)	(0.7)	(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:
  - $W \to c(d/s)$
  - $\mu\mu qq$

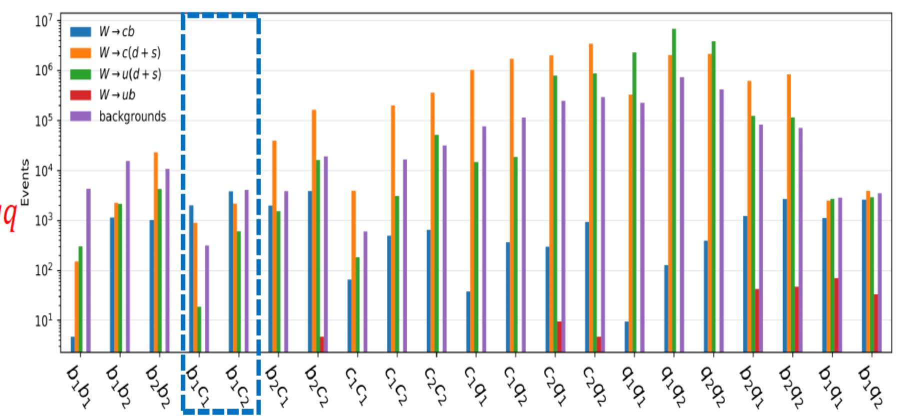
Vcb from W decay



$\mu\nu qq$



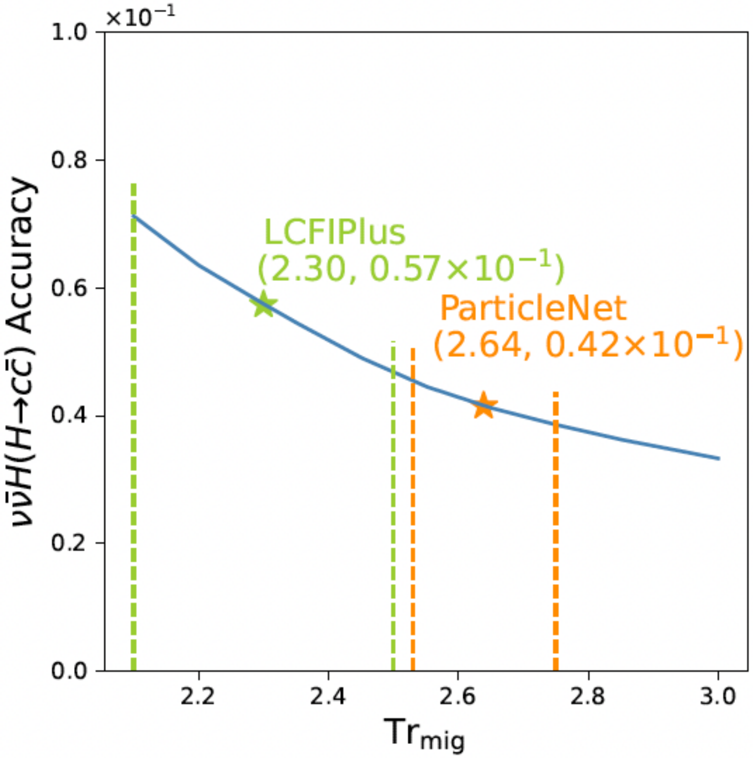
$e\nu qq$

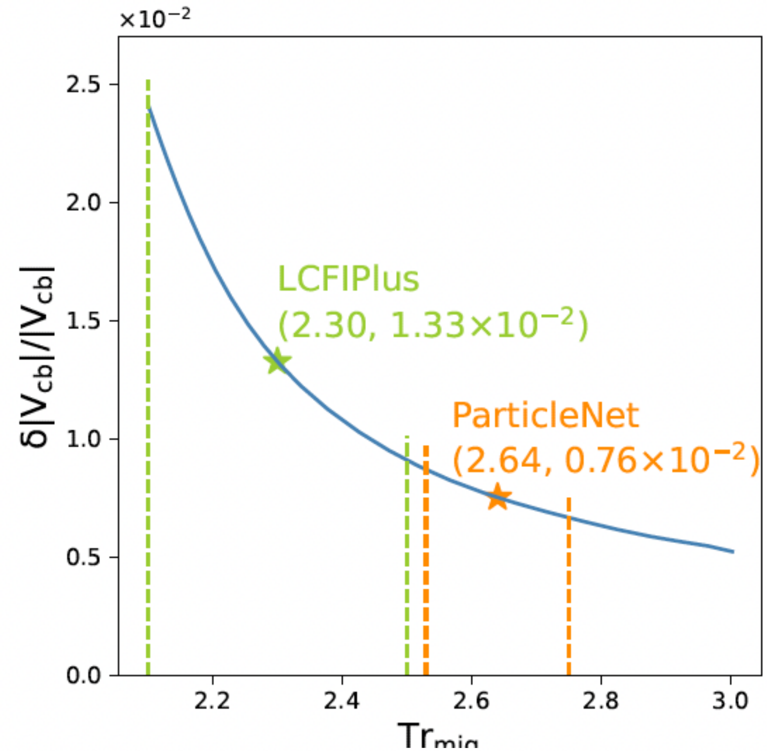


quark \ tag	b_1	b_2	c_1	c_2	q_1	q_2
b	0.47	0.378	0.0197	0.0965	0.00397	0.0315
c	0.00042	0.078	0.298	0.373	0.0682	0.182
uds	0.000104	0.00477	0.00145	0.054	0.538	0.401

- $\mu\nu qq$
  - 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%

Impact on physics benchmarks





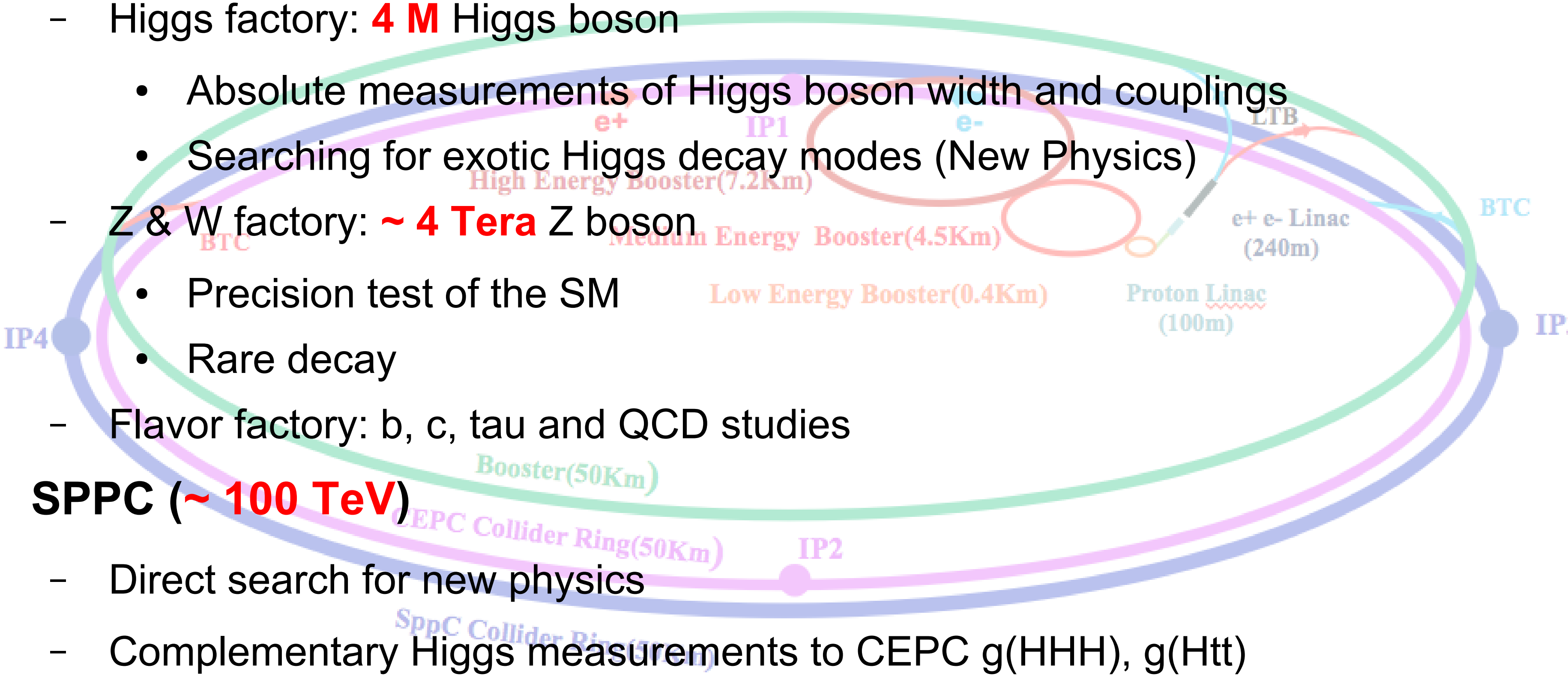
Conservative/Aggressive:

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

Key parameters of the CEPC-SPPC

- **Tunnel ~ 100 km**
- **CEPC (90 – 240 GeV)**
  - Higgs factory: **4 M** Higgs boson
    - Absolute measurements of Higgs boson width and couplings
    - Searching for exotic Higgs decay modes (New Physics)
  - Z & W factory: **~ 4 Tera** Z boson
    - Precision test of the SM
    - Rare decay
  - Flavor factory: b, c, tau and QCD studies
- **SPPC (~ 100 TeV)**
  - Direct search for new physics
  - Complementary Higgs measurements to CEPC g(HHH), g(Htt)
  - ...
- **Heavy ion, e-p collision...**



Complementary