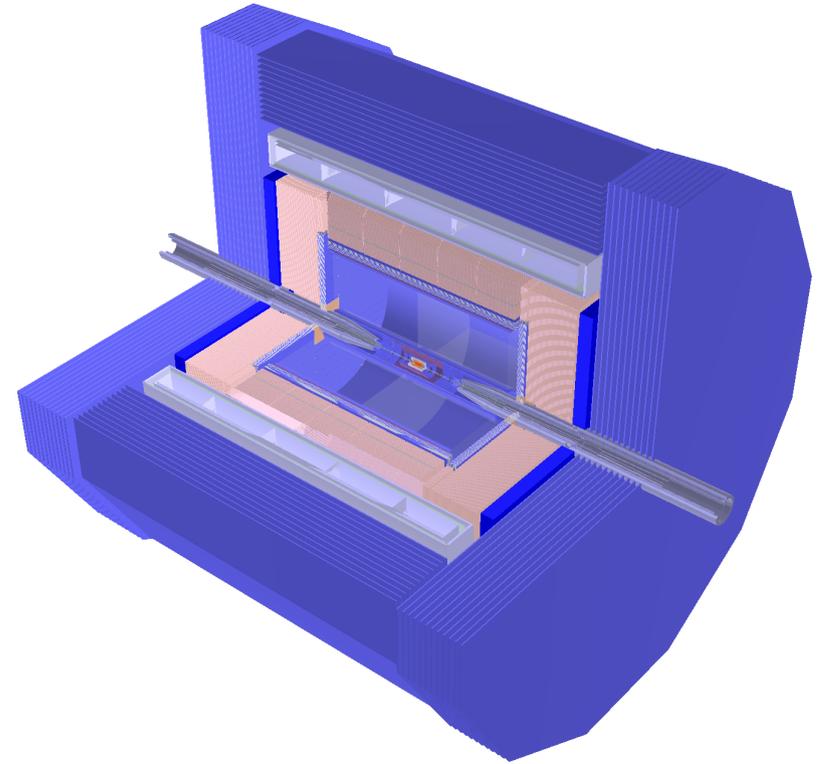
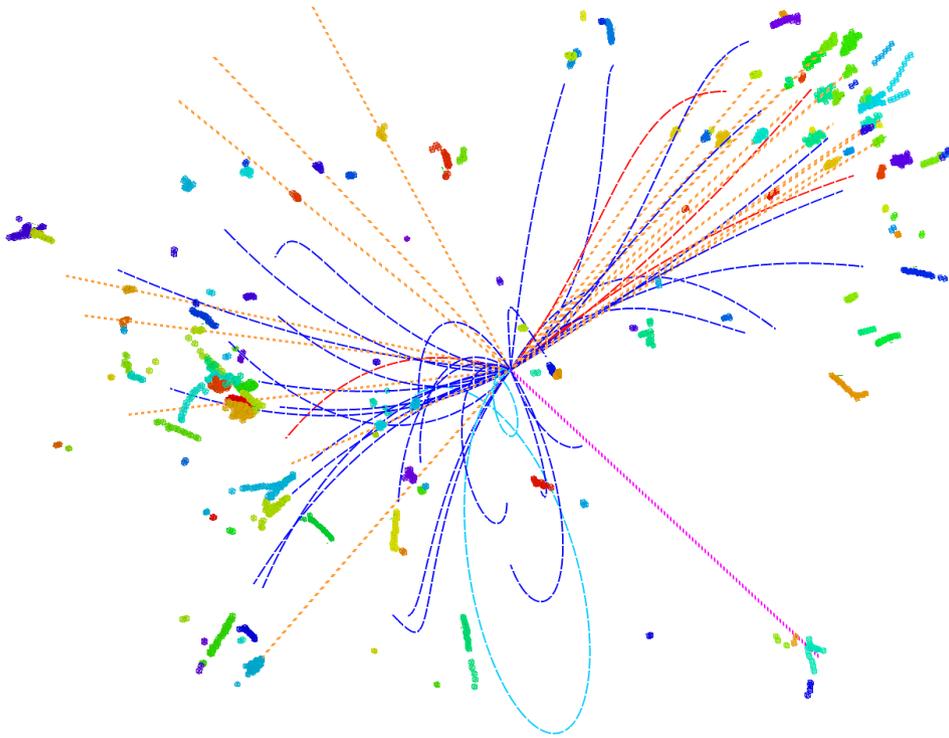




***From Jet Flavor Tagging to Origin  
Identification: AI usage on  
hadronic event of electron positron  
Higgs factories***

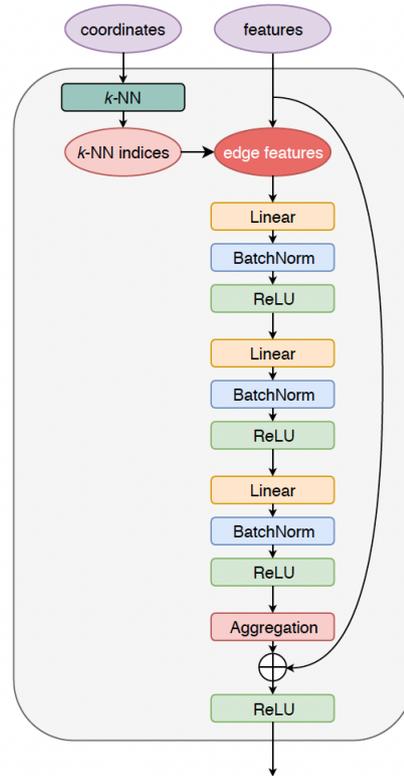
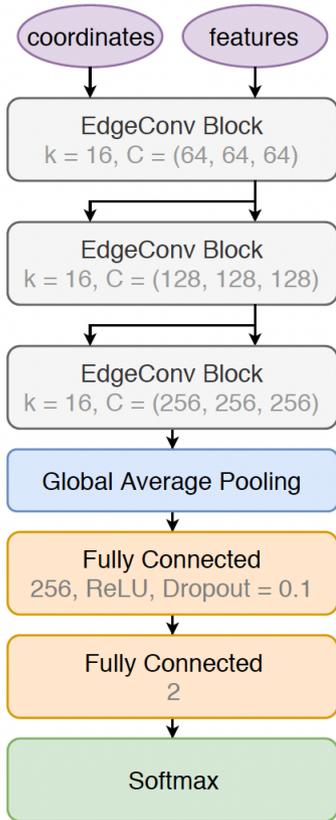
Yongfeng Zhu, Hao Liang, Chen Zhou, Huilin Qu,  
Manqi Ruan, etc

# Geo. & Tools



- Jet Flavor Tagging: 3 categories (b, c, light)
- Jet origin identification: 11 categories (5 quarks + 5 anti quarks + gluon)
  - Jet Flavor Tagging + Jet Charge measurements + ...
- Full Simulated Z/H to dijet sample at CEPC-v4 detector, reconstructed with Arbor.

# Particle Net: IO

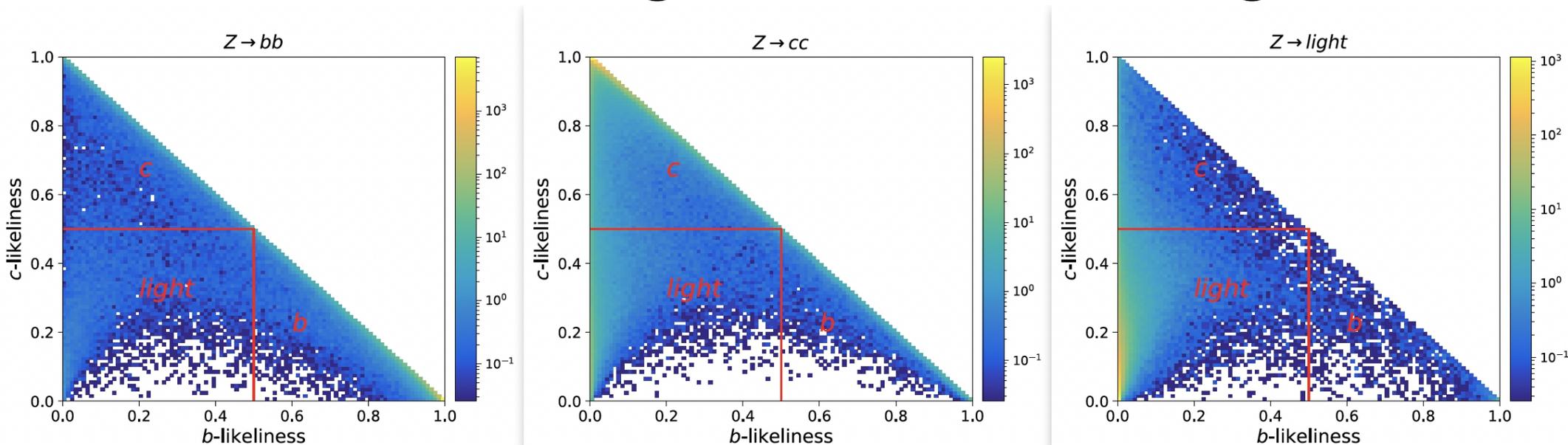


Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$\log p_T$	logarithm of the particle's $p_T$
$\log E$	logarithm of the particle's energy
$\log \frac{p_T}{p_T(jet)}$	logarithm 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
$\Delta R$	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
isChargedKaon	if the particle is a charged Kaon
isChargedPion	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.

- Output: likelihoods to different categories

# Three categories: b, c, & light



Hadronic Z pole sample

1 M  $Z \rightarrow bb, cc, (uds)$  each

60/20/20% for

training/validating/testing.

Result on Testing sample

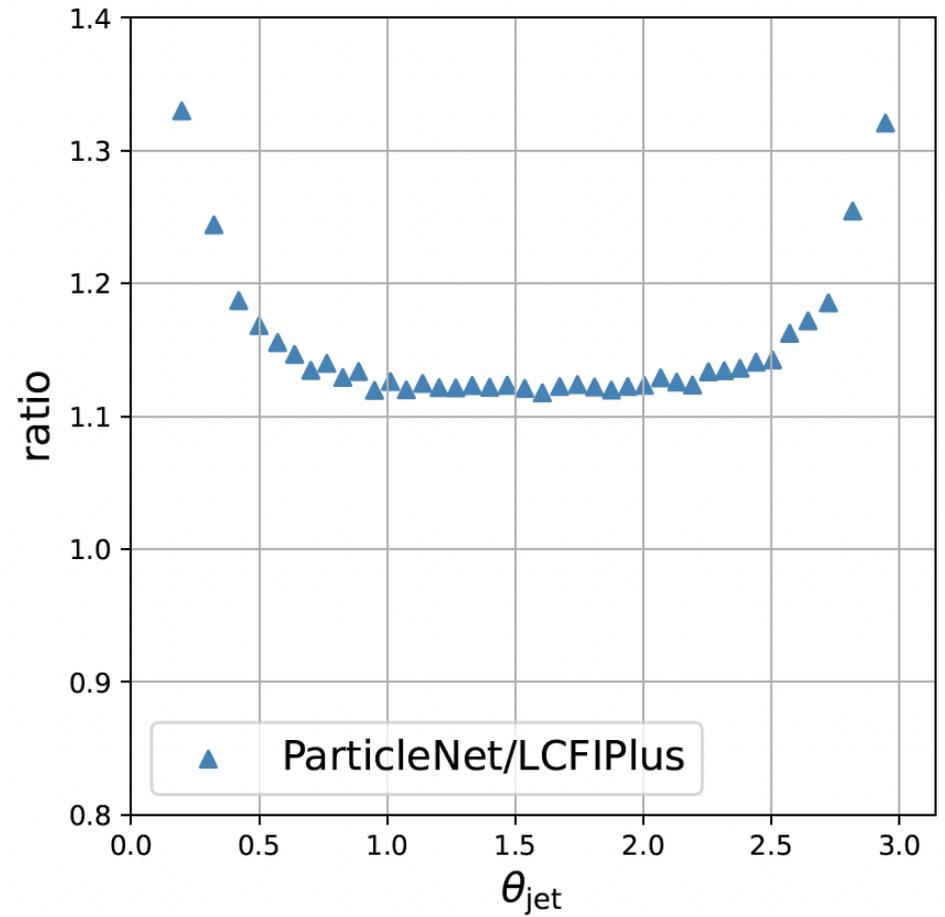
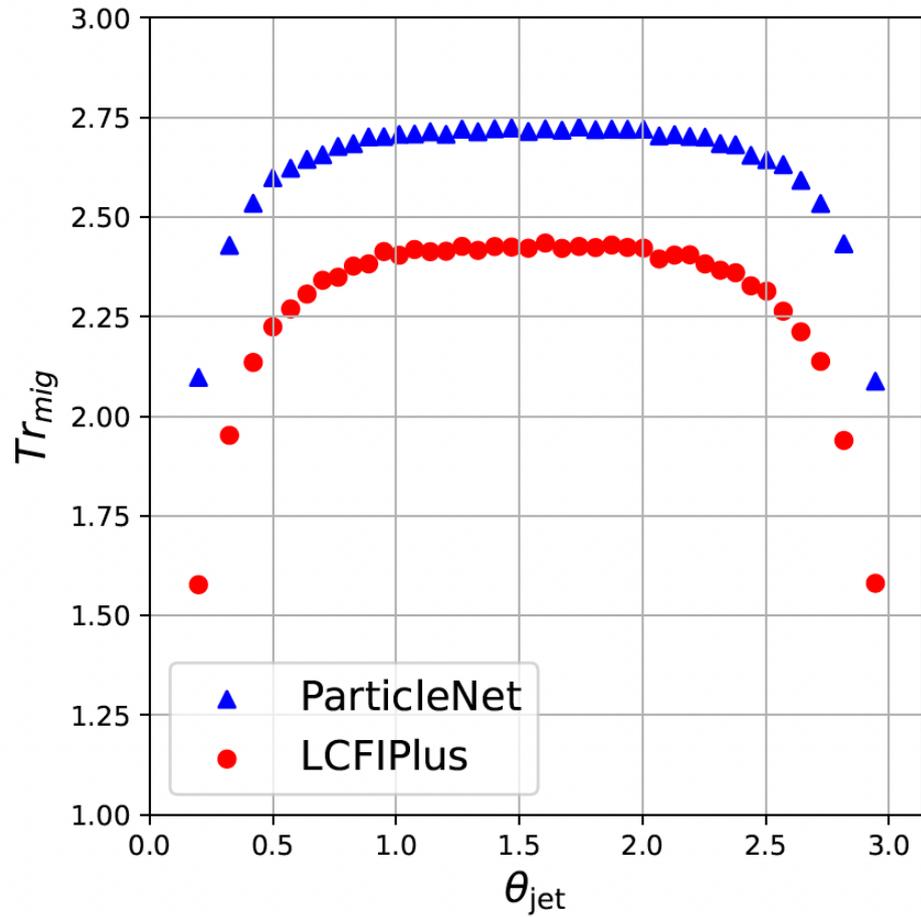
		predicted		
		b	c	uds
truth	b	0.911	0.059	0.031
	c	0.039	0.784	0.177
	uds	0.005	0.051	0.944

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

5/9/2023

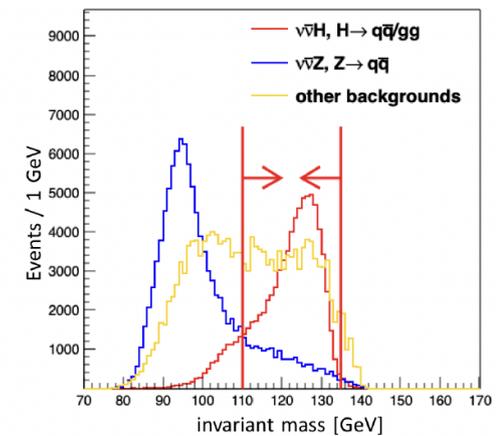
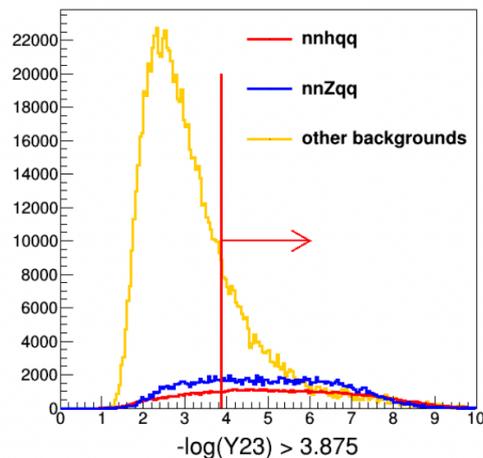
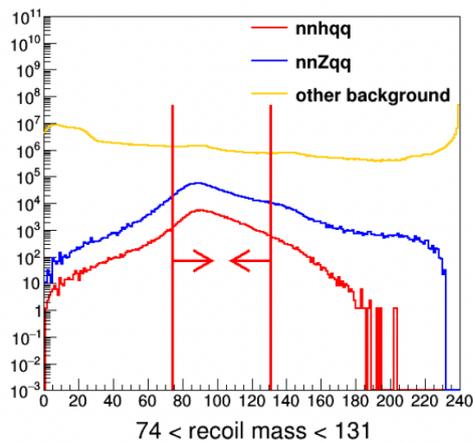
Figure 7. The migration matrix of ParticleNet (left) and LCFIPlus (right) at the CEPC.

# Dependence on polar angle

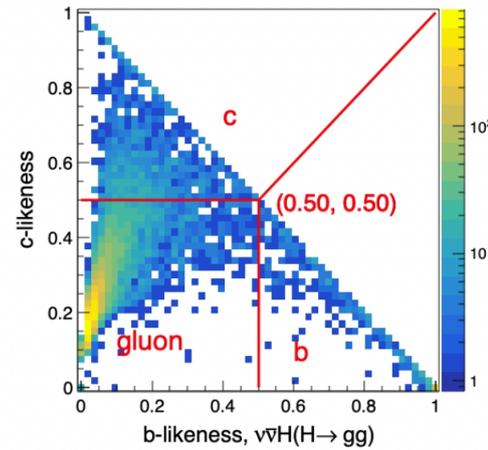
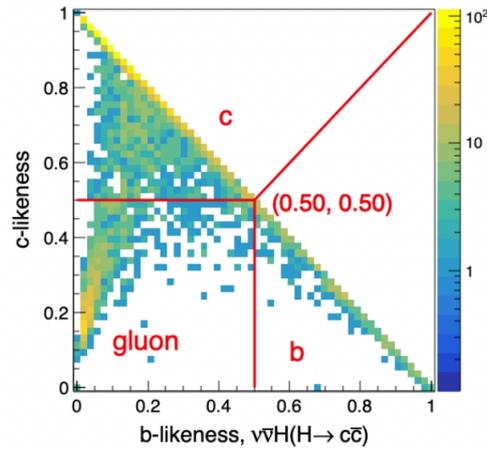
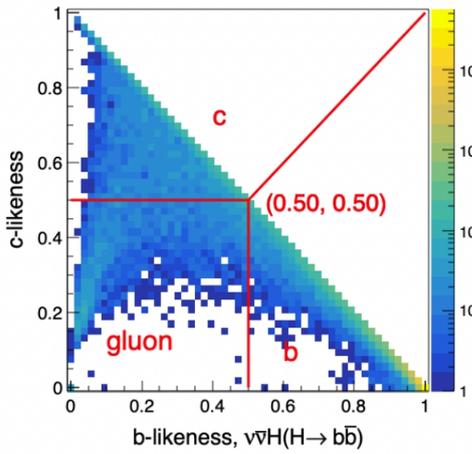


# Impact on benchmark: $\nu\nu H$ , $H \rightarrow \text{jets}$

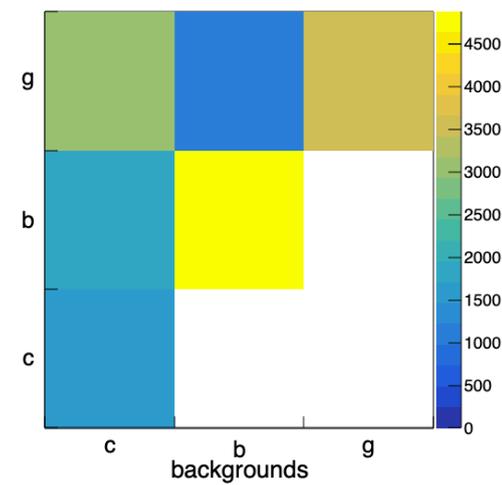
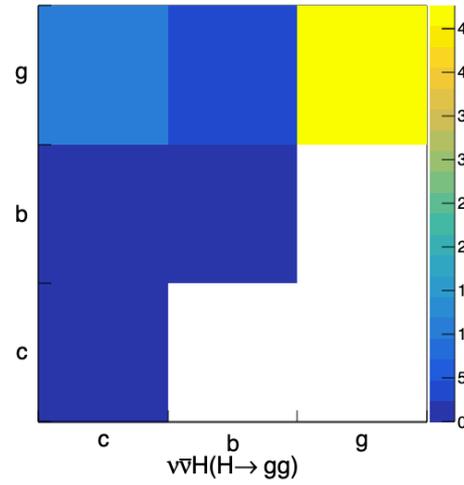
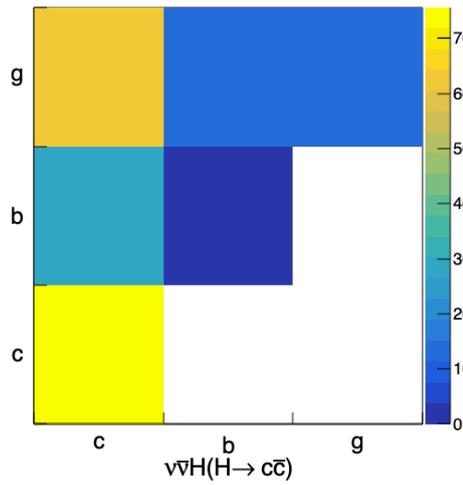
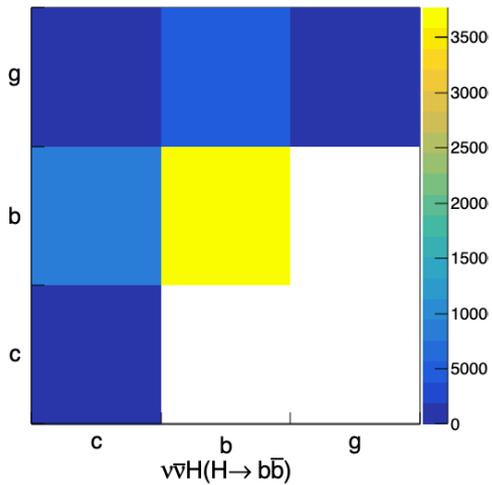
	$\nu\nu H q\bar{q}/gg$	2f	SW	SZ	WW	ZZ	Mixed	ZH	$\frac{\sqrt{S+B}}{S}$ (%)
total	178890	8.01E8	1.95E7	9.07E6	5.08E7	6.39E6	2.18E7	961606	16.86
recoilMass (GeV) $\in (74, 131)$	157822	5.11E7	2.17E6	1.38E6	4.78E6	1.30E6	1.08E6	74991	4.99
visEn (GeV) $\in (109, 143)$	142918	2.37E7	1.35E6	8.81E5	3.60E6	1.03E6	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
PI (GeV) $\in (0, 50)$	118109	52784	1.05E5	74936	7.30E5	1.13E5	847	34279	0.94
$-\log_{10}(Y23)$ $\in (3.375, +\infty)$	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



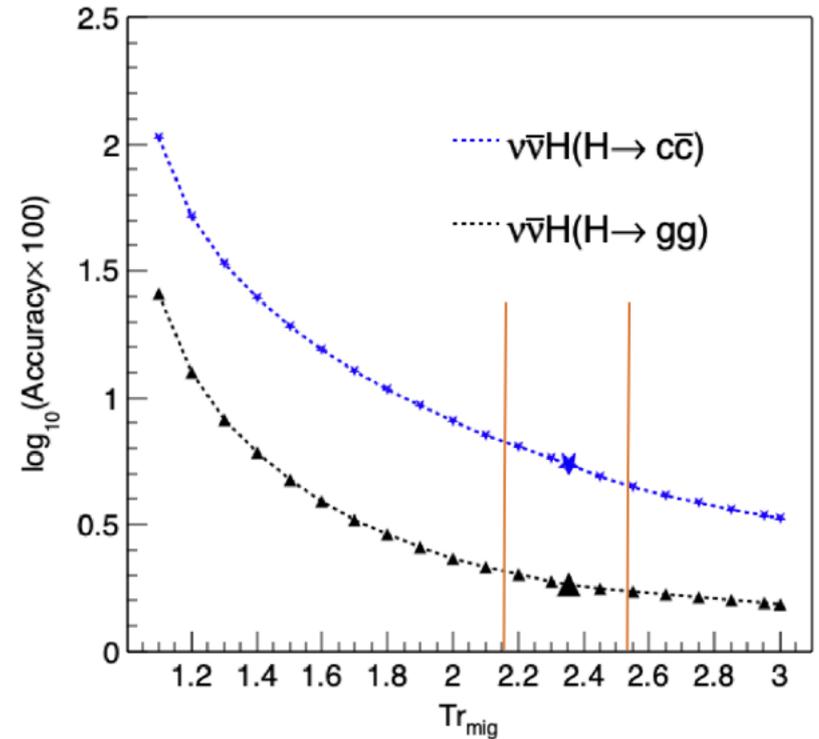
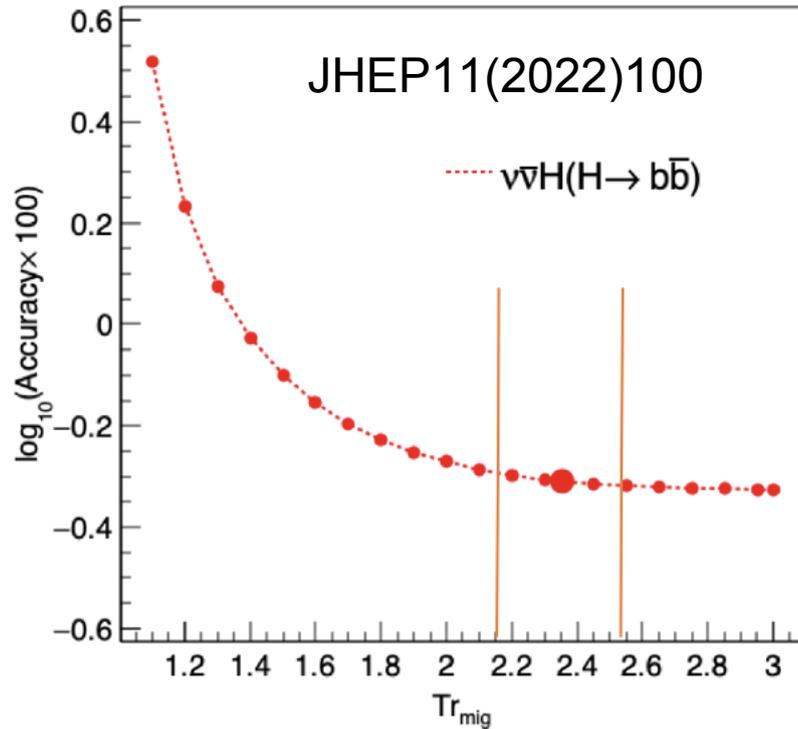
# Flavor tagging @ $\nu\nu H$



true \ identified as	b	c	g
b	0.8675	0.0887	0.0437
c	0.1136	0.6263	0.2601
g	0.0411	0.1007	0.8582



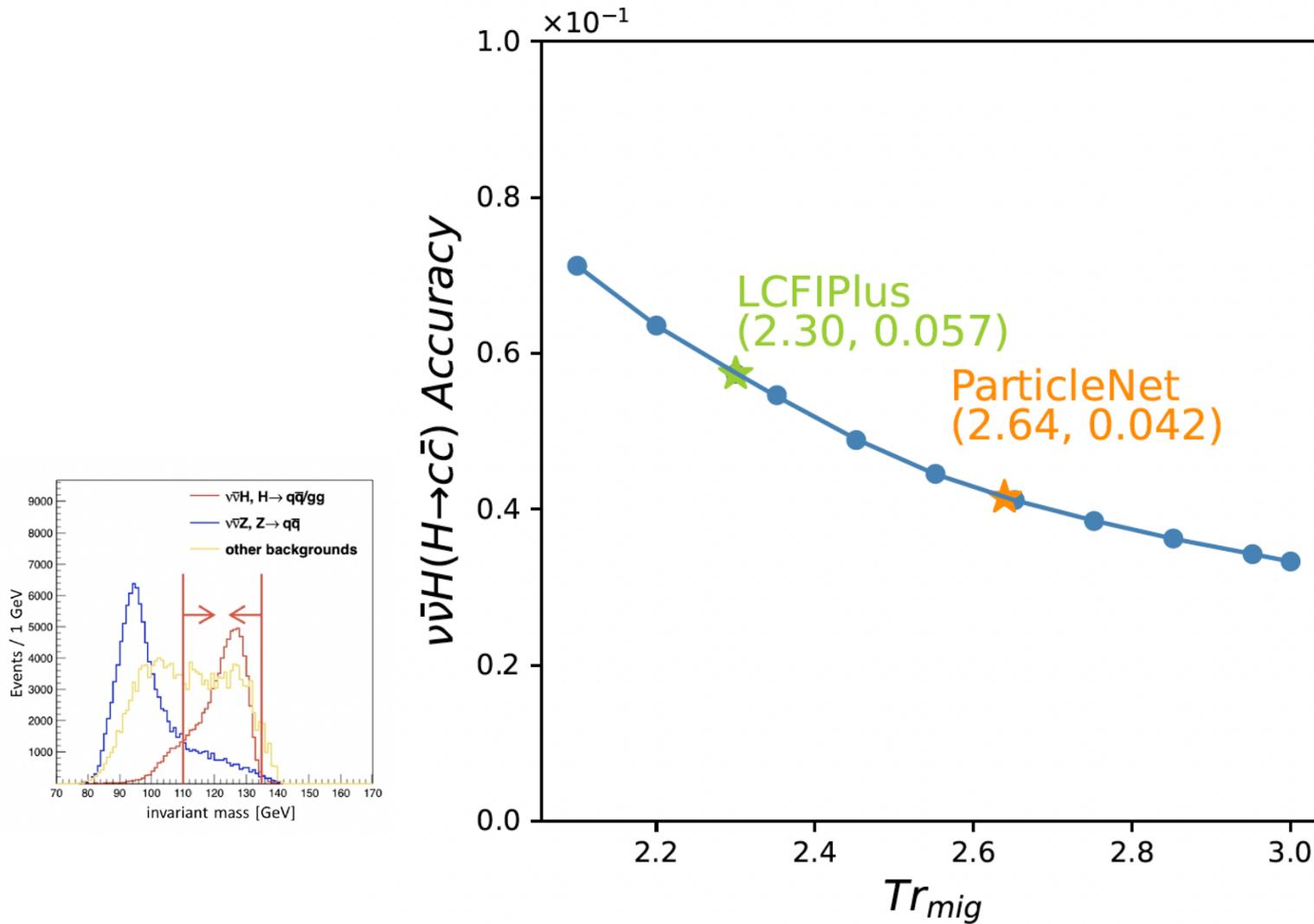
# Vertex



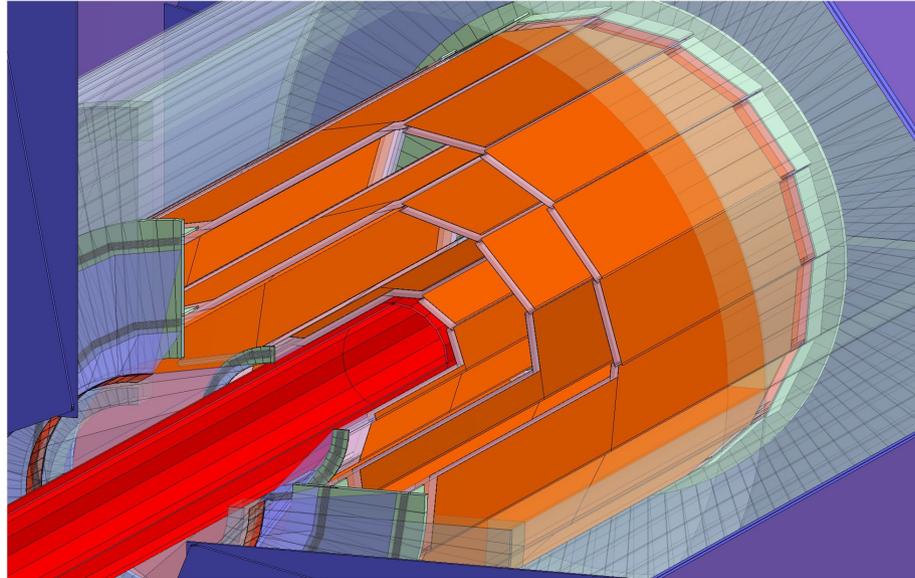
$$Tr_{\text{mig}} = 2.35 + 0.05 \cdot \log_2 \frac{R_{\text{material}}^0}{R_{\text{material}}} + 0.04 \cdot \log_2 \frac{R_{\text{resolution}}^0}{R_{\text{resolution}}} + 0.10 \cdot \log_2 \frac{R_{\text{radius}}^0}{R_{\text{radius}}}.$$

- Vertex: track impact para & 2<sup>nd</sup> vertex reconstruction: Flavor Tagging, etc
  - As close to the IP as possible
  - Limited by the beam induced background (~ beam energy & B-Field)

# Impact on Physics benchmark

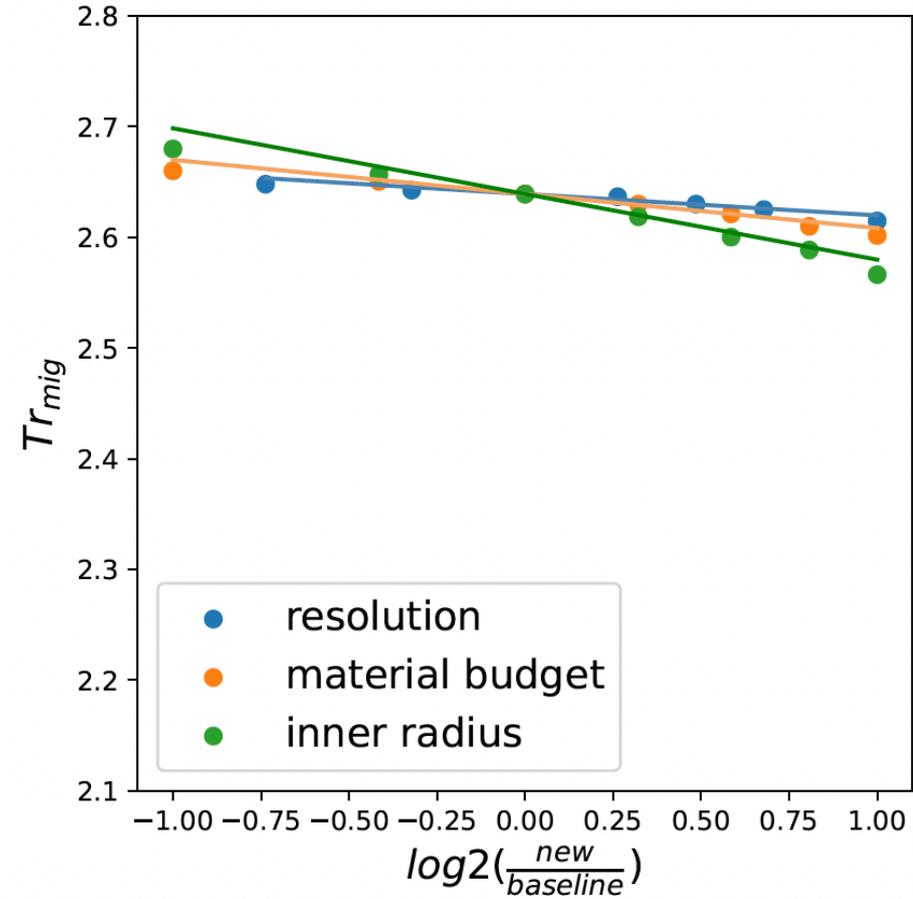
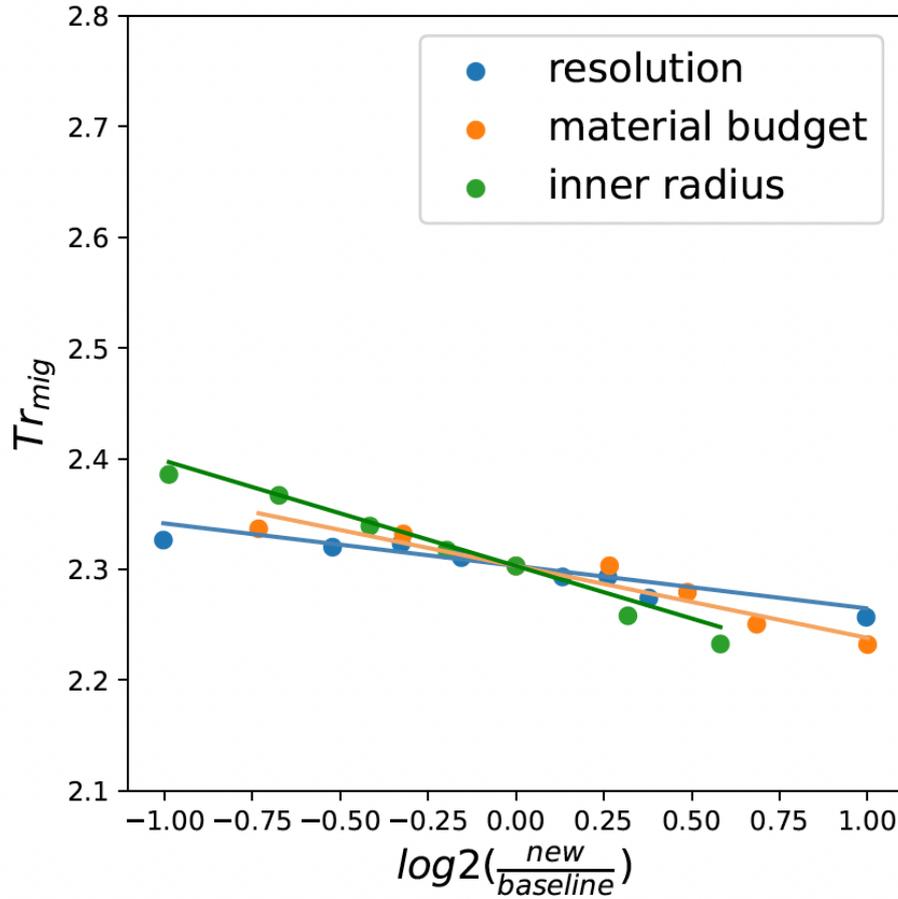


# Comparison on Det. Optimization



	R (mm)	single-point resolution ( $\mu m$ )	material budget
Layer 1	16	2.8	0.15%/X <sub>0</sub>
Layer 2	18	6	0.15%/X <sub>0</sub>
Layer 3	37	4	0.15%/X <sub>0</sub>
Layer 4	39	4	0.15%/X <sub>0</sub>
Layer 5	58	4	0.15%/X <sub>0</sub>
Layer 6	60	4	0.15%/X <sub>0</sub>

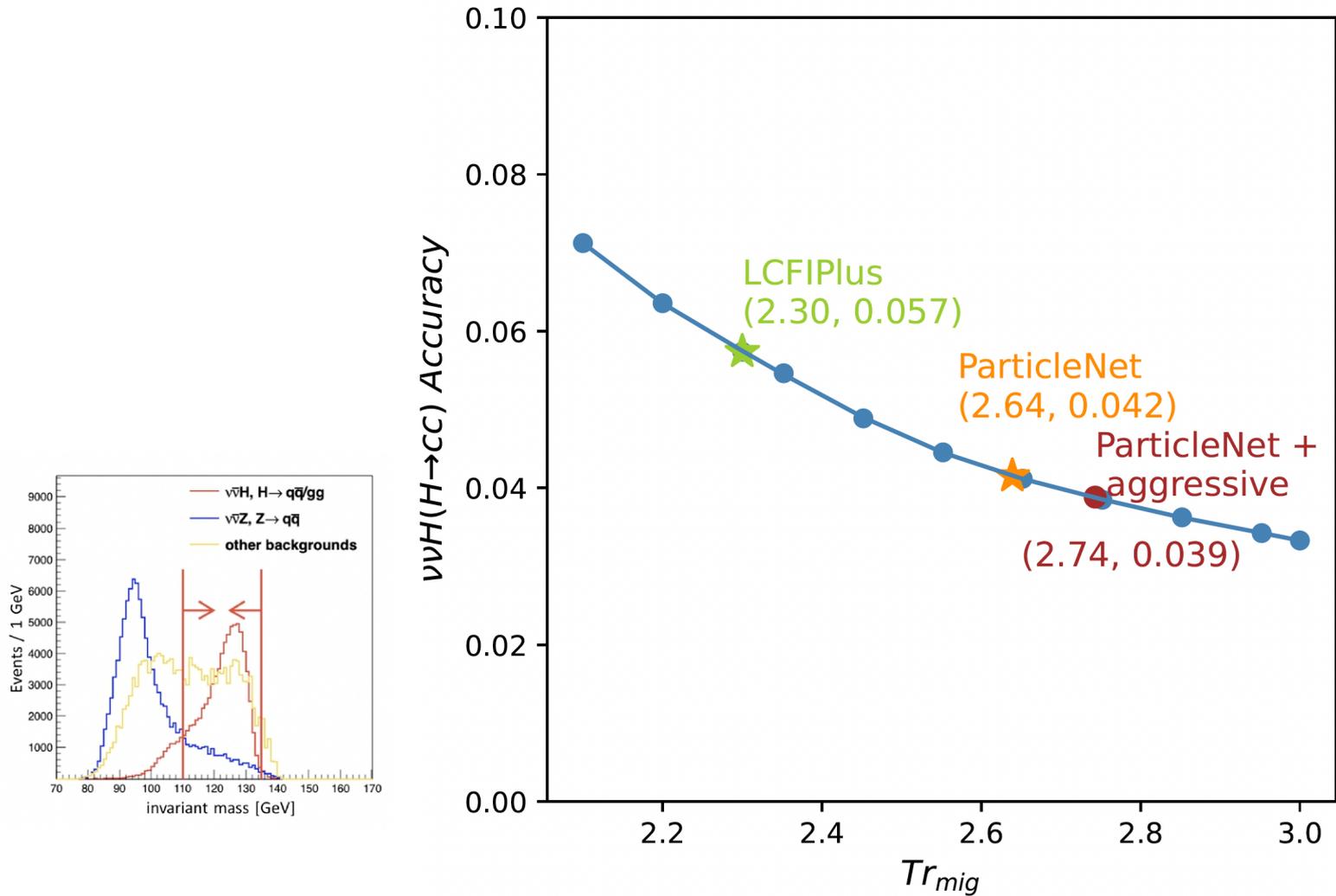
# Comparison on Det. Optimization



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

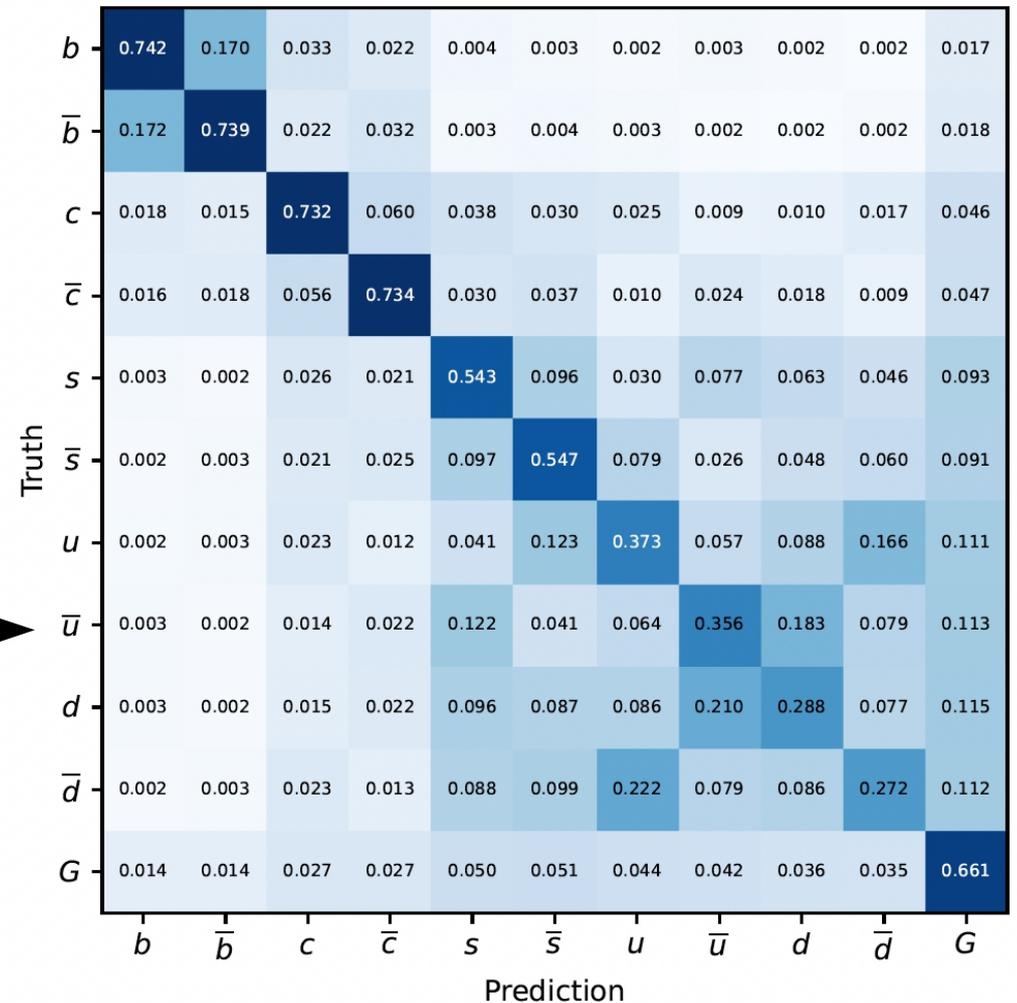
$$Tr_{mig} = 2.64 + 0.03 \cdot \log_2 \frac{R_{material}^0}{R_{material}} + 0.02 \cdot \log_2 \frac{R_{resolution}^0}{R_{resolution}} + 0.06 \cdot \log_2 \frac{R_{radius}^0}{R_{radius}} \quad (4.2)$$

# On the same benchmark...

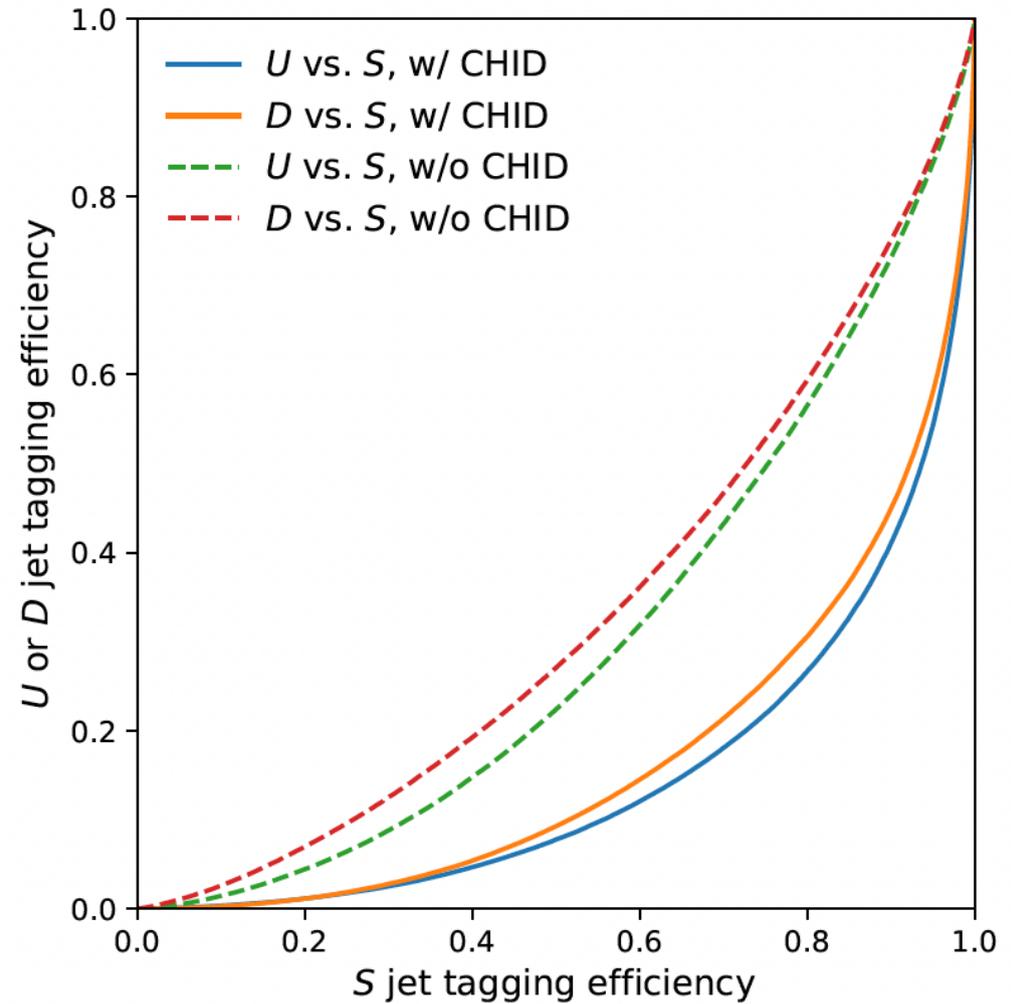
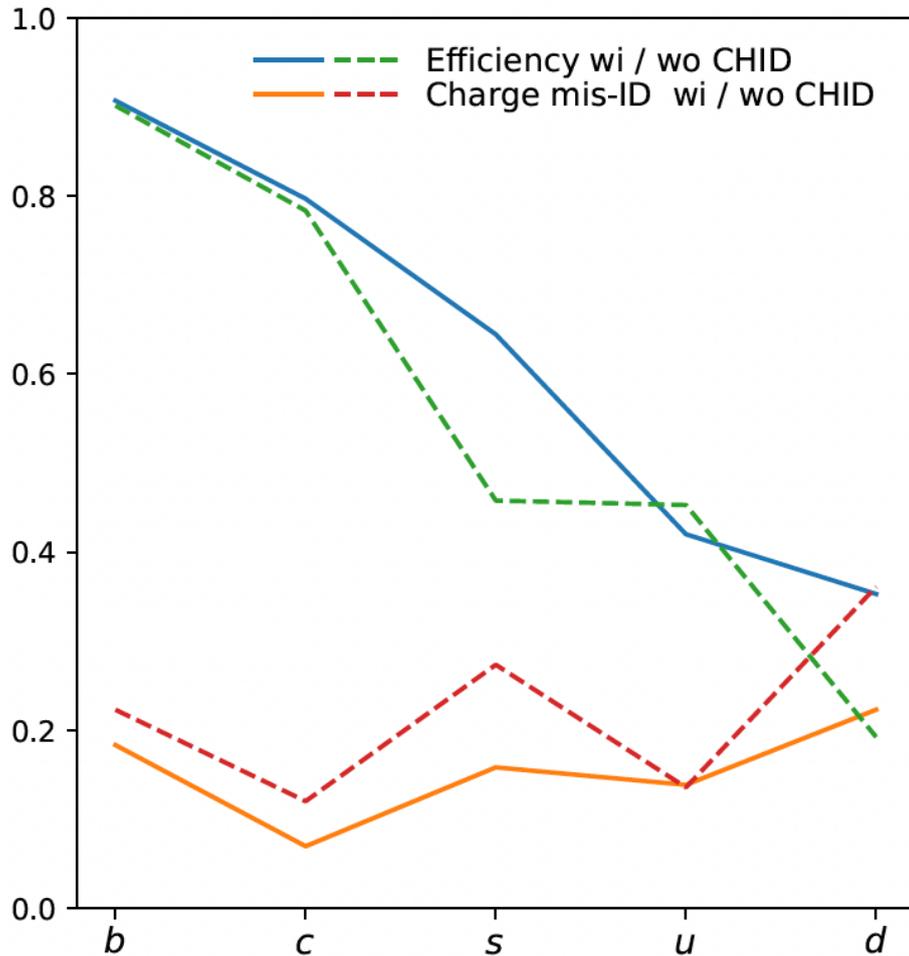


# Jet origin id: 11 categories

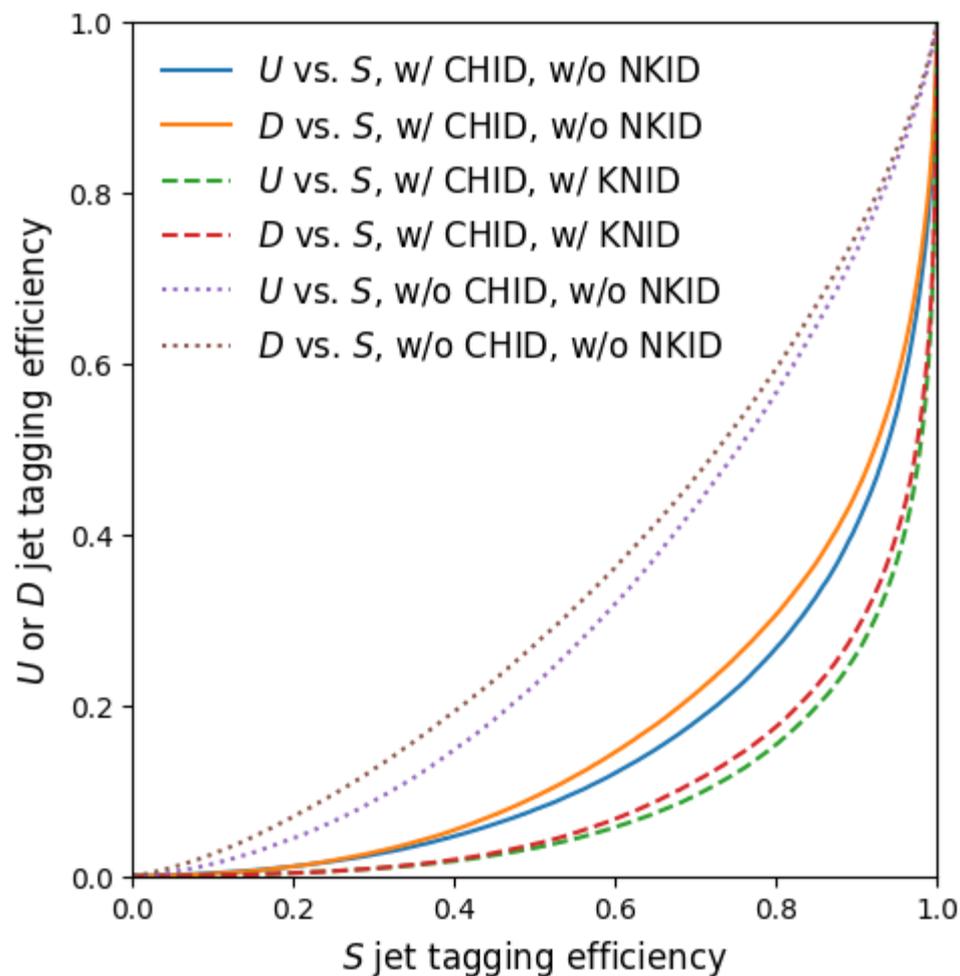
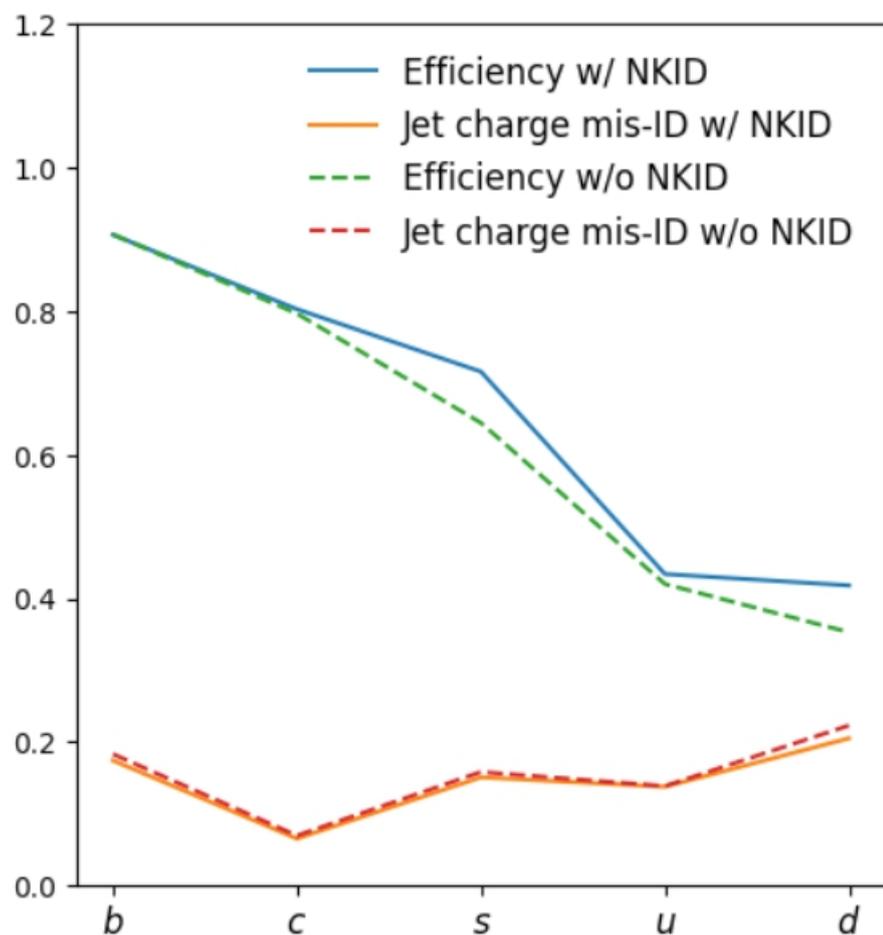
- vvH sample, with Higgs decays into different species of colored particle: 5 quark, 5 antiquark & gluon
  - 1 Million of each type
  - 60/20/20% for training, validating, and testing, result corresponding to testing sample
- Pid: ideal Pid – three categories
  - Lepton identification
  - Charged Kaon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



# Impact of charged kaon id

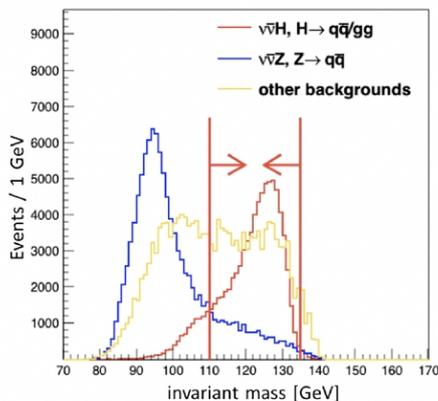
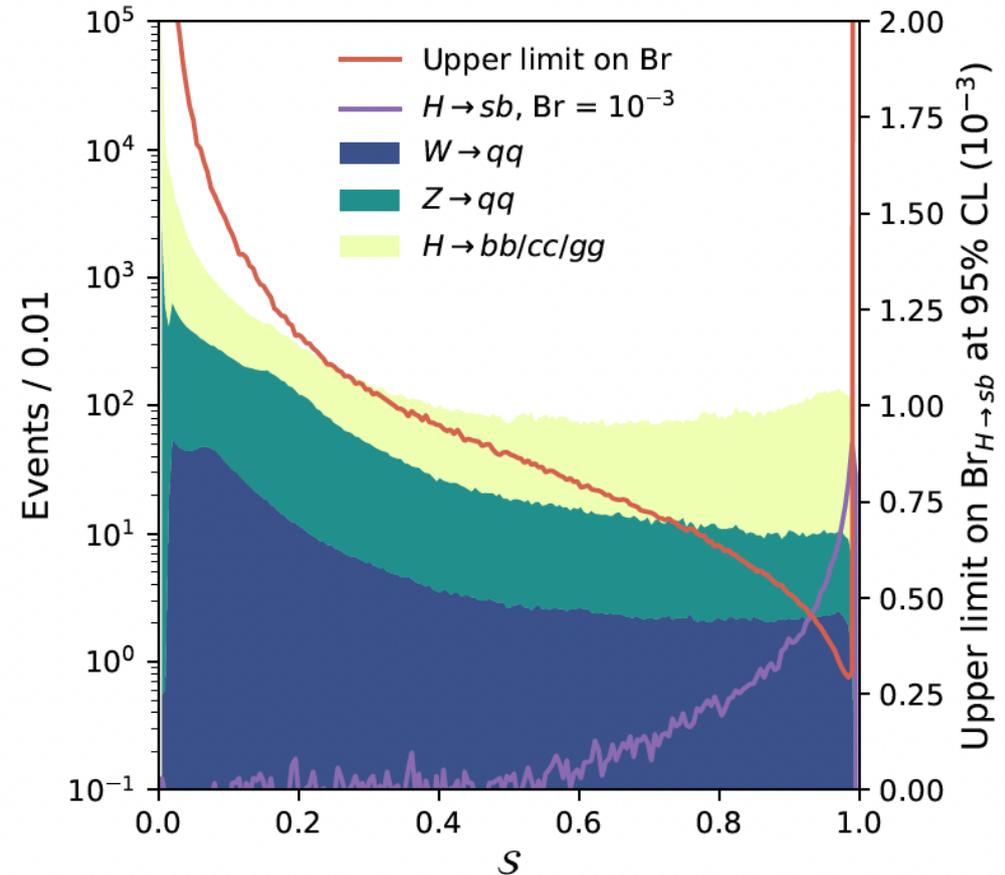
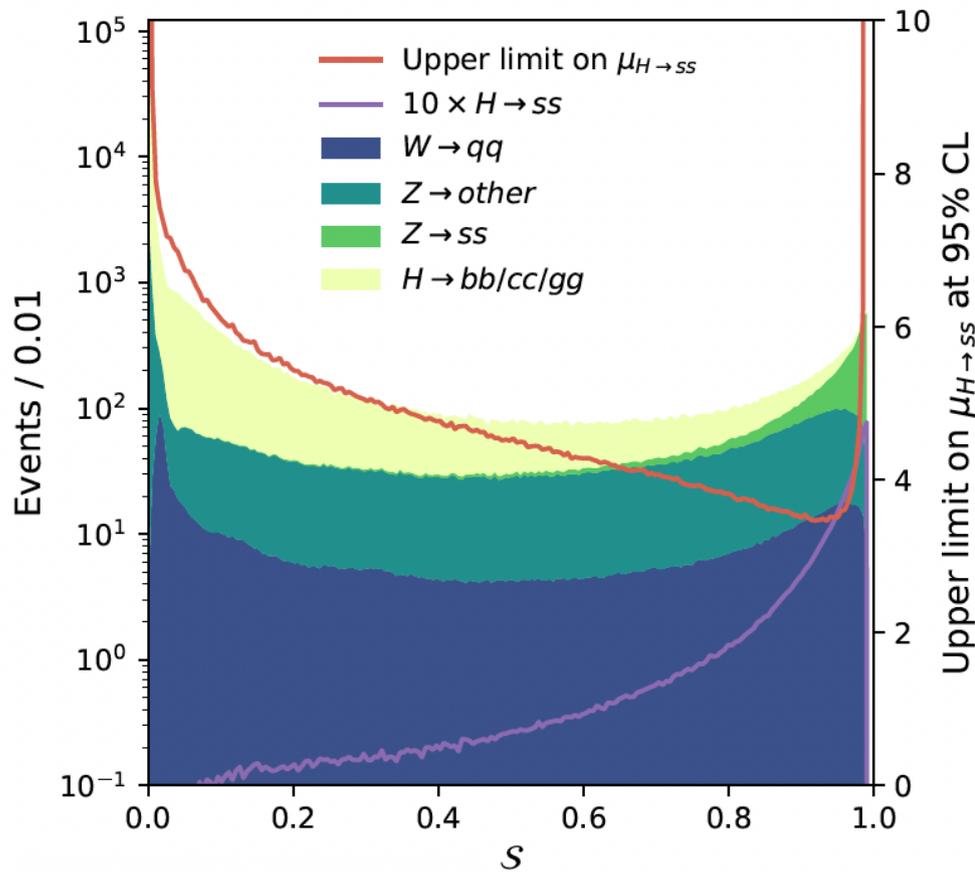


# Neutral Kaon id



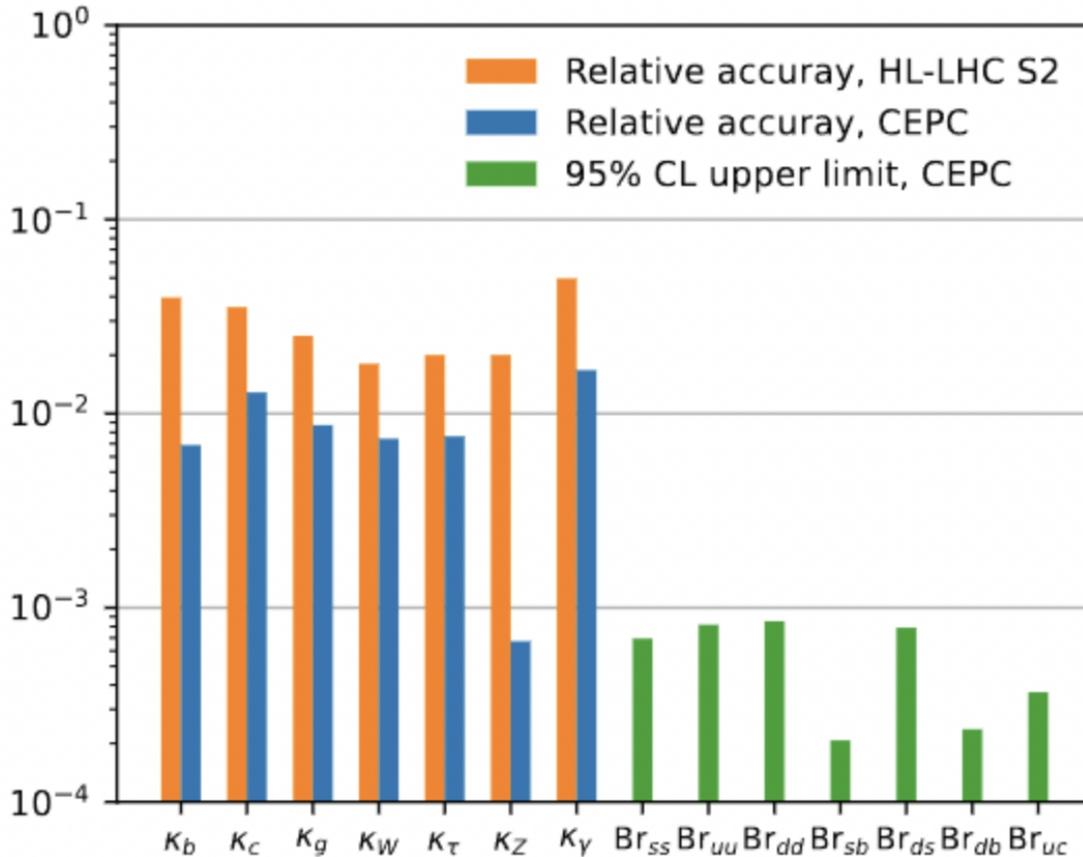
- Current tool (PN) is not clever enough to figure out  $K_s \rightarrow 2 \pi$ , etc

# Benchmark analyses using Jet origin ID



Applied to quasi-data of  $\nu\bar{\nu}H$ ;  
 $H \rightarrow ss$ : be limited to  $3 \times SM$  using  $\nu\bar{\nu}H$  +  $llH$  at 20 iab  
 $H \rightarrow sb$ : up limit of  $2E-4$  at 95% C.L.

# Benchmark analyses using Jet origin ID



	Bkg. ( $10^3$ )			Upper limit ( $10^{-3}$ )						
	$H$	$Z$	$W$	$ss$	$uu$	$dd$	$sb$	$ds$	$db$	$uc$
$\nu\nu H$	152	16	1.7	0.74	0.86	0.92	0.25	0.84	0.27	0.44
$\mu\mu H$	50	25	0	2.5	3.1	3.1	0.5	2.9	0.6	1.0
$ee H$	26	16	0	3.9	4.6	4.8	0.7	4.4	0.8	1.6
Comb.	-	-	-	0.68	0.81	0.85	0.20	0.81	0.23	0.38

TABLE II: Summary of background events of  $H \rightarrow bb/cc/gg$ ,  $Z$ , 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.

# Summary

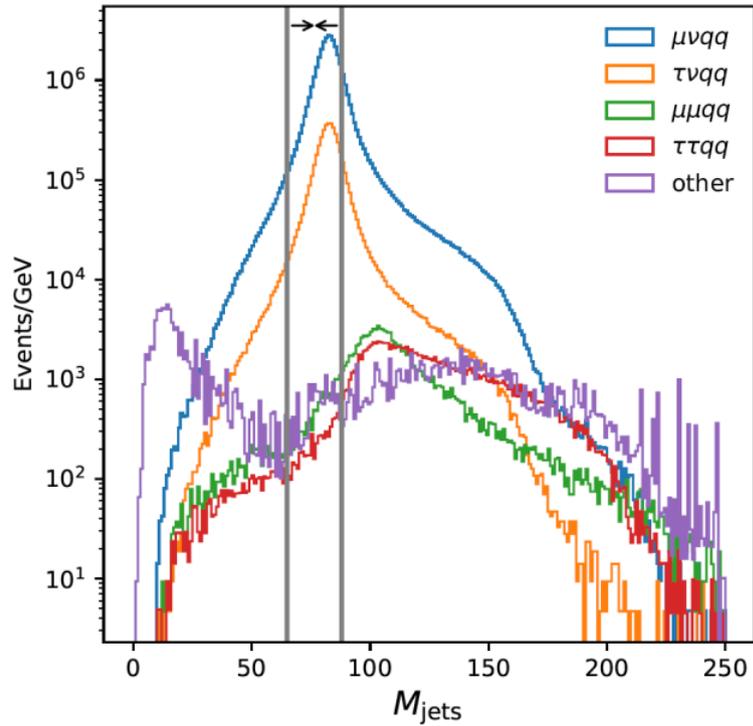
- Reco & Measurements of hadronic events is critical to e+e- Higgs factories. At current baseline detector & ParticleNet, jet origin identification (combines jet flavor tagging + charge measurements) is possible
  - M11 established, with encouraging performances
  - Compared to LCFIPlus, ParticleNet has significant better performance
    - Improves H->cc accuracy ~ 40%
    - Leads to consistent conclusion on Det. Optimization
  - Higgs exotic/FCNC processes with hadronic final states can be limited to the Branching ratio of 1E-3 to 1E-4; H->ss limited to 3 times SM prediction (vvH + llH only)
  - *Yet, it cannot figure out Ks decays into 2 pion...*
- Vision (long long term)
  - Jet origin id as Pid
  - Access to g(Hss) at future Higgs factory

# Backup

# Summary

- Endeavor of 11 years: CEPC is technologically ready for construction
- CEPC supports extremely rich physics program, leads to stringent requirement to its detector system, Multiple detector concepts are proposed
- Key physics requirements for the Calorimeter system
  - EM resolution  $\sim 3\%/\sqrt{E}$
  - BMR  $< 3\%$
  - ToF  $< 50$  ps (at Cluster level)
- As for our collaboration
  - High quality glasses with cheap enough price
  - High homogeneity ( $< 10\%$ )
  - High density (5-6)
  - High light yield: to support Hadronic energy resolution of  $40\%/\sqrt{E}$

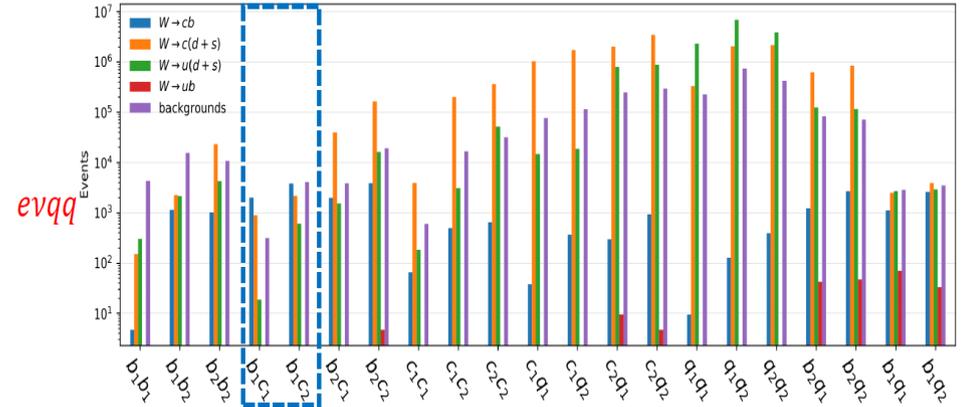
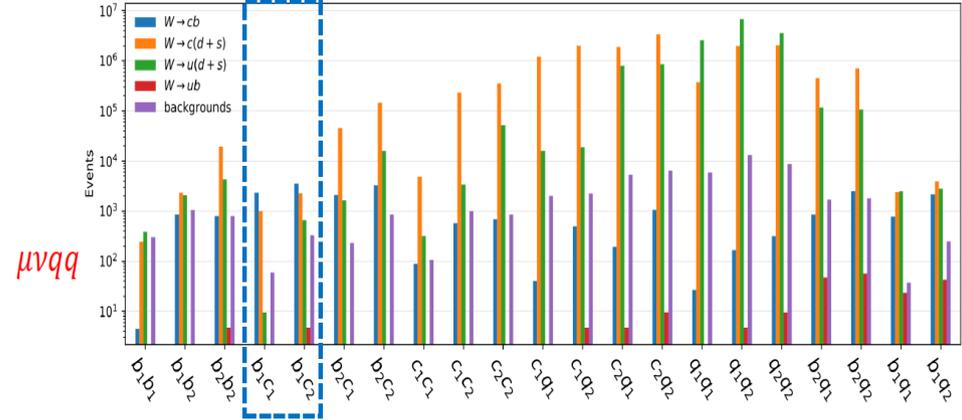
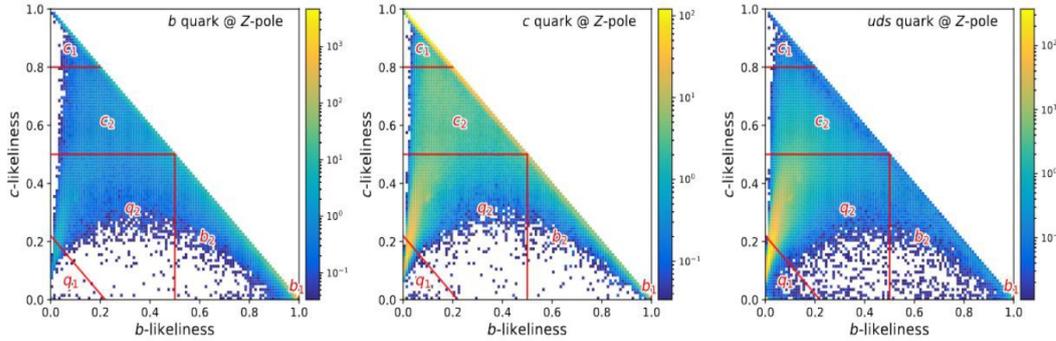
# Vcb from W decay



	$\mu\nu W, W \rightarrow$				$\tau(\mu\nu)\nu_\tau W, W \rightarrow$				$\tau\nu_\tau qq, \tau \rightarrow$					
	cb	ub	c(d/s)	u(d/s)	cb	ub	c(d/s)	u(d/s)	e2ν	had.ντ	ττqq	μμqq	Higgs	others
w/o selections	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
$E_{L\mu} > 12\text{GeV}$	37.9K	330	22.6M	22.6M	5.59K	56	2.98M	2.97M	133K	687K	422K	2.82M	645K	186.3M
$R_{L\mu} > 0.85$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$\cos(\theta_{L\mu})$	35.3K	302	21.1M	21.1M	5.01K	46	2.73M	2.73M	1.55K	43.2K	266K	1.82M	308K	128.8M
$q_{L\mu} \cos(\theta_{L\mu}) < 0.20$	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 $\ell$ 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 $\geq 15$	32.8K	283	19.5M	19.4M	4.7K	42	2.56M	2.55M	1.23K	39.6K	153K	522K	118K	185K
Missing $P_T > 9.5 \text{ GeV}/c$	31.5K	264	18.7M	18.6M	4.38K	37	2.4M	2.39M	1.18K	37.2K	136K	118K	92.6K	97.7K
$M_{\text{jets}} > 65 \text{ GeV}/c^2$	29.4K	254	18.1M	18.3M	4.15K	32	2.33M	2.35M	978	36.0K	132K	112K	85.3K	24.5K
$M_{\text{jets}} < 88 \text{ GeV}/c^2$	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
$M_{\text{jets, recoil}} < 115 \text{ GeV}/c^2$	20.2K	184	13.0M	13.1M	2.96K	23	1.72M	1.73M	505	22.6K	3.57K	6.86K	536	3.02K
$M_{L\mu S\mu} < 75 \text{ GeV}/c^2$	19.6K	184	12.9M	13.0M	2.95K	23	1.72M	1.73M	505	22.6K	3.56K	5.78K	414	3.0K
$M_{\ell\nu} > 12 \text{ GeV}/c^2$	19.6K	184	12.9M	13.0M	2.7K	18	1.54M	1.55M	416	19.5K	2.08K	5.16K	390	1.81K
$\epsilon_{\text{kin}} (\%)$	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)
$b_1 c_{1,2}$	5.14K	4	2.79K	571	632	0	407	65	0	14	67	228	0	0
$\epsilon_{b_1 c_{1,2}} (\%)$	12.8	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:
  - $W \rightarrow c(d/s)$
  - $\mu\mu qq$

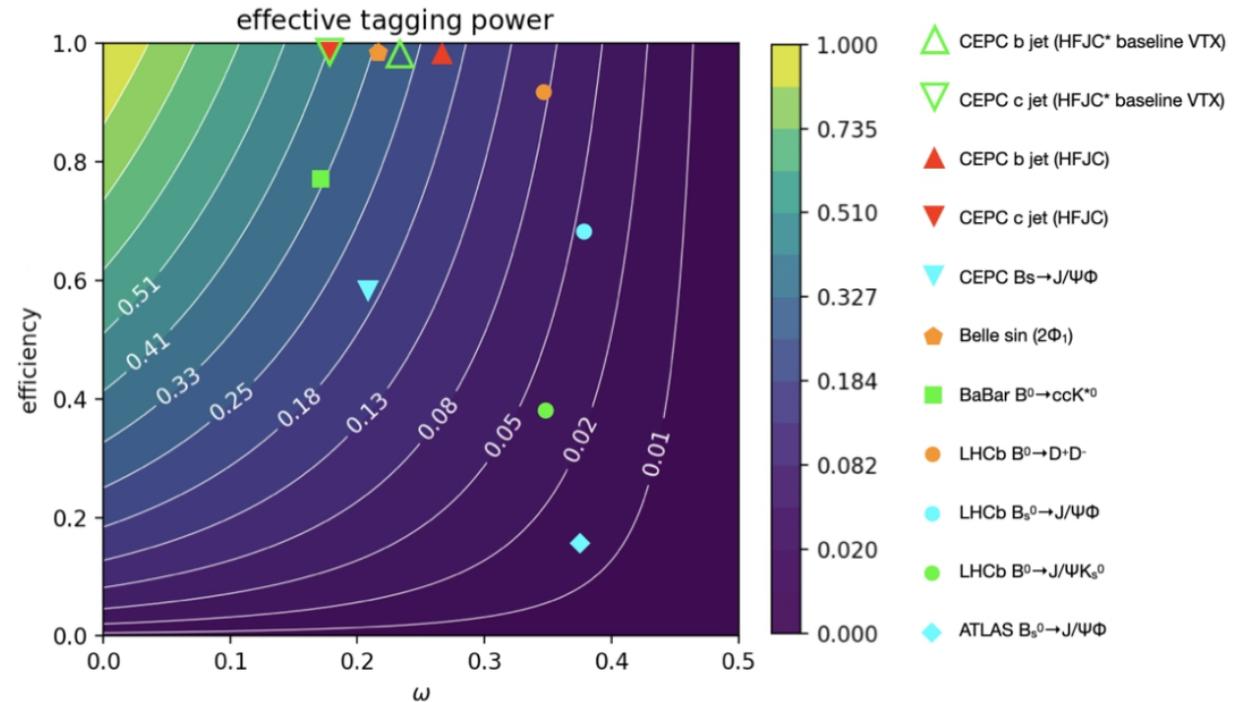
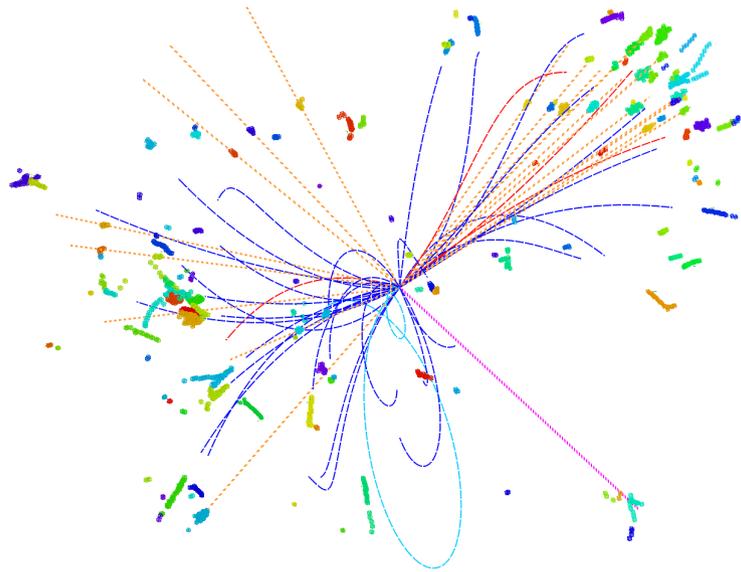
# Vcb from W decay



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%

# Jet charge measurement using Leading jet particle & weighted jet charge



Eff. Tag. Power:  
~ 40%/20% for c/b jet  
with 45 GeV energy

arXiv > hep-ex > arXiv:2306.14089

High Energy Physics – Experiment

[Submitted on 25 Jun 2023 (v1), last revised 13 Jul 2023 (this version, v3)]

## Jet charge identification in ee-Z-qq process at Z pole operation

Hanhua Cui, Mingrui Zhao, Yuexin Wang, Hao Liang, Manqi Ruan

Accurate jet charge identification is essential for precise electroweak and flavor measurements at the high-energy frontier. We propose a novel method called the Leading Particle Jet Charge method (LPJC) to determine the jet charge based on information about the leading charged particle. Tested on Z – bb and Z – cc samples at a center-of-mass energy of 91.2GeV, the LPJC achieves an effective tagging power of 20%/9% for the c/b jet, respectively. Combined with the Weighted Jet Charge method (WJC), we develop a Heavy Flavor Jet Charge method (HFJC), which achieves an effective tagging power of 39%/20% for c/b jet, respectively. This paper also discusses the dependencies between jet charge identification performance and the fragmentation process of heavy flavor jets, and critical detector performances.

# Summary

- We propose CHLOE, using
  - GSHCAL
  - Xbar ECAL + Position/timing layer of
    - **Silicon**
    - MGPRC
  - 2.5 Tracker Scenarios:
    - **Gas Tracker  $R_{in/out} \sim 25/175$  cm,  $Z \sim 500$  cm**
    - Improved 4<sup>th</sup>: Fwd RHIC
    - Full Silicon with Pid ( $dE/dx \sim 3\%$ ...)
  - 3 VTX Scenarios
    - $R_{in} \sim 10$  mm
    - **Vin**
    - Vin Portable
- Anticipated Performance
  - Acceptance:  $\cos(\theta) \sim 0.995$
  - **BMR  $\sim 3\%$**
  - **EM resolution  $3\%/\sqrt{E}$ , const. term  $< 1\%$**
  - Timing resolution  $\sim o(50)$  ps
  - $dP/P \sim 0.1\%$  in the barrel
  - Pid: **eff/purity  $> 96\%$**  for charged Kaon at hadronic Z event
  - Jet Flavor Tagging:
    - Tr(Mig): from  $\sim 2.4$  to  $\sim 2.7$
    - **Enhance the  $g(H_{cc})$  and  $|V_{cb}|$  measurements by 60% - 100%...**
  - **Fulfill the requirements of not only Higgs, but also Flavor & New Physics**

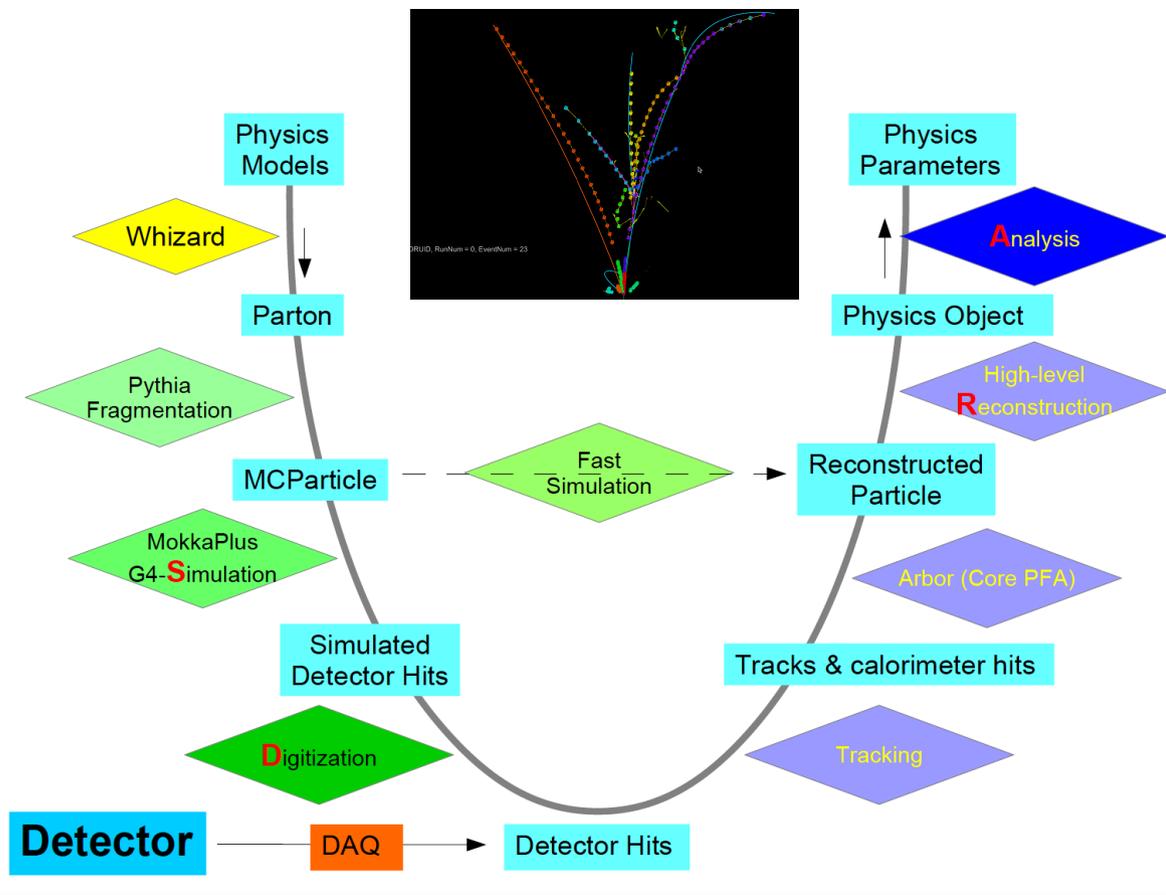
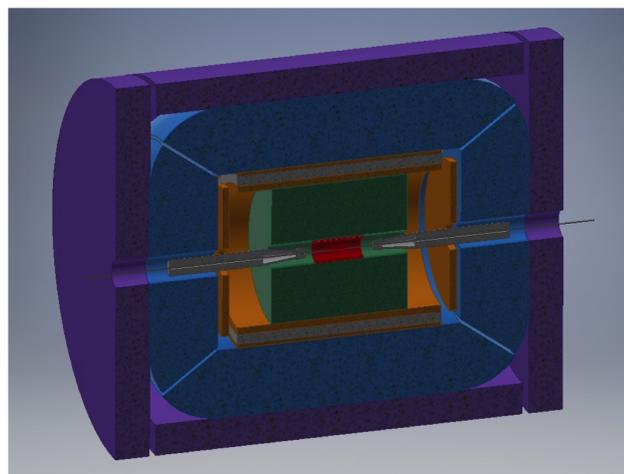
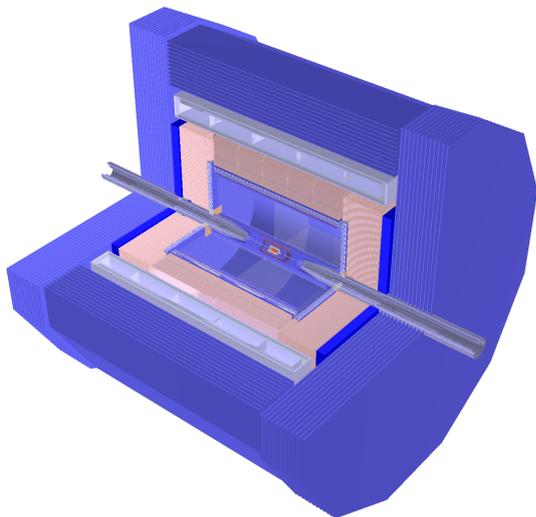
# Summary

- Critical Challenges
  - Boundary conditions to determine sub-detector technology & configuration...
    - Impact of beam background on sub detectors, especially gaseous one
    - MDI design, installation & integration
  - Vin
    - Power & Signal
    - Integration - Hom heat & radiation bkgrd, coating...
    - Vacuum level - material requirements
    - Large curvature stitch tech...
  - ECAL
    - Xstal:
      - Homogeneity, light yield – SiPM coupling, saturation;
      - Non cuboid Xstal manufactory & response
      - Energy/Position reconstruction & correction algorithm

# Summary

- Critical Challenges
  - ECAL
    - Position layer optimization:
      - specification (time, position, and potentially energy),
      - cooling requirement – material budget
  - HCAL
    - Requirement on homogeneity light yield & coupling to SiPM
    - Mass production of glass
  - **Need to understand the in-time leakage & off-time pile up**
- Action items
  - Optimization of geometry parameters via Detailed simulation + algorithm development... with machine learning, etc
  - R&D to address challenges...
  - Integration study

# Detector & Software



Full simulation reconstruction Chain with Arbor, iterating/validation with hardware studies



*Jet origin identification at  
future electron positron  
Higgs factory*

Yongfeng Zhu, Hao Liang, Chen Zhou, Huilin Qu,  
Manqi Ruan, etc