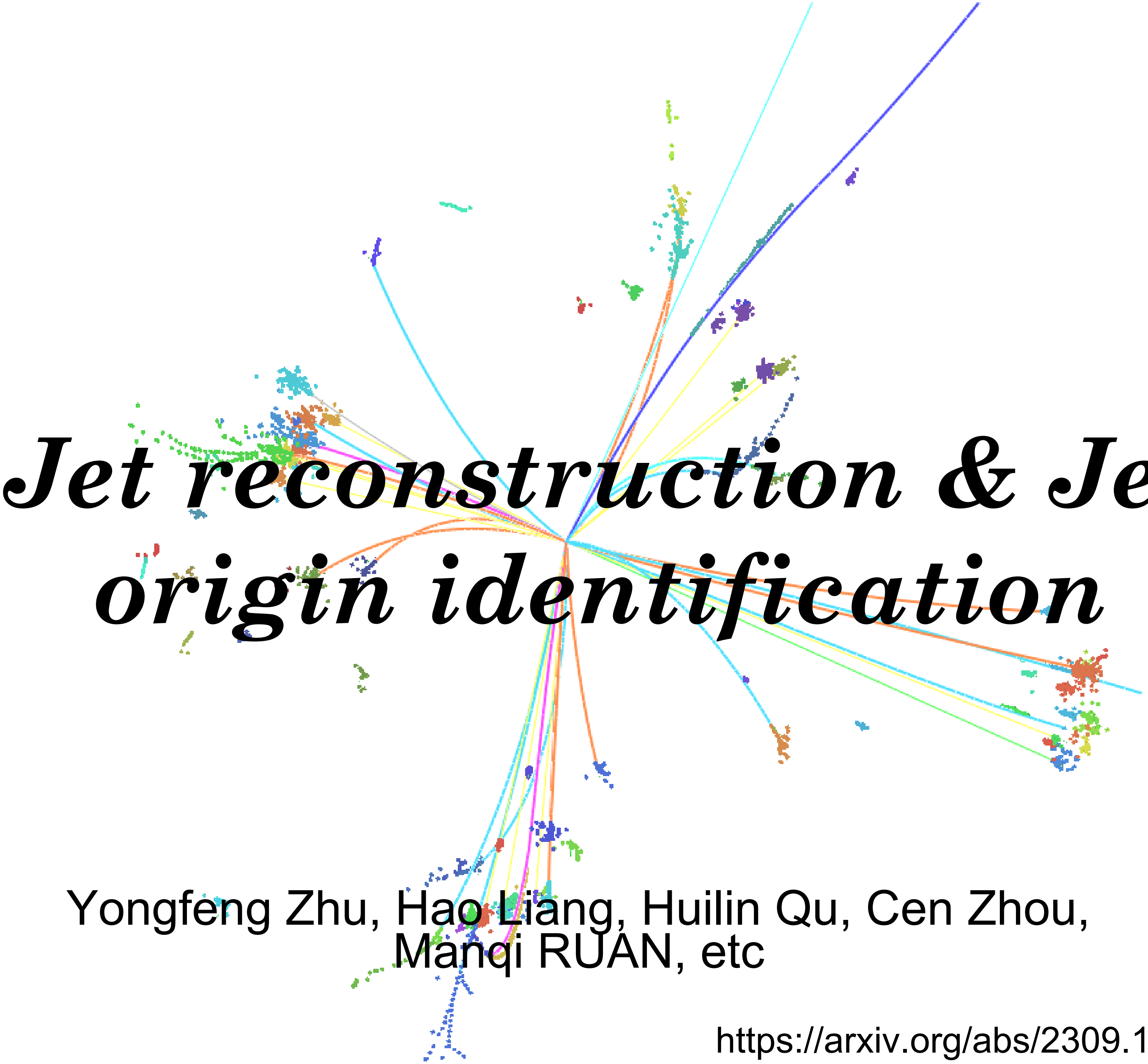
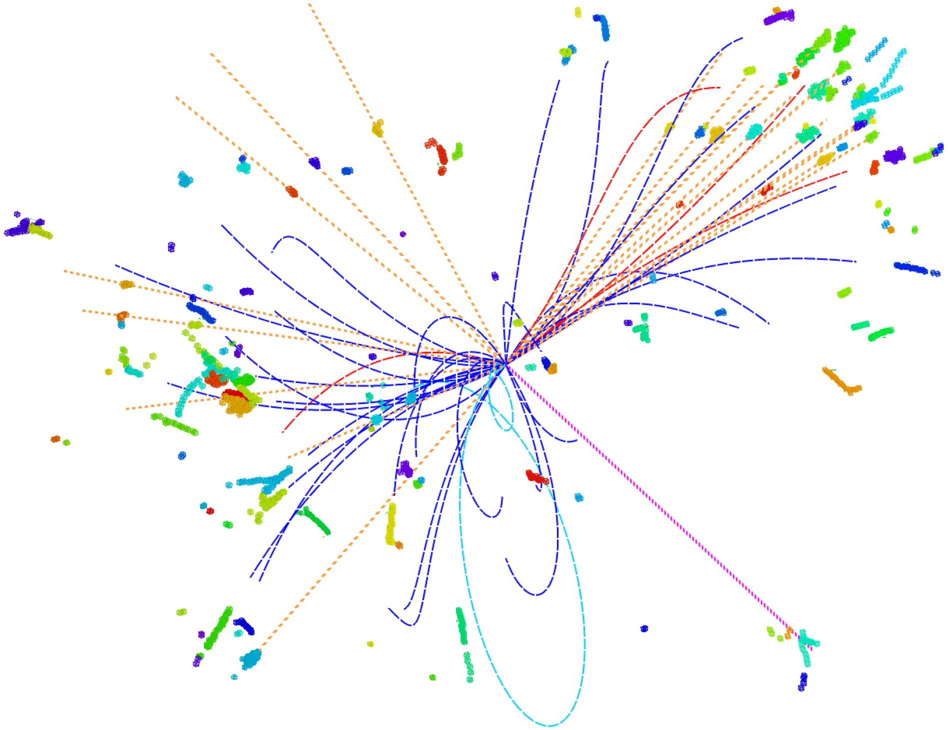


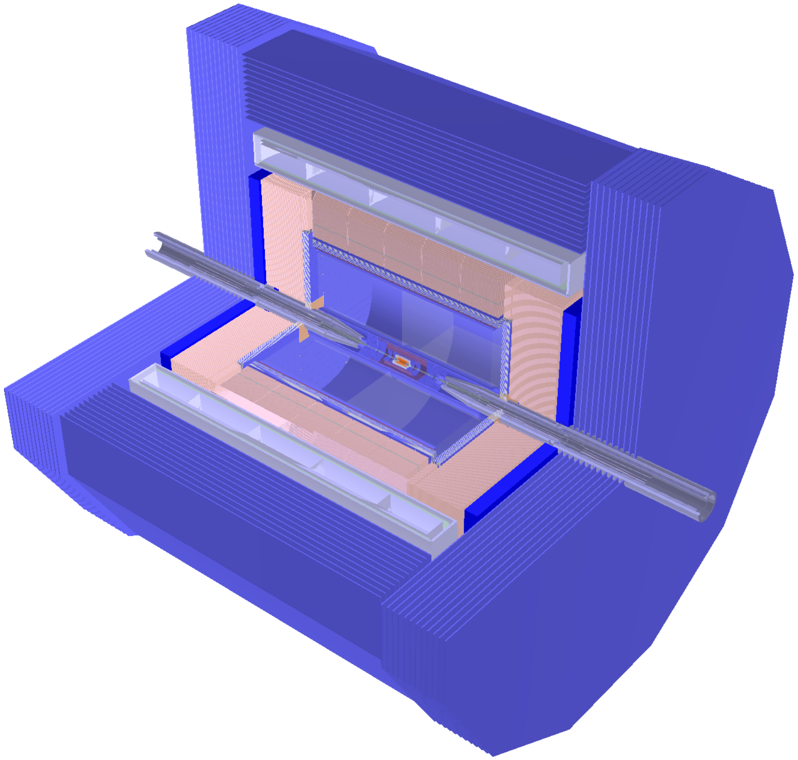
Jet reconstruction & Jet origin identification

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

https://arxiv.org/abs/2309.13231

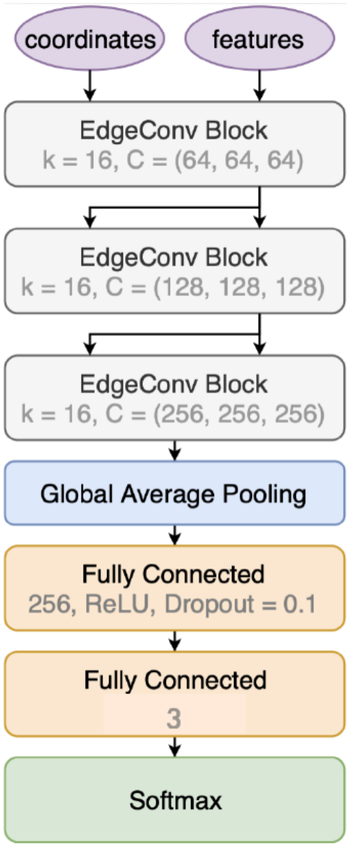
Geo. & Tools

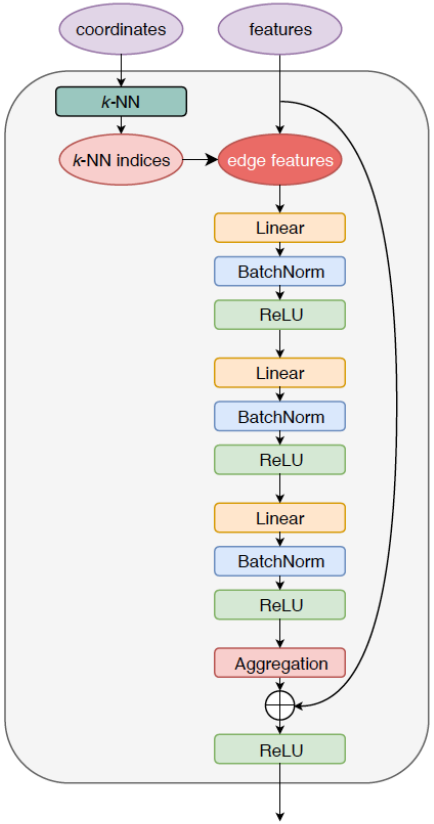


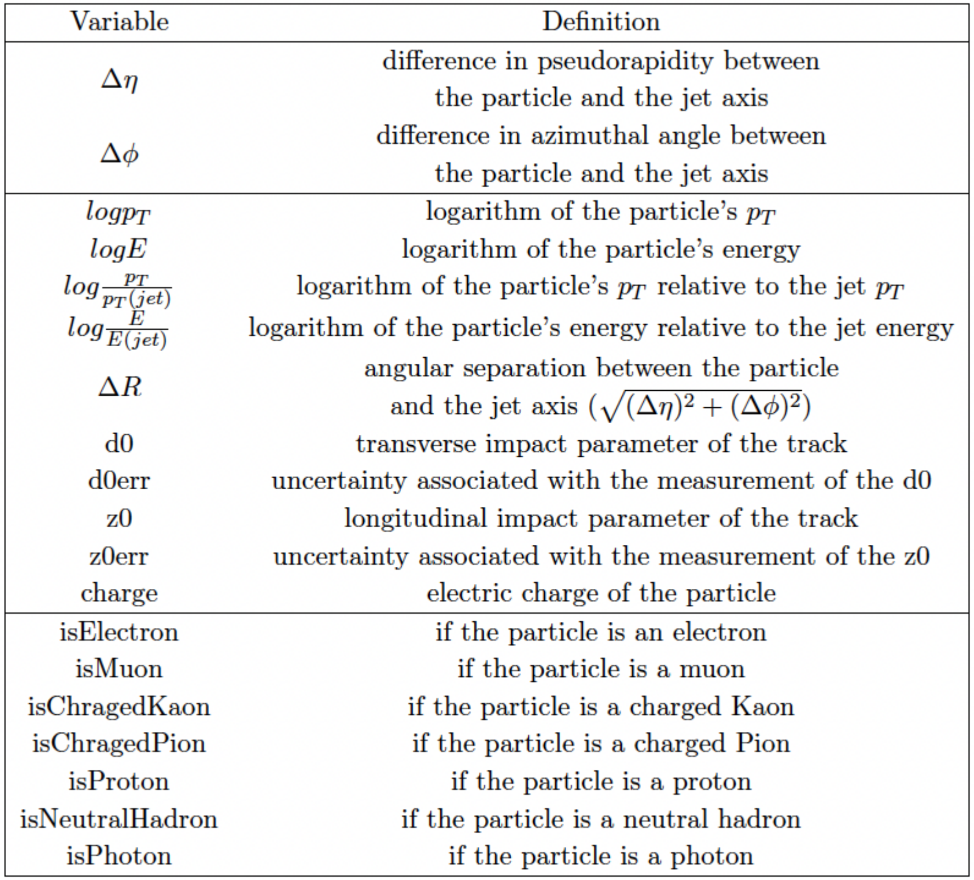


- Jet Flavor Tagging: 3 categories (b, c, light)
- Jet origin identification: 11 categories (5 quarks + 5 anti quarks + gluon)
  - Jet Flavor Tagging + Jet Charge measurements + ...
- Full Simulated Z/H to dijet sample at CEPC-v4 detector, reconstructed with Arbor.

Particle Net: IO


coordinates
features
EdgeConv Block
k = 16, C = (64, 64, 64)
EdgeConv Block
k = 16, C = (128, 128, 128)
EdgeConv Block
k = 16, C = (256, 256, 256)
Global Average Pooling
Fully Connected
256, ReLU, Dropout = 0.1
Fully Connected
3
Softmax


coordinates
features
k-NN
k-NN indices
edge features
Linear
BatchNorm
ReLU
Linear
BatchNorm
ReLU
Linear
BatchNorm
ReLU
Aggregation
ReLU

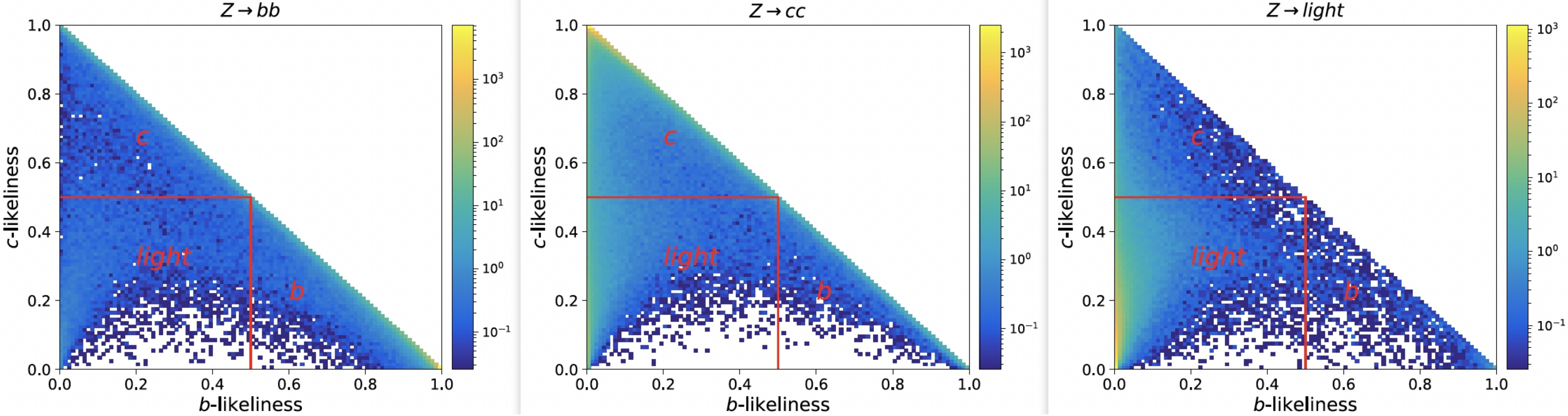


Variable	Definition
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis
$log p_T$	logarithm of the particle's p_T
$log E$	logarithm of the particle's energy
$log \frac{p_T}{p_T(jet)}$	logarithm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

- Output: likelihoods to different categories

Three categories: b, c, & light



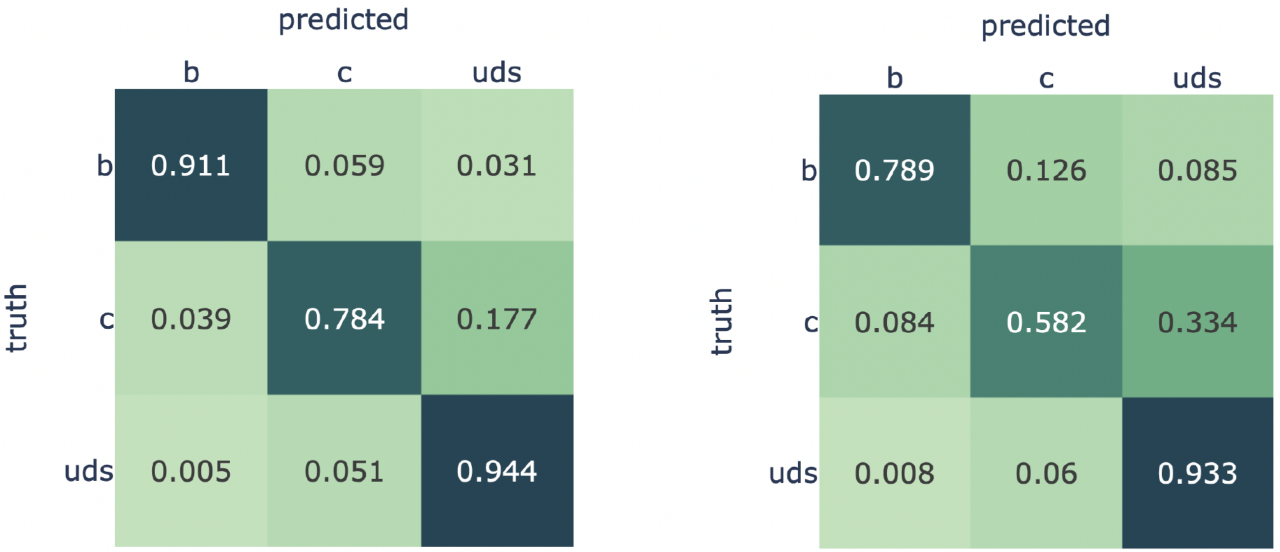
Hadronic Z pole sample

1 M Z→bb, cc, (uds) each

60/20/20% for

training/validating/testing.

Result on Testing sample

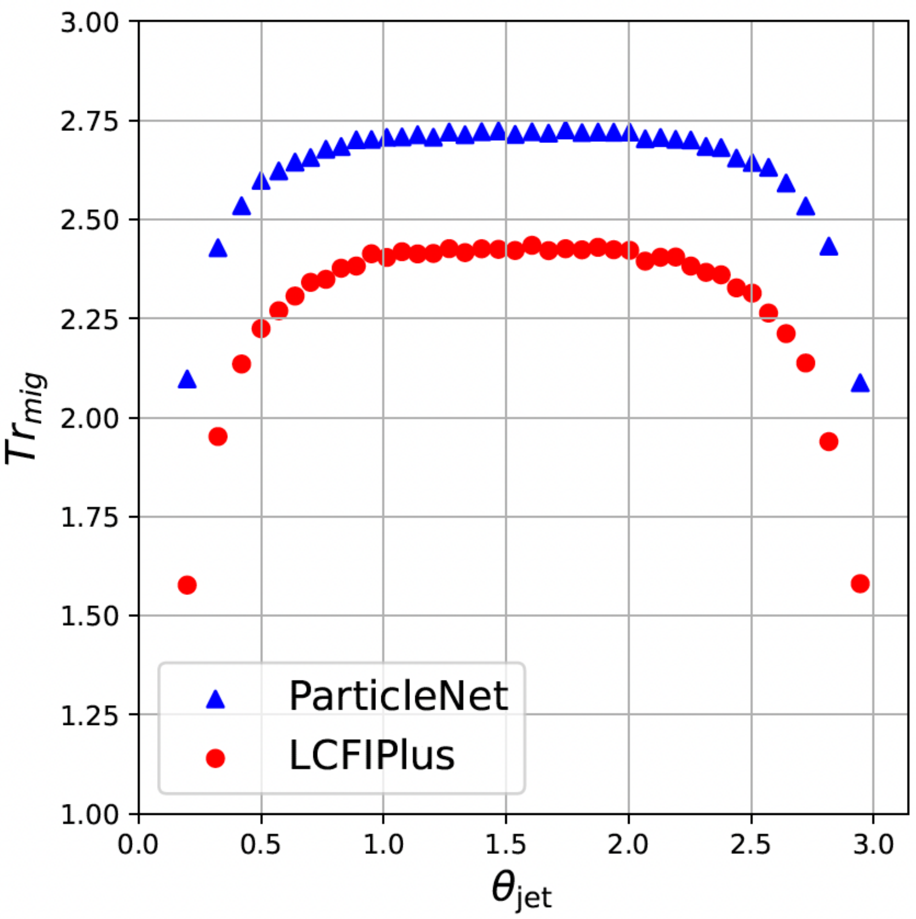


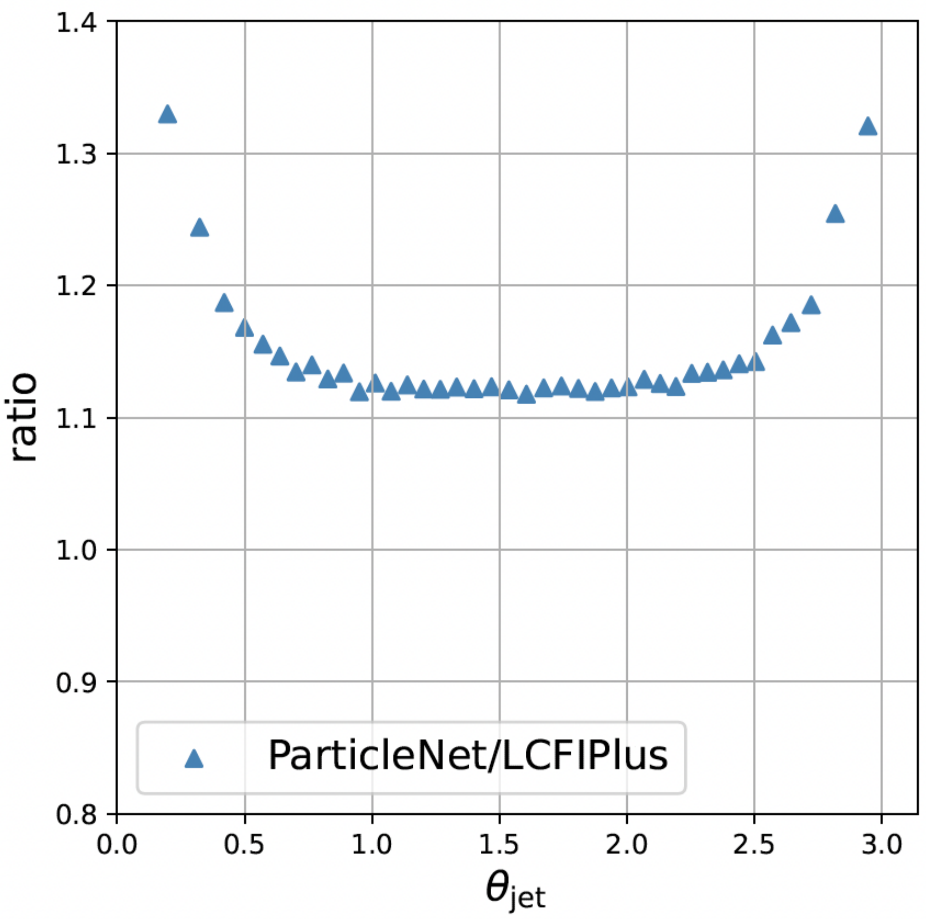
truth \ predicted	b	c	uds
b	0.911	0.059	0.031
c	0.039	0.784	0.177
uds	0.005	0.051	0.944

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

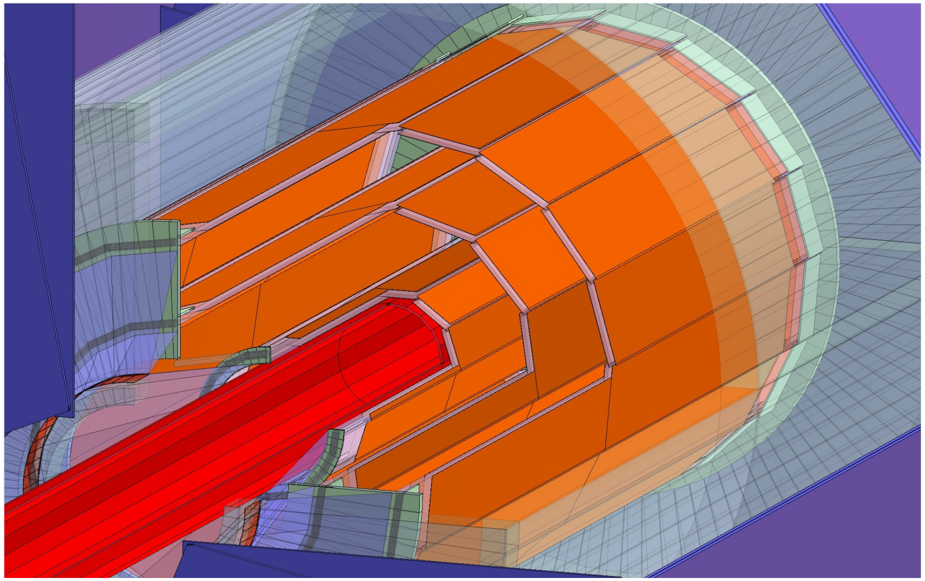
Figure 7. The migration matrix of ParticleNet (left) and LCFIPlus (right) at the CEPC.

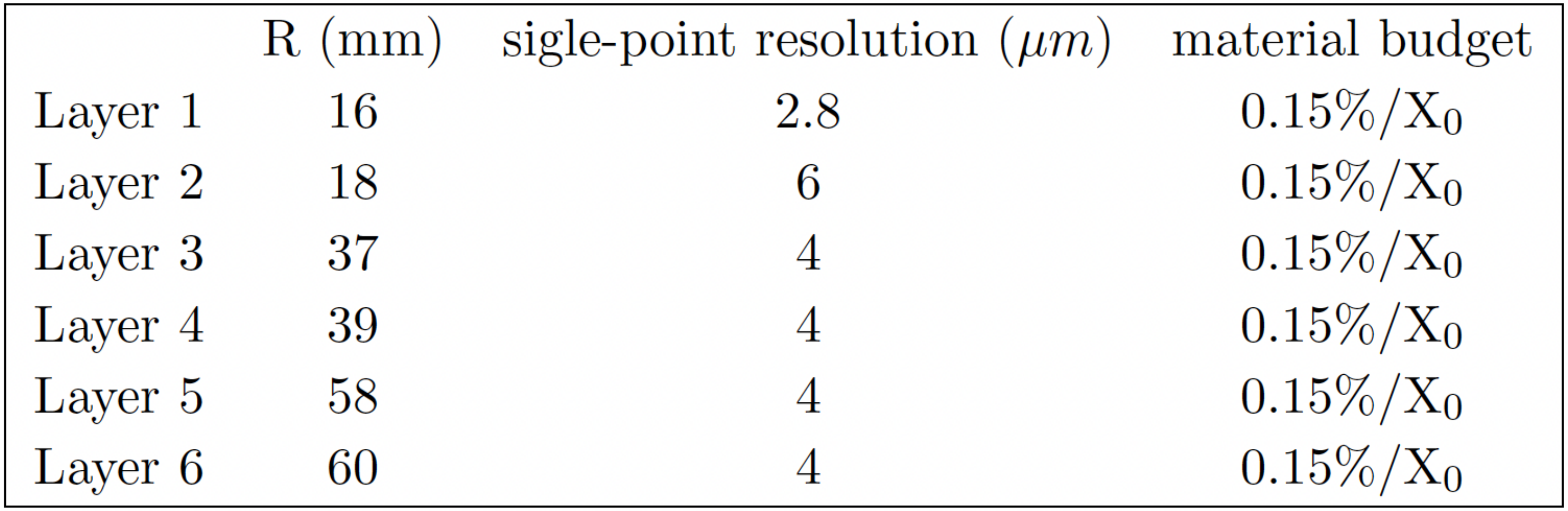
Dependence on polar angle





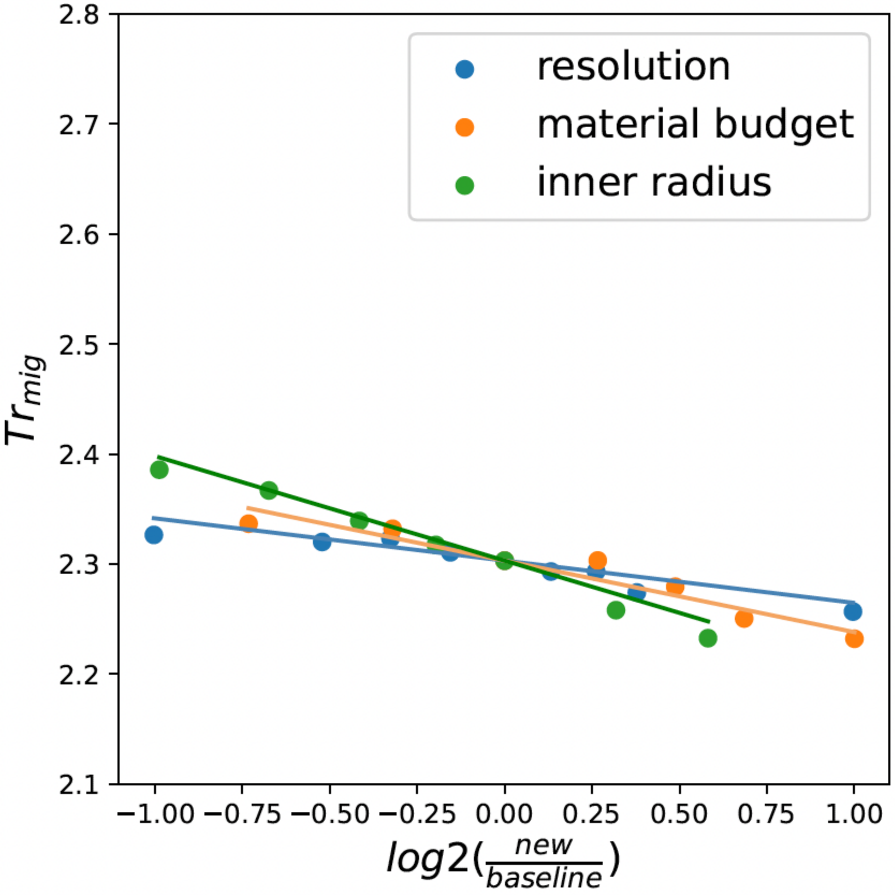
Comparison on Det. Optimization

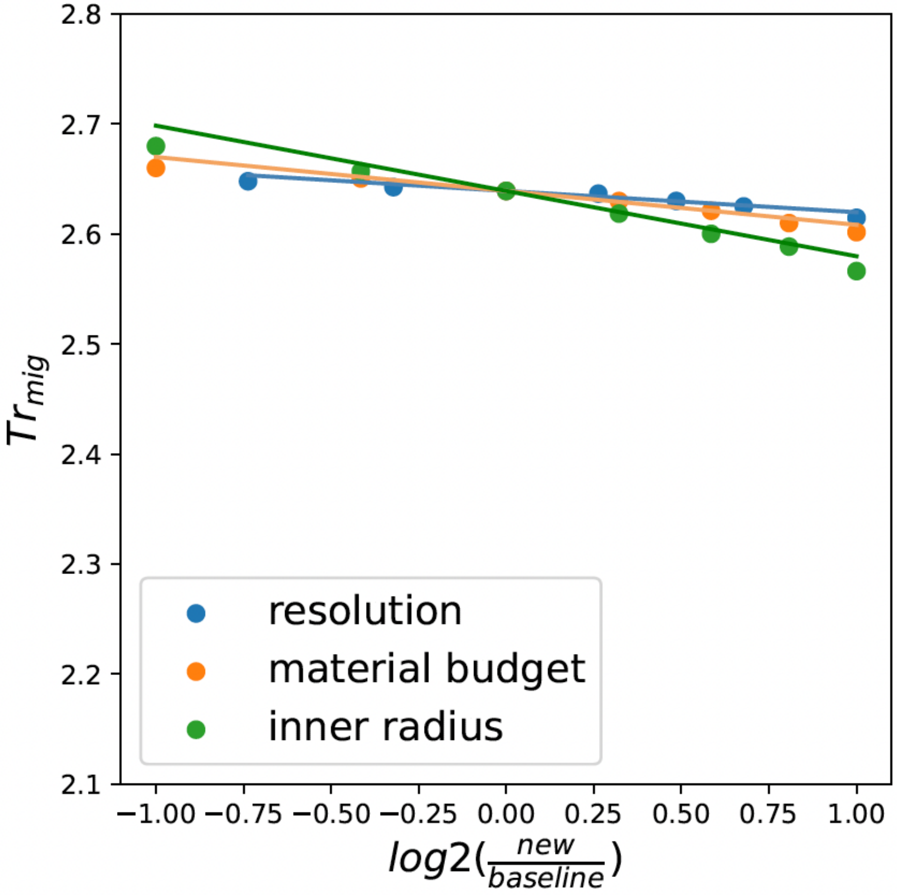




	R (mm)	sigle-point resolution (μm)	material budget
Layer 1	16	2.8	0.15%/X_0
Layer 2	18	6	0.15%/X_0
Layer 3	37	4	0.15%/X_0
Layer 4	39	4	0.15%/X_0
Layer 5	58	4	0.15%/X_0
Layer 6	60	4	0.15%/X_0

Comparison on Det. Optimization



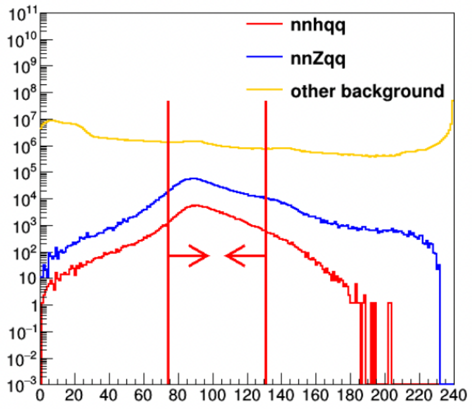


$$Tr_{mig} = 2.30 + 0.06 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2 \frac{R^0_{radius}}{R_{radius}} \qquad (4.1)$$

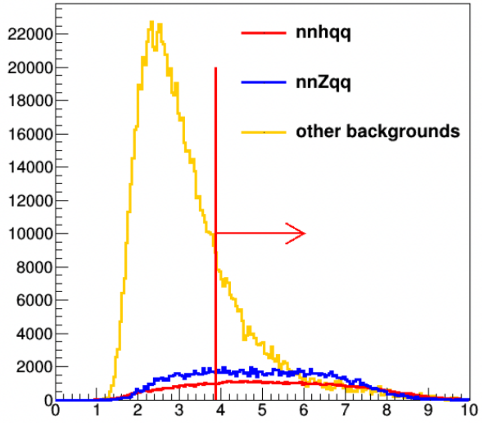
$$Tr_{mig} = 2.64 + 0.03 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.02 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.06 \cdot log_2 \frac{R^0_{radius}}{R_{radius}} \qquad (4.2)$$

Impact on benchmark: ννH, H→jets

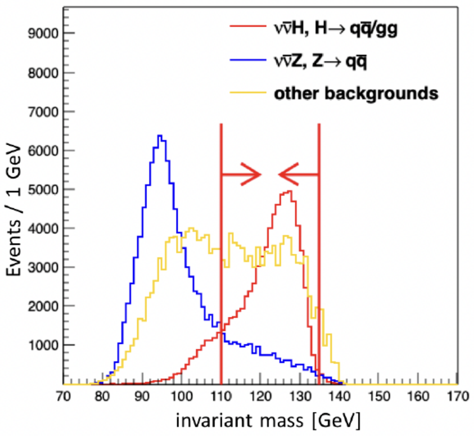
	$\nu\nu Hq\bar{q}/gg$	2f	SW	SZ	WW	ZZ	Mixed	ZH	$\frac{\sqrt{S+B}}{S}$ (%)
total	178890	8.01E8	1.95E7	9.07E6	5.08E7	6.39E6	2.18E7	961606	16.86
recoilMass (GeV) $\in (74, 131)$	157822	5.11E7	2.17E6	1.38E6	4.78E6	1.30E6	1.08E6	74991	4.99
visEn (GeV) $\in (109, 143)$	142918	2.37E7	1.35E6	8.81E5	3.60E6	1.03E6	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

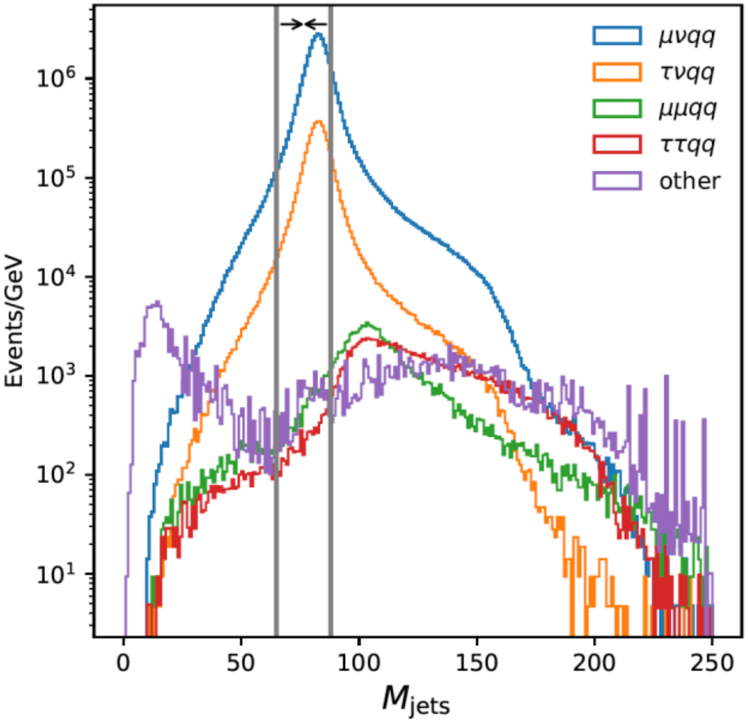


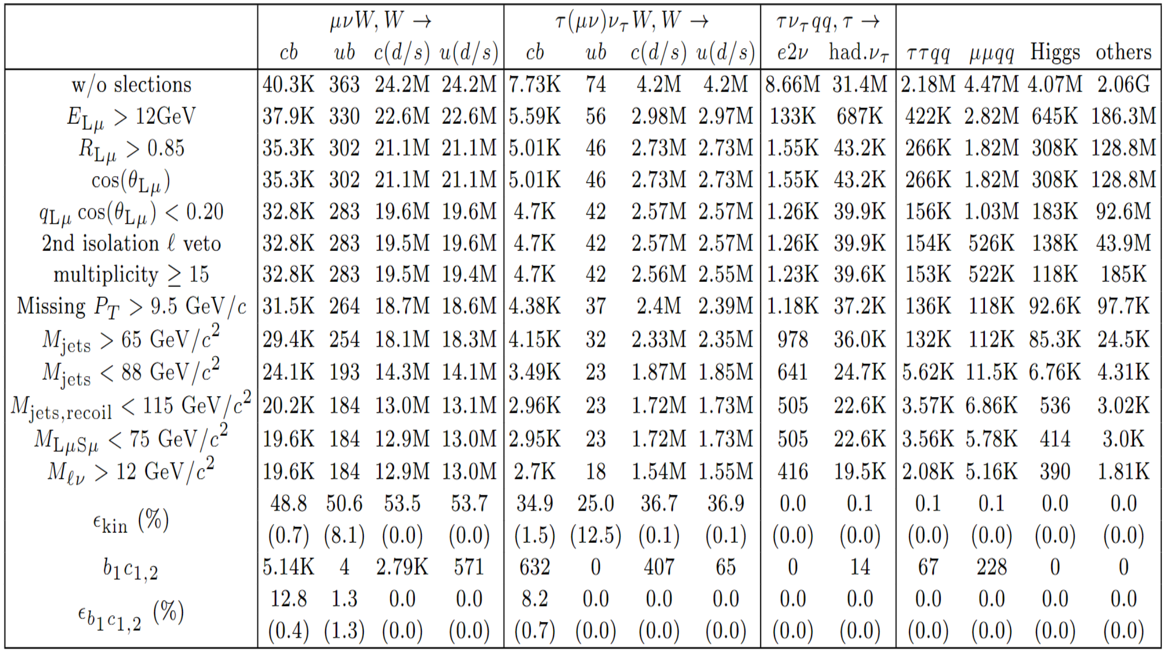
-log(Y23) > 3.875



Thanks to BMR ~ 3.8%!

Vcb from W decay

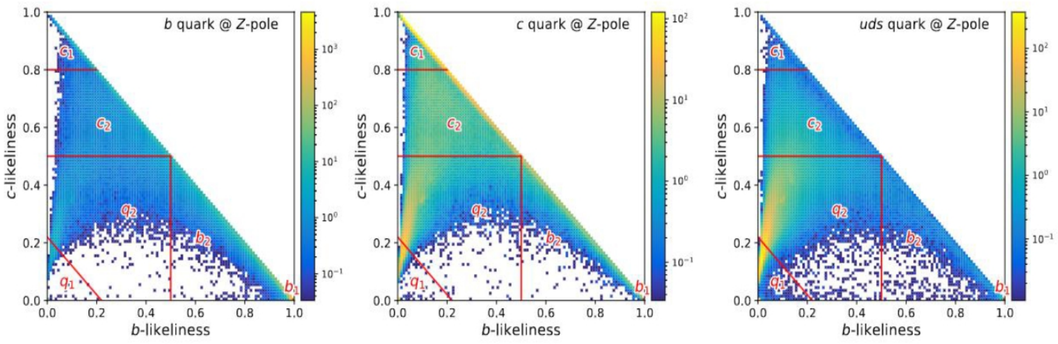




	$\mu\nu W, W \rightarrow$				$\tau(\mu\nu)\nu_\tau W, W \rightarrow$				$\tau\nu_\tau qq, \tau \rightarrow$					
	cb	ub	$c(d/s)$	$u(d/s)$	cb	ub	$c(d/s)$	$u(d/s)$	$e2\nu$	had.ν_τ	$\tau\tau qq$	$\mu\mu qq$	Higgs	others
w/o slections	40.3K	363	24.2M	24.2M	7.73K	74	4.2M	4.2M	8.66M	31.4M	2.18M	4.47M	4.07M	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)
$b_1 c_{1,2}$	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)

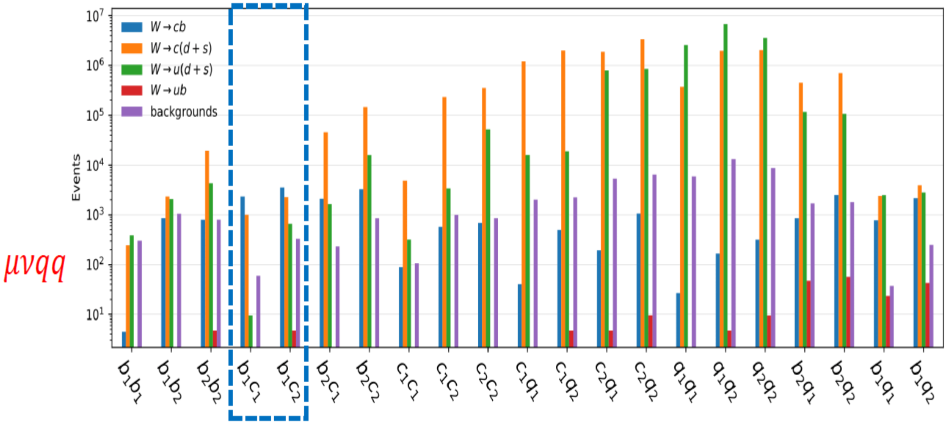
- Purity > 99.5% at Eff. 50% for $\mu\nu qq$ and 34% for $\tau(\mu 2\nu)\nu qq$
- Main backgrounds include:
  - $W \rightarrow c(d/s)$
  - $\mu\mu qq$

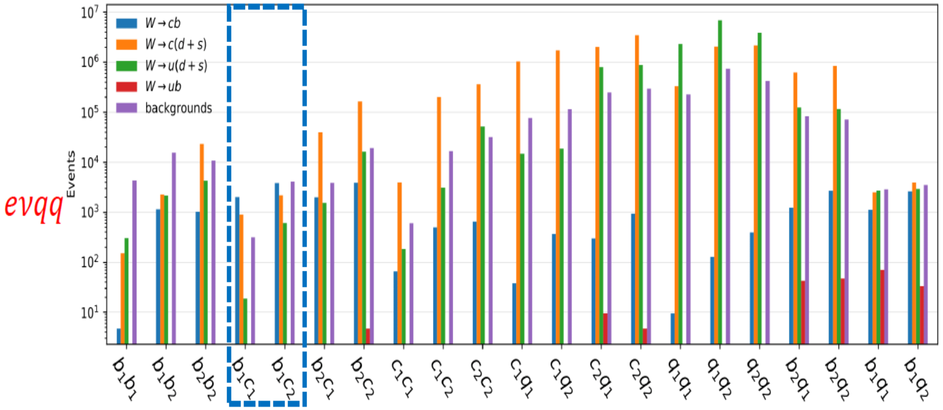
Vcb from W decay



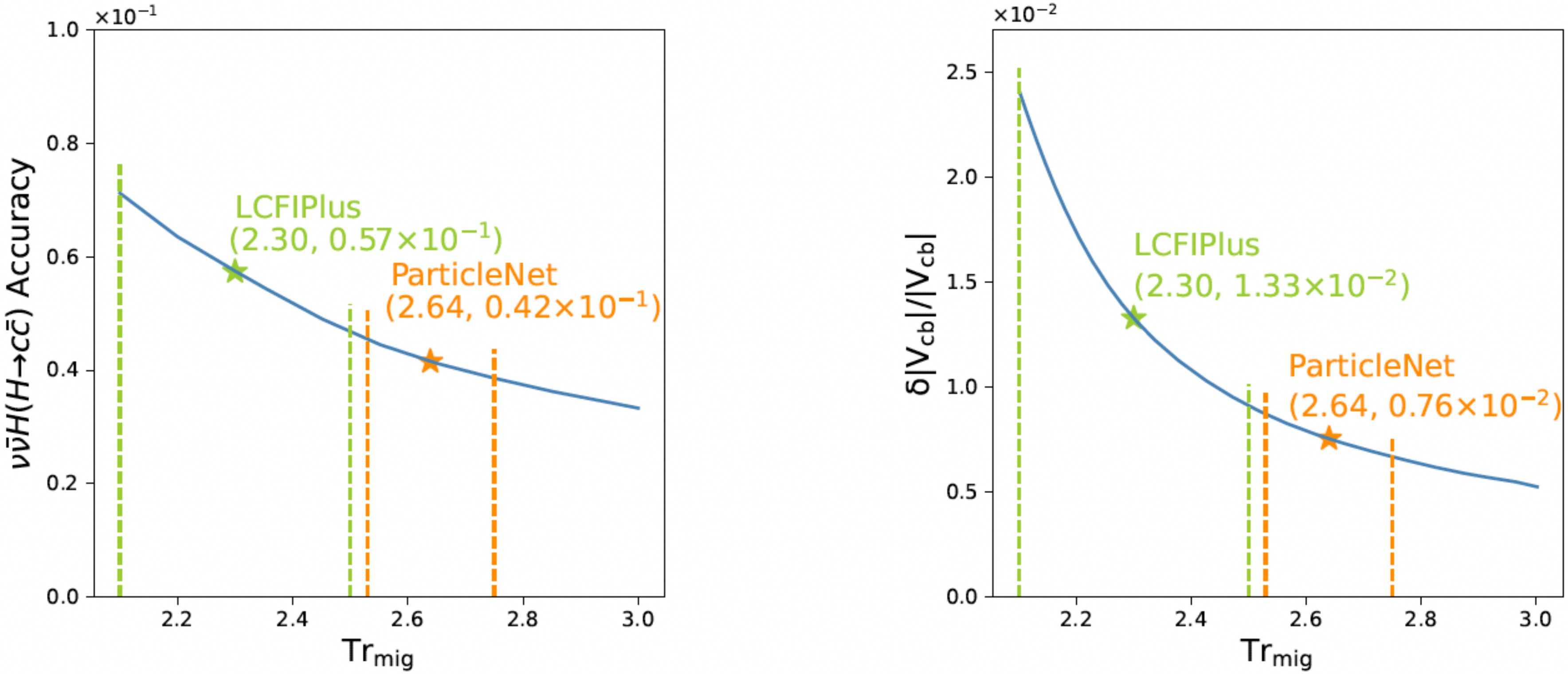
quark \ tag	b_1	b_2	c_1	c_2	q_1	q_2
b	0.47	0.378	0.0197	0.0965	0.00397	0.0315
c	0.00042	0.078	0.298	0.373	0.0682	0.182
uds	0.000104	0.00477	0.00145	0.054	0.538	0.401

- $\mu\nu qq$
  - 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%


$\mu\nu qq$


$e\nu qq$

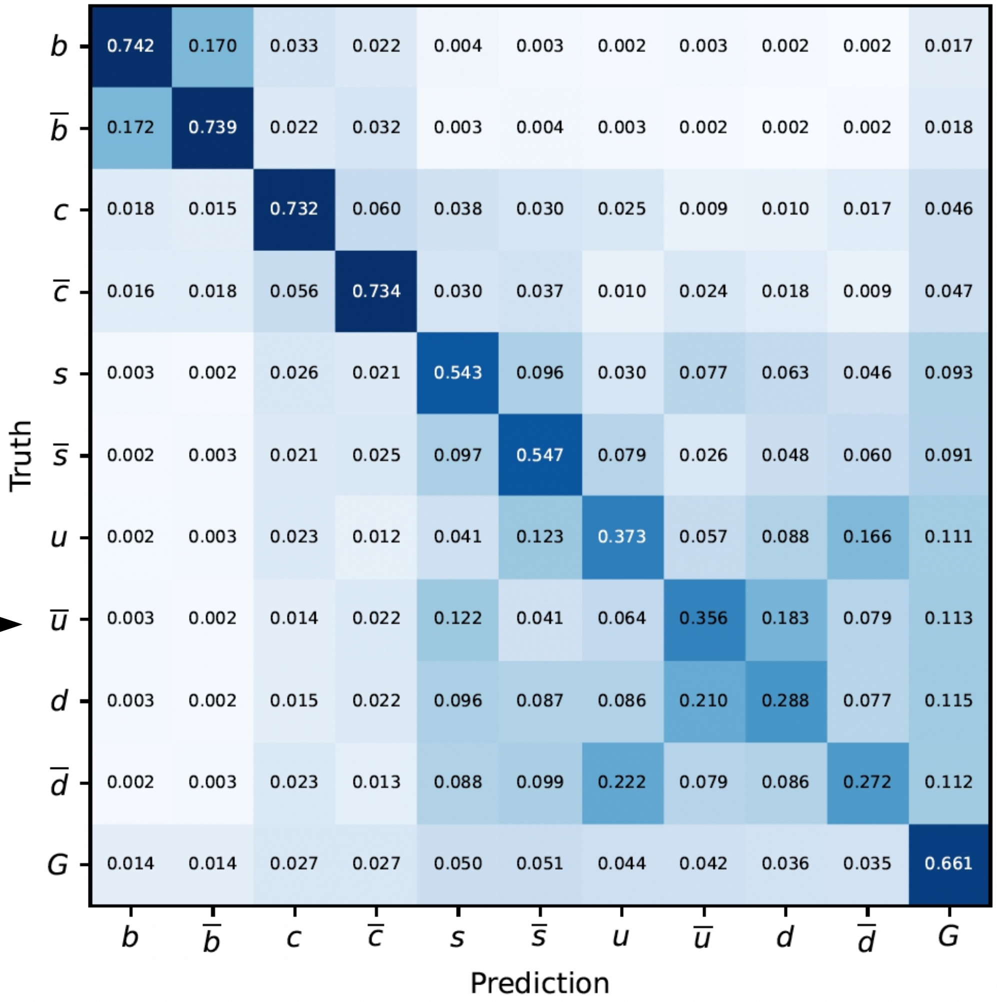
Impact on physics benchmarks



		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

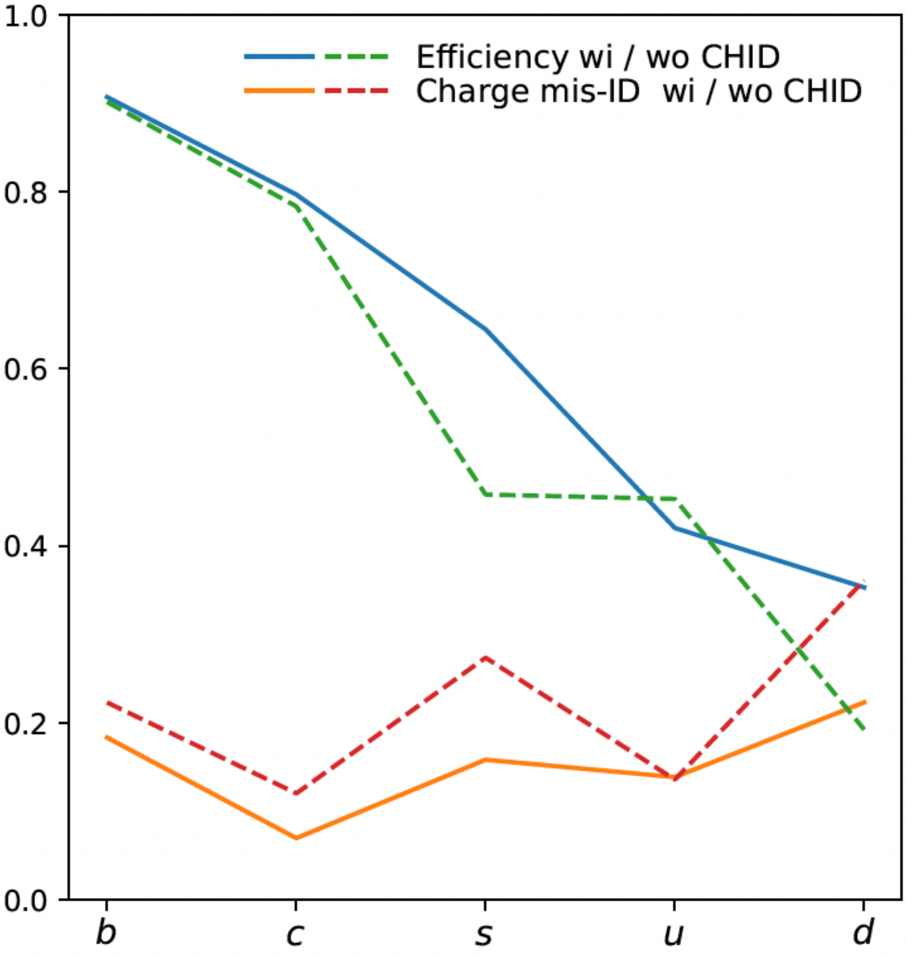
Jet origin id: 11 categories

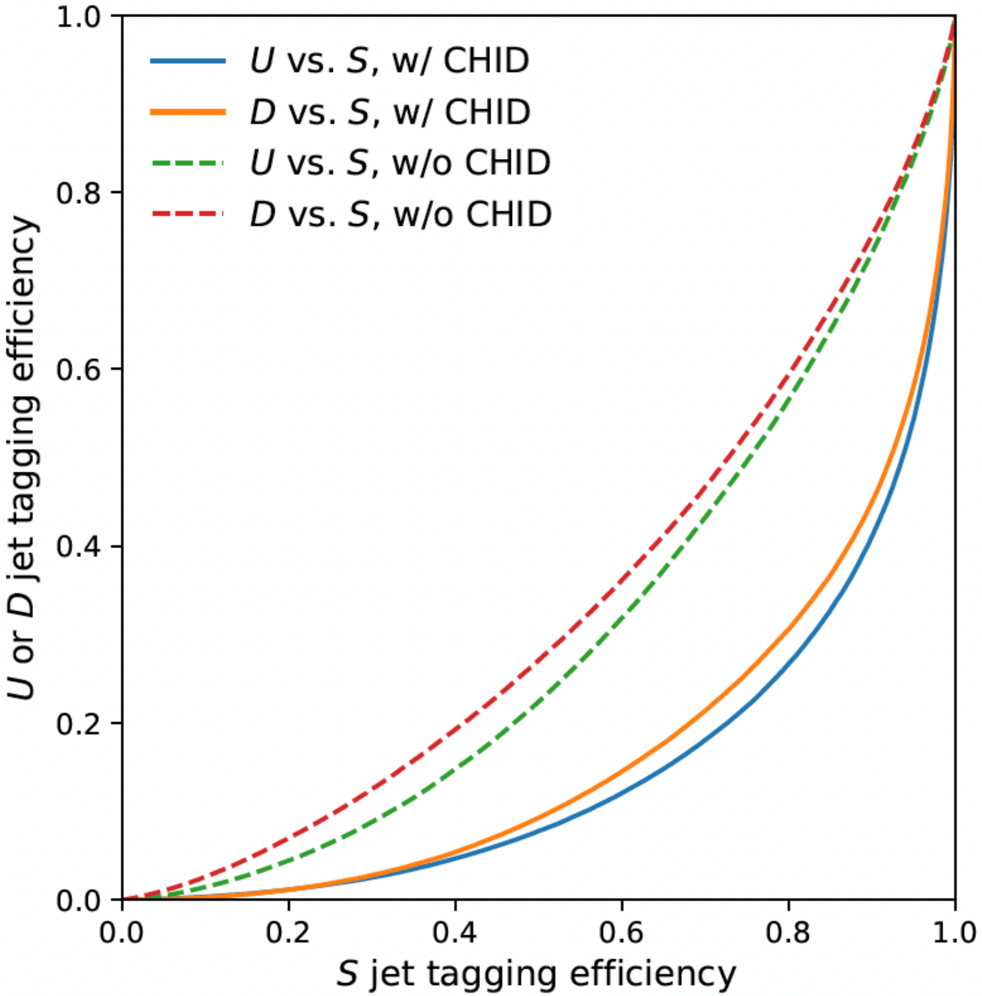
- vvH sample, with Higgs decays into different species of colored particle: 5 quark, 5 antiquark & gluon
  - 1 Million of each type
  - 60/20/20% for training, validating, and testing, result corresponding to testing sample
- Pid: ideal Pid – three categories
  - Lepton identification
  - Charged Kaon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



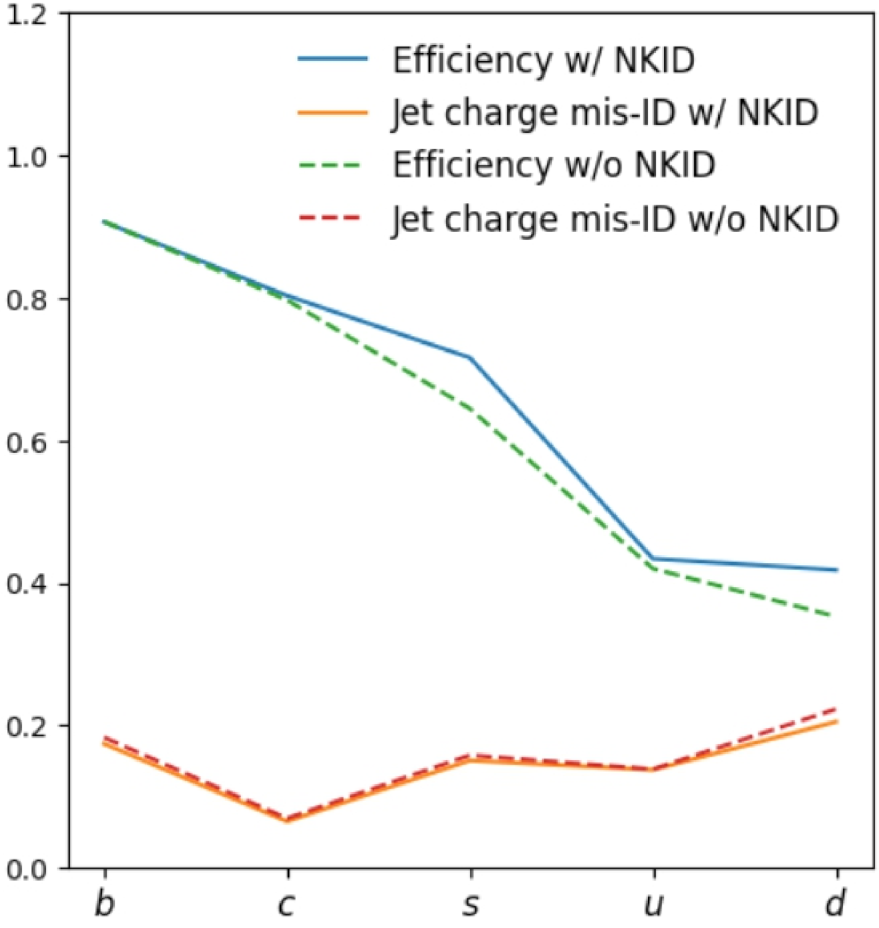
Truth \ Prediction	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.742	0.170	0.033	0.022	0.004	0.003	0.002	0.003	0.002	0.002	0.017
$\bar{b}$	0.172	0.739	0.022	0.032	0.003	0.004	0.003	0.002	0.002	0.002	0.018
c	0.018	0.015	0.732	0.060	0.038	0.030	0.025	0.009	0.010	0.017	0.046
$\bar{c}$	0.016	0.018	0.056	0.734	0.030	0.037	0.010	0.024	0.018	0.009	0.047
s	0.003	0.002	0.026	0.021	0.543	0.096	0.030	0.077	0.063	0.046	0.093
$\bar{s}$	0.002	0.003	0.021	0.025	0.097	0.547	0.079	0.026	0.048	0.060	0.091
u	0.002	0.003	0.023	0.012	0.041	0.123	0.373	0.057	0.088	0.166	0.111
$\bar{u}$	0.003	0.002	0.014	0.022	0.122	0.041	0.064	0.356	0.183	0.079	0.113
d	0.003	0.002	0.015	0.022	0.096	0.087	0.086	0.210	0.288	0.077	0.115
$\bar{d}$	0.002	0.003	0.023	0.013	0.088	0.099	0.222	0.079	0.086	0.272	0.112
G	0.014	0.014	0.027	0.027	0.050	0.051	0.044	0.042	0.036	0.035	0.661

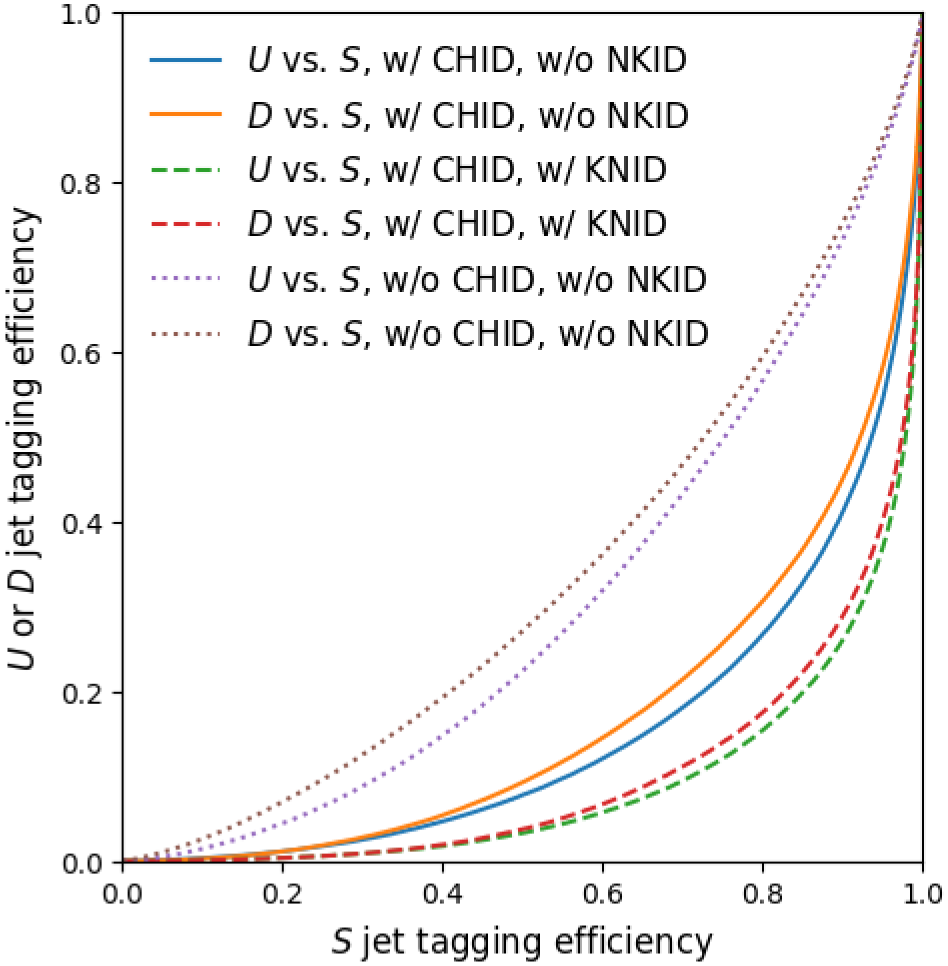
Impact of charged kaon id





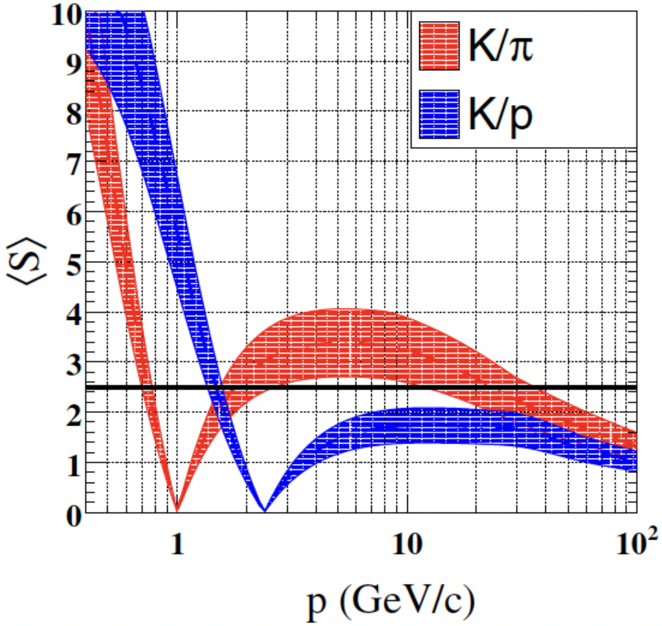
Neutral Kaon id

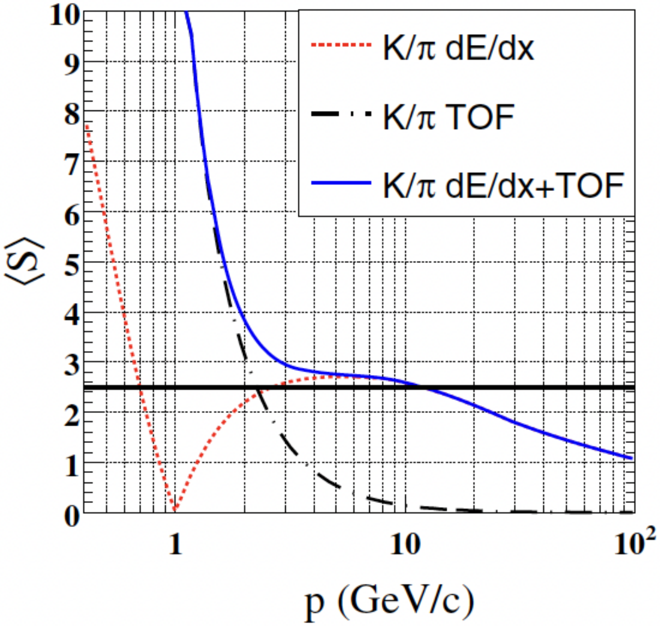


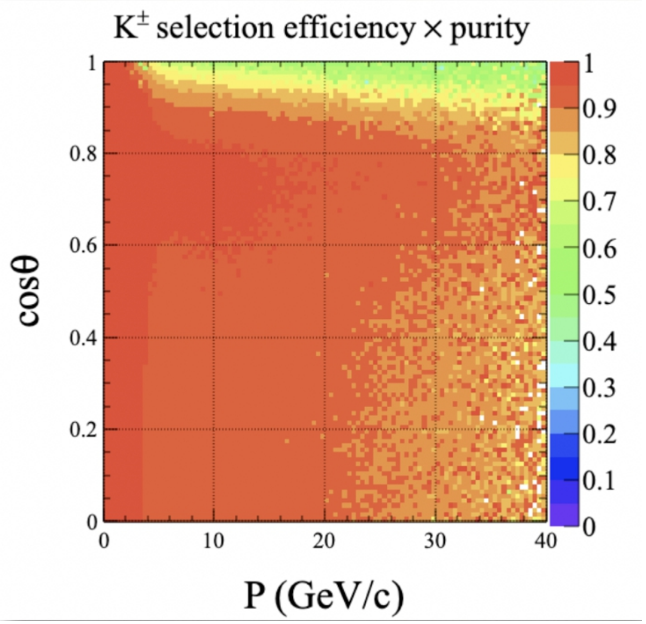


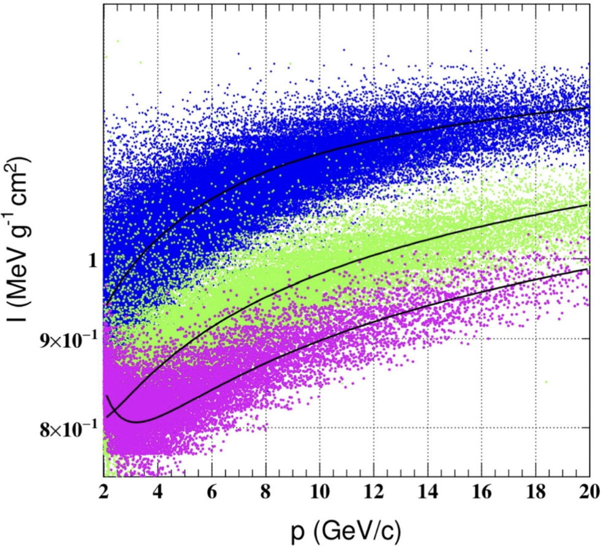
- Current tool (PN) is not clever enough to figure out Ks->2 pi, etc

Tracker: Pid









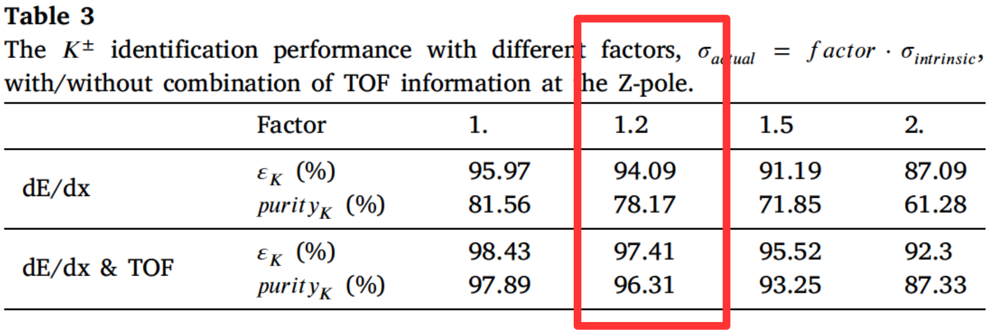


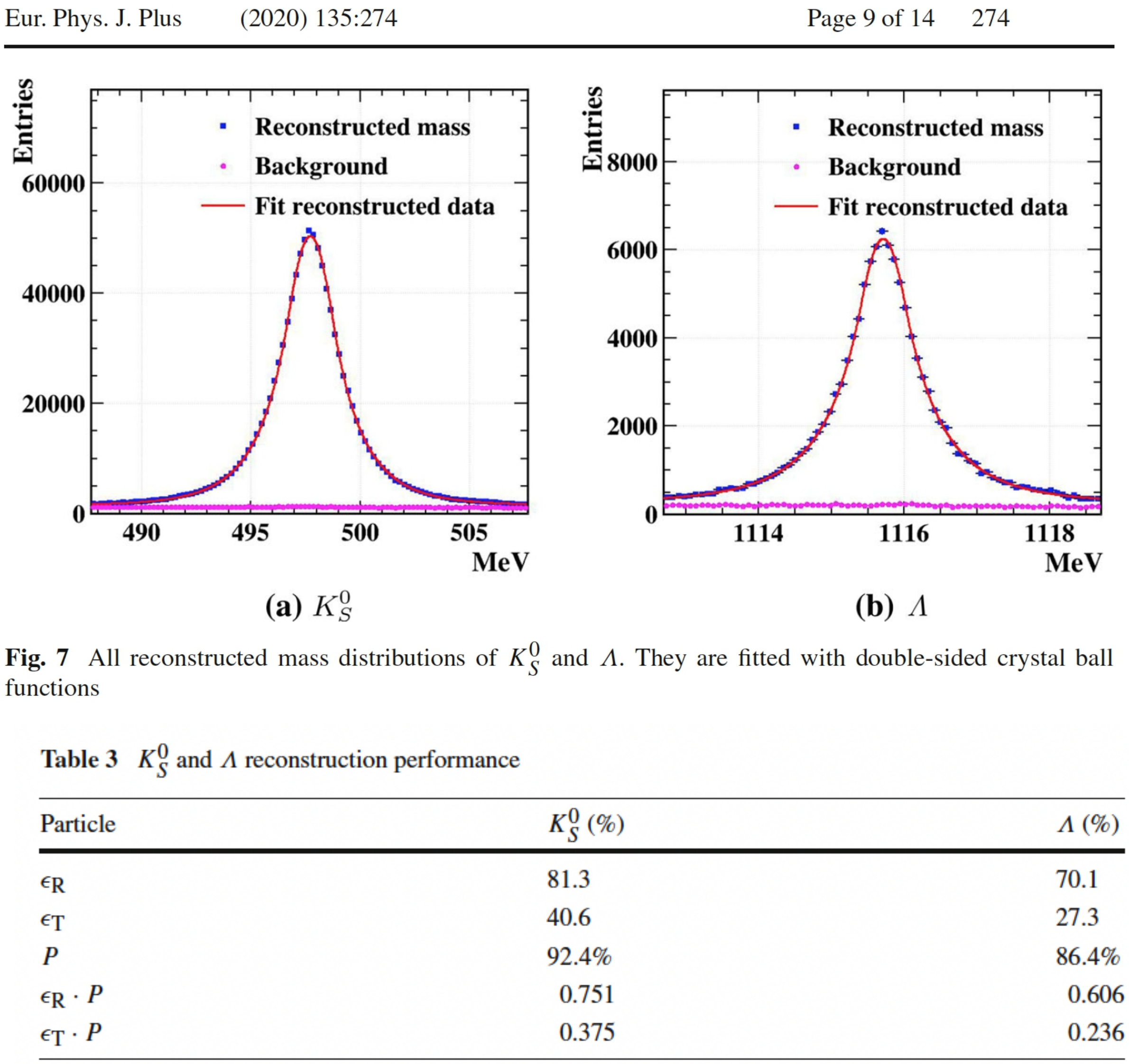
Table 3

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

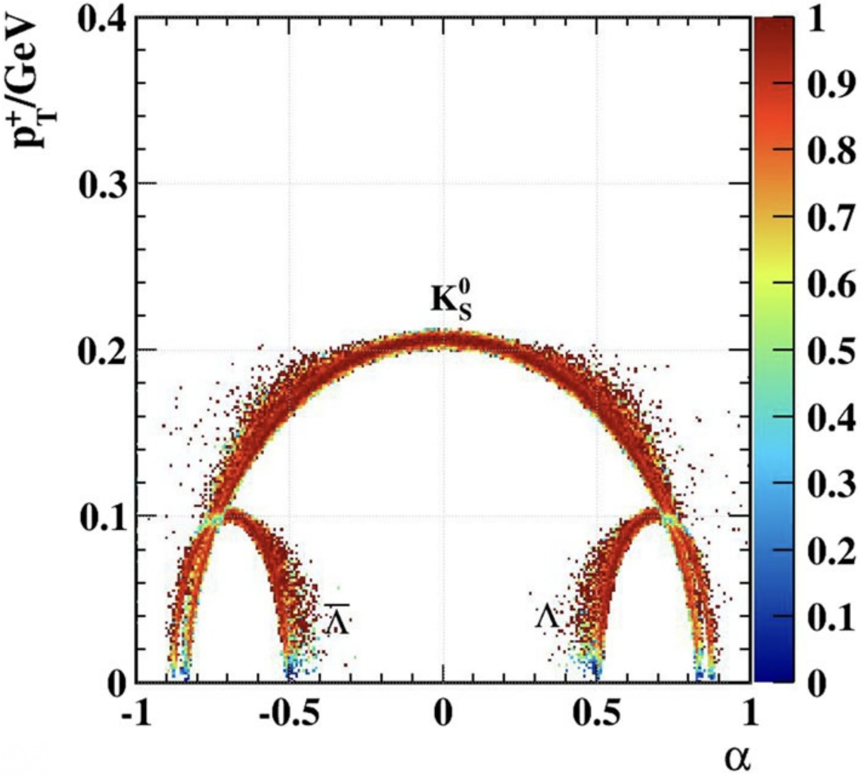

Eur. Phys. J. Plus (2020) 135:274 Page 9 of 14 274

(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

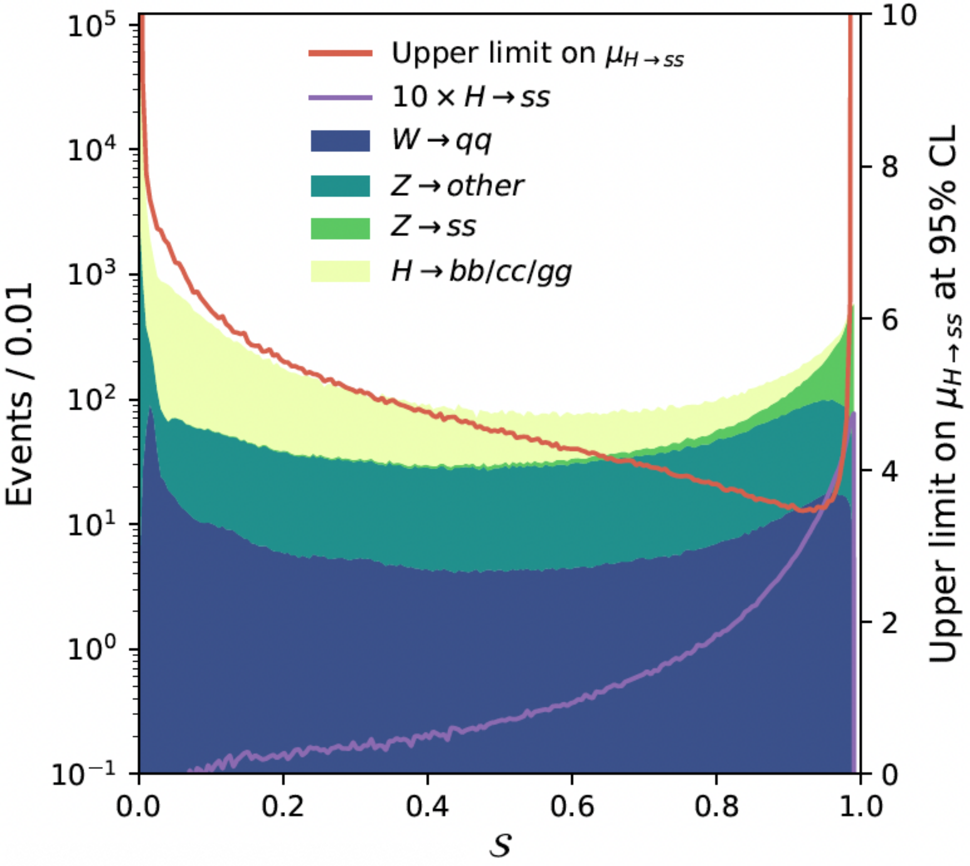
Table 3 K_S^0 and Λ reconstruction performance

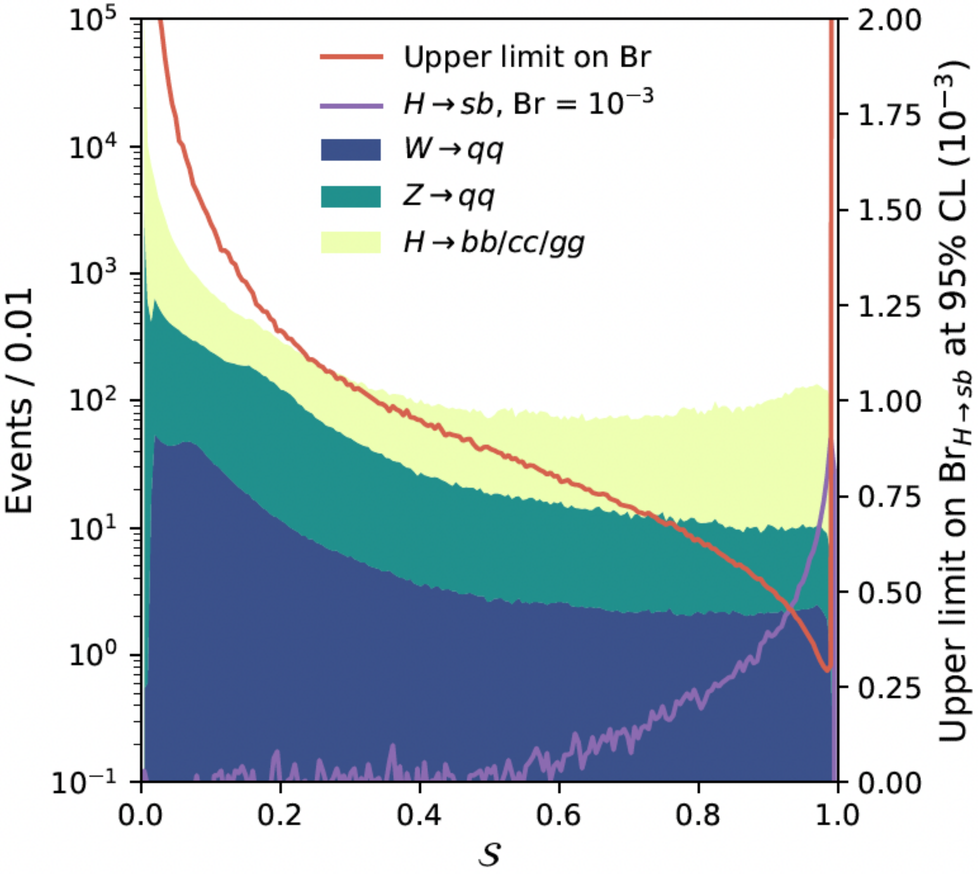
Particle	K_S^0 (%)	Λ (%)
ϵ_R	81.3	70.1
ϵ_T	40.6	27.3
P	92.4%	86.4%
$\epsilon_R \cdot P$	0.751	0.606
$\epsilon_T \cdot P$	0.375	0.236

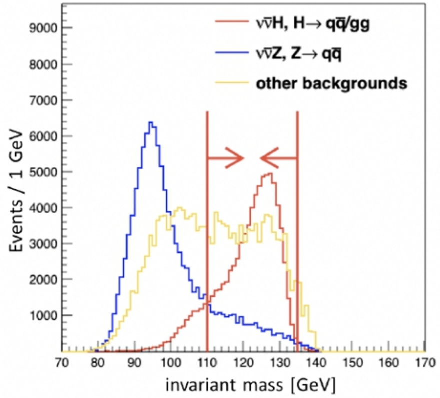


High eff/purity reco. of charged Final states...

Benchmark analyses using Jet origin ID

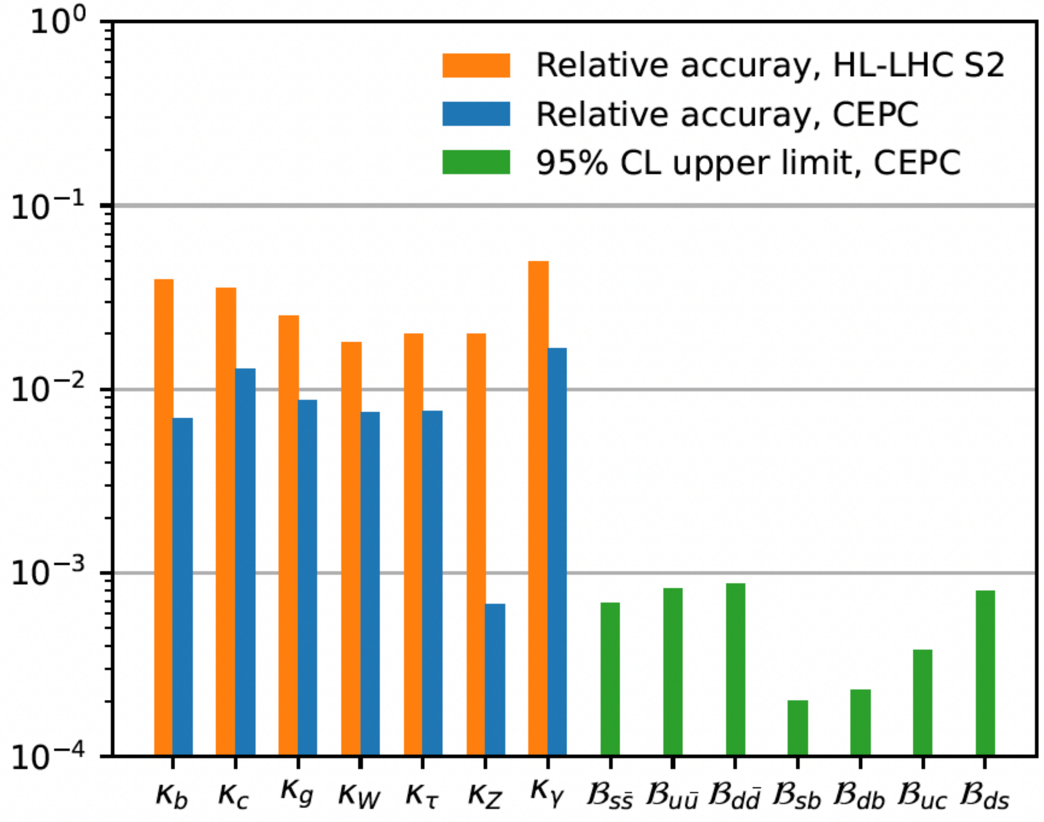






Applied to quasi-data of vvH;
H → ss: be limited to 3*SM using vvH + llH at 20 iab
H → sb: up limit of 2E-4 at 95% C.L.

Benchmark analyses using Jet origin ID



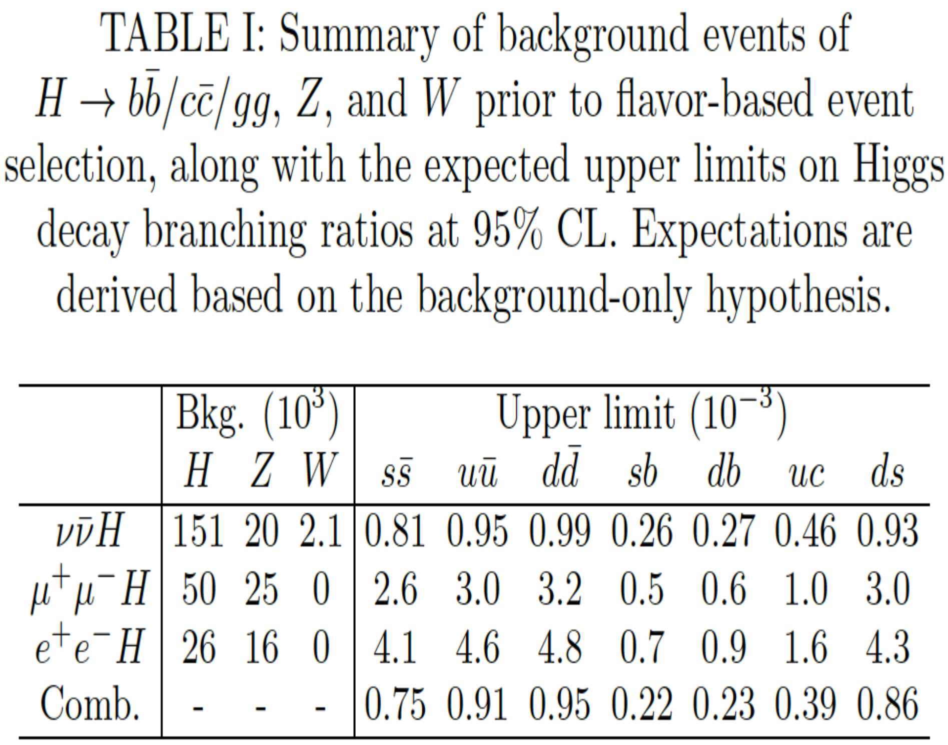


TABLE I: Summary of background events of $H \to b\bar{b}/c\bar{c}/gg$, Z, and W prior to flavor-based event selection, along with the expected upper limits on Higgs decay branching ratios at 95% CL. Expectations are derived based on the background-only hypothesis.

	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

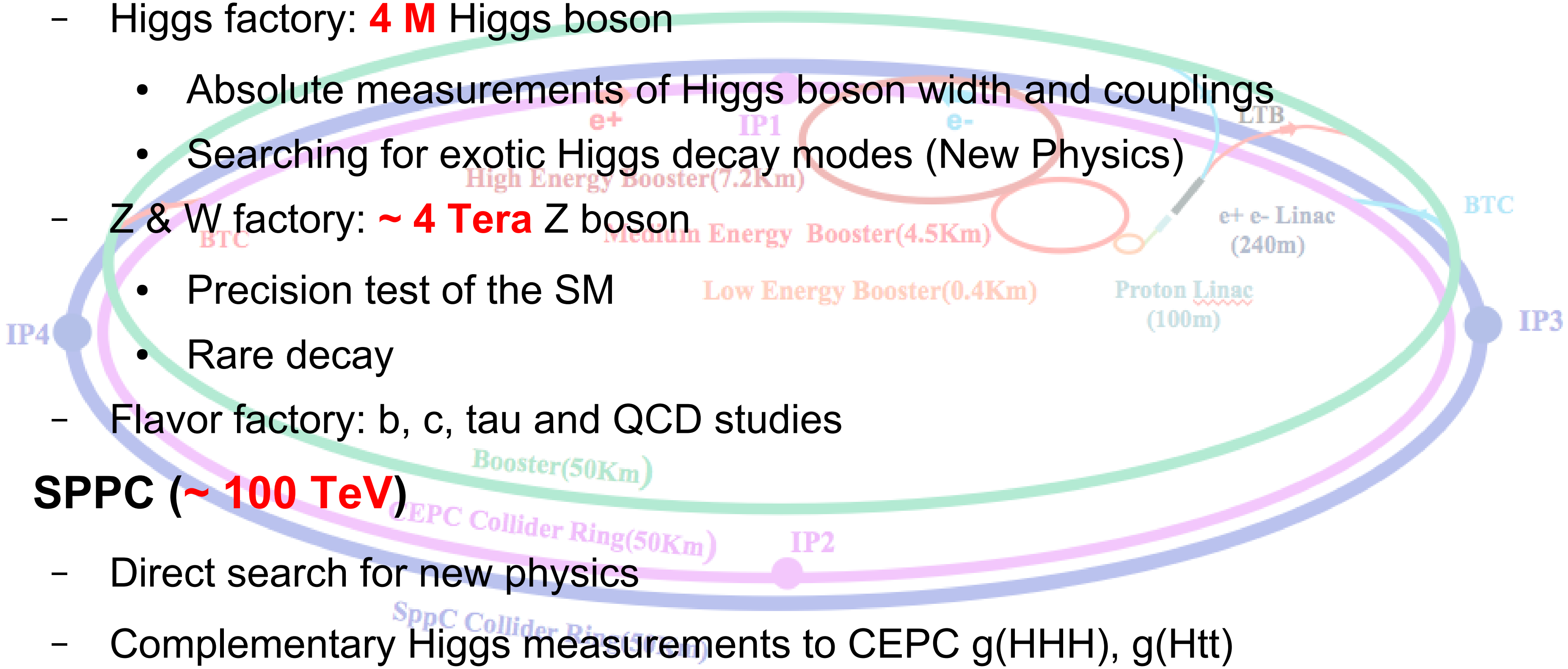
Summary

- 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 (combines jet flavor tagging + charge measurements) is possible and has encouraging performances
  - Compared to LCFIPlus, ParticleNet has significant better performance
    - Improves H->cc accuracy ~ 40%
    - Consistent conclusion on Det. Optimization
  - Higgs exotic/FCNC processes with hadronic final states limited to the BRs of 1E-3 to 1E-4; H→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

Backup

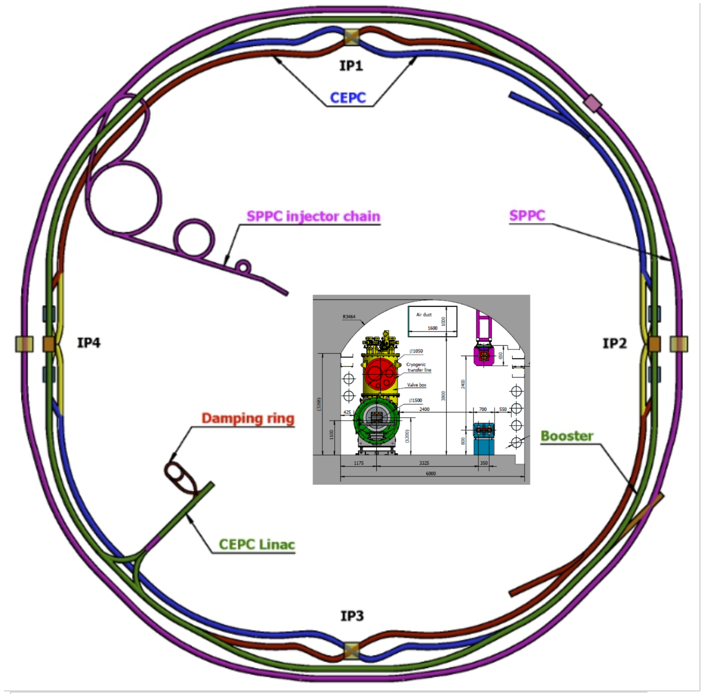
Key parameters of the CEPC-SPPC

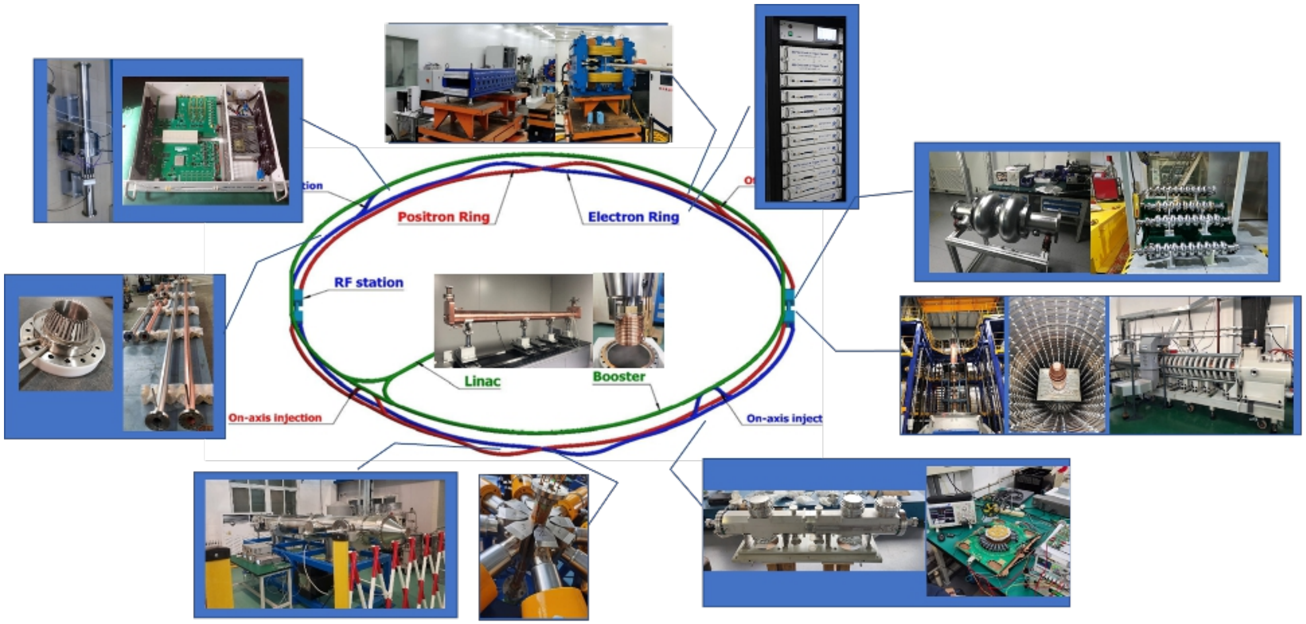
- **Tunnel ~ 100 km**
- **CEPC (90 – 240 GeV)**
  - Higgs factory: **4 M** Higgs boson
    - Absolute measurements of Higgs boson width and couplings
    - Searching for exotic Higgs decay modes (New Physics)
  - Z & W factory: **~ 4 Tera** Z boson
    - Precision test of the SM
    - Rare decay
  - 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...**

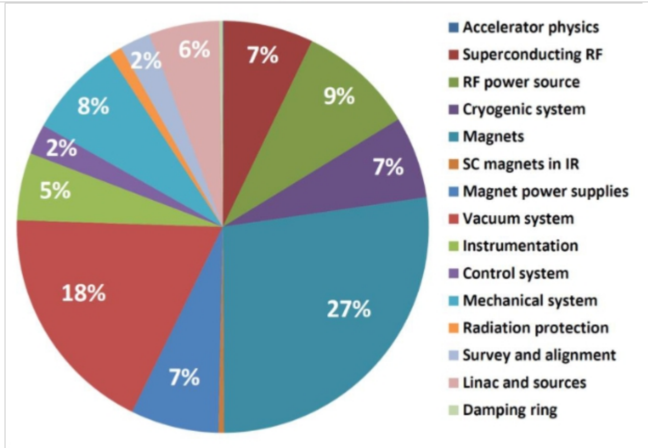


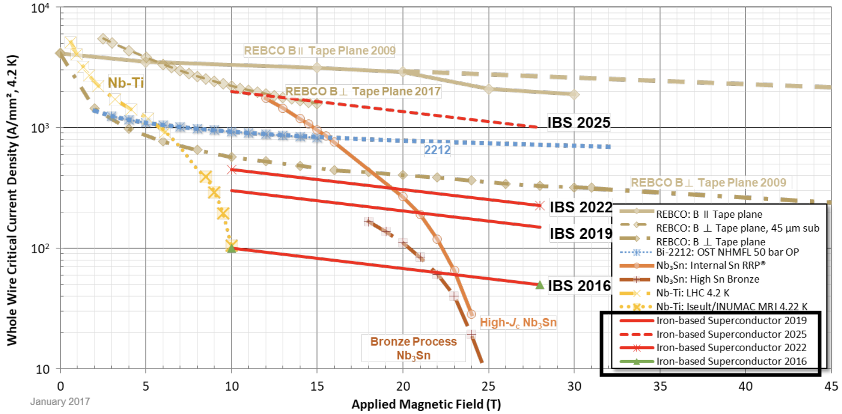
Complementary

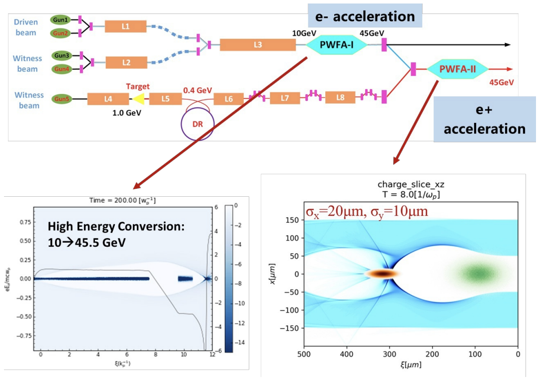
Accelerator at 2023











Platform for key technology R&D

- Magnet R&D
- Mechanical & alignment
- Klystron Test

Beam test hall

- SRF test hall
- Vacuum assembly & coating

Cryogenic System

SRF R&D facility

Magnet R&D

Vacuum R&D

Beam test lab

Alignment

Klystron R&D

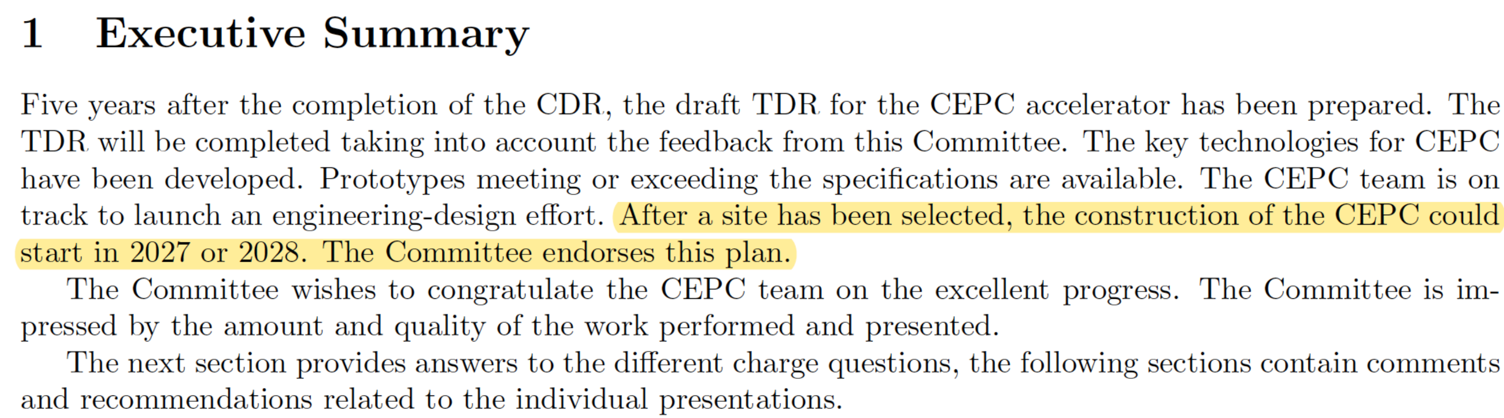
Accelerator key technology R&D platform was established:

- ➢ SRF cavity and module
- ➢ High precision magnet
- ➢ Vacuum assembly & coating
- ➢ High efficiency Klystron
- ➢ Mechanics and alignment
- ➢ Beam test facility

12-16. June. 2023, Hongkong, CEPC Accelerator TDR International Review

38

TDR review: HK June 2023



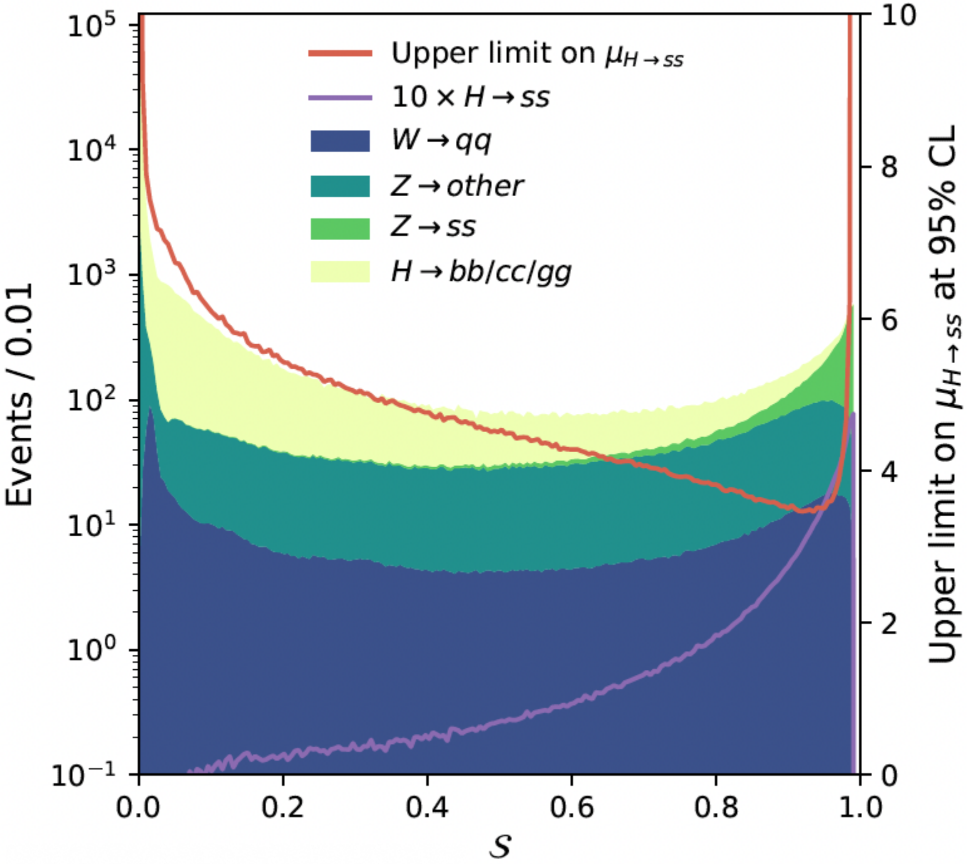
1 Executive Summary

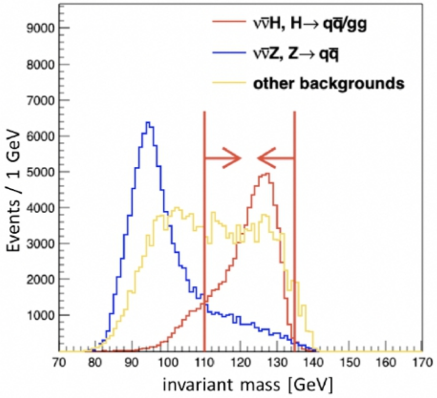
Five years after the completion of the CDR, the draft TDR for the CEPC accelerator has been prepared. The TDR will be completed taking into account the feedback from this Committee. The key technologies for CEPC have been developed. Prototypes meeting or exceeding the specifications are available. The CEPC team is on track to launch an engineering-design effort. After a site has been selected, the construction of the CEPC could start in 2027 or 2028. The Committee endorses this plan.

The Committee wishes to congratulate the CEPC team on the excellent progress. The Committee is impressed by the amount and quality of the work performed and presented.

The next section provides answers to the different charge questions, the following sections contain comments and recommendations related to the individual presentations.

Benchmark analyses using Jet origin ID





H → ss:

be limited to 3*SM using vvH + llH at 20 iab

Potential improvement:

Better event selection;
Including qqH channel;
Including neutral Kaon ID;

1 – 2 *SM @ 20 iab

1E8 Higgs → 5-sigma discovery...

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 ~ 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)

Summary

- We propose CHLOE, using
  - GSHCAL
  - Xbar ECAL + Position/timing layer of
    - **Silicon**
    - MGPRC
  - 2.5 Tracker Scenarios:
    - **Gas Tracker $R_{in/out}$ ~ 25/175 cm, Z ~ 500 cm**
    - Improved 4th: Fwd RHIC
    - Full Silicon with Pid (dE/dx ~ 3%...)
  - 3 VTX Scenarios
    - Rin ~ 10 mm
    - **Vin**
    - Vin Portable

- Anticipated Performance
  - Acceptance: cos(θ)~0.995
  - **BMR ~ 3%**
  - **EM resolution 3%/sqrt(E), const. term < 1%**
  - Timing resolution ~ o(50) ps
  - dP/P ~ 0.1% in the barrel
  - Pid: **eff/purity > 96%** for charged Kaon at hadronic Z event
  - Jet Flavor Tagging:
    - Tr(Mig): from ~2.4 to ~2.7
    - Enhance the g(Hcc) and |Vcb| measurements by 60% - 100%...
  - Fulfill the requirements of not only Higgs, but also Flavor & New Physics

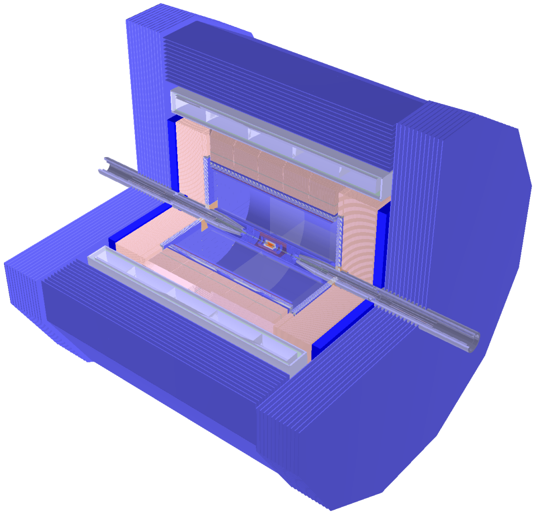
Summary

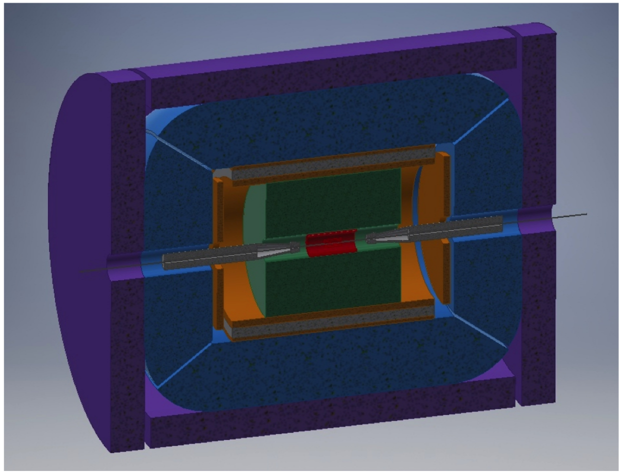
- Critical Challenges
  - Boundary conditions to determine sub-detector technology & configuration...
    - Impact of beam background on sub detectors, especially gaseous one
    - MDI design, installation & integration
  - Vin
    - Power & Signal
    - Integration - Hom heat & radiation bkgrd, coating...
    - 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

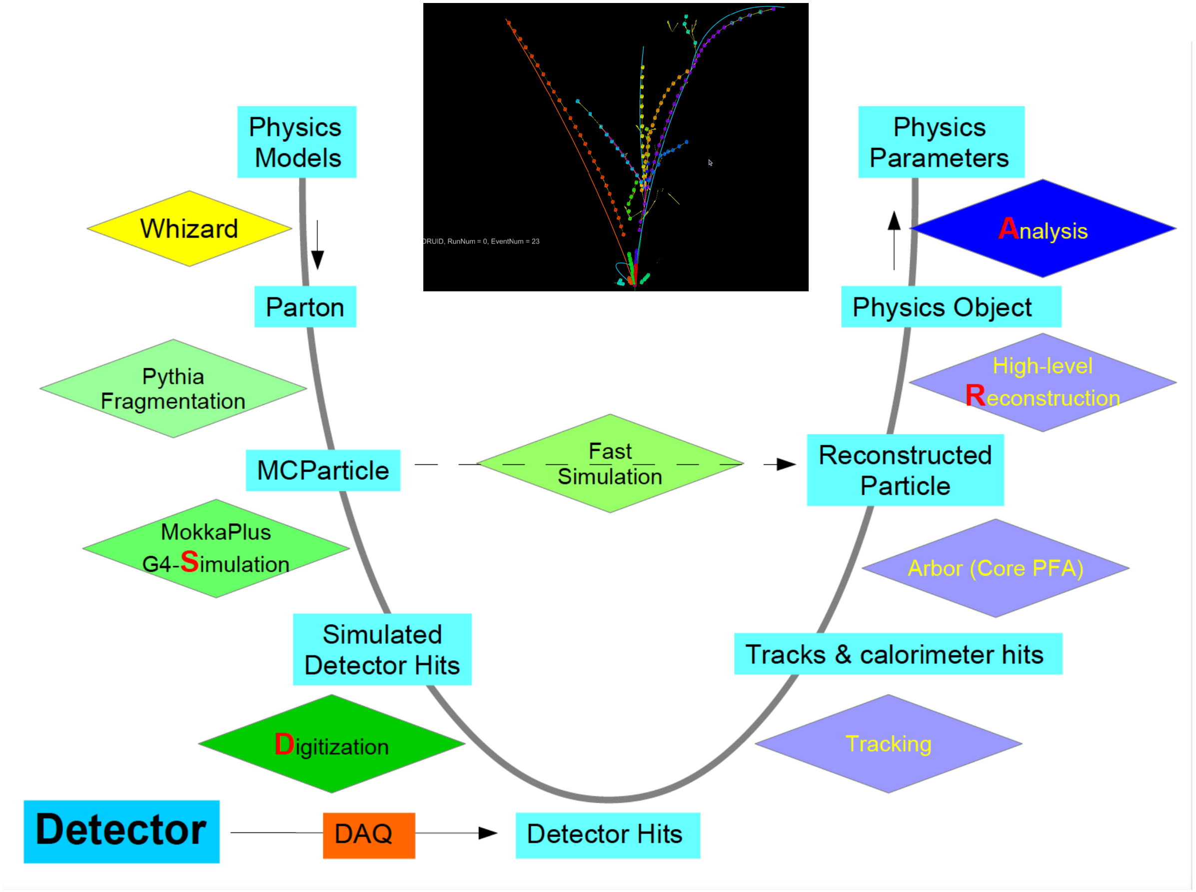
Summary

- 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







Physics Models

Whizard

Parton

Pythia Fragmentation

MCParticle

MokkaPlus G4-Simulation

Simulated Detector Hits

Digitization

Detector

DAQ

Detector Hits

Fast Simulation

Tracking

Tracks & calorimeter hits

Arbor (Core PFA)

Reconstructed Particle

High-level Reconstruction

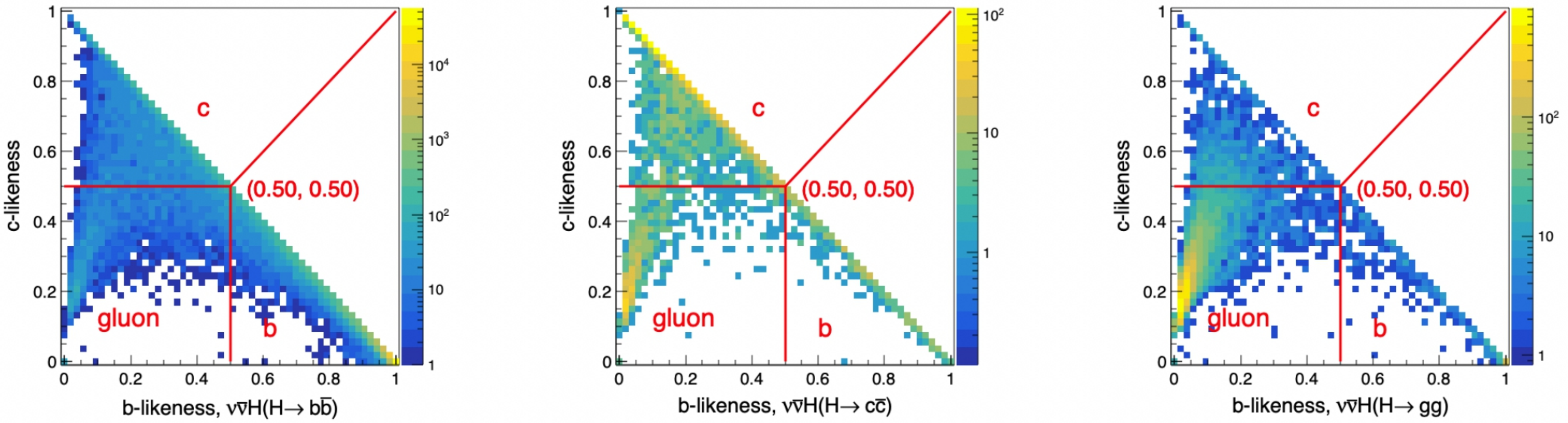
Physics Object

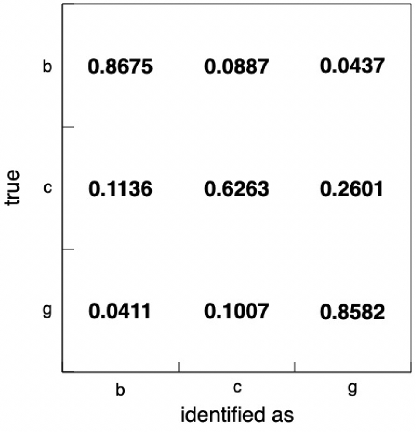
Analysis

Physics Parameters

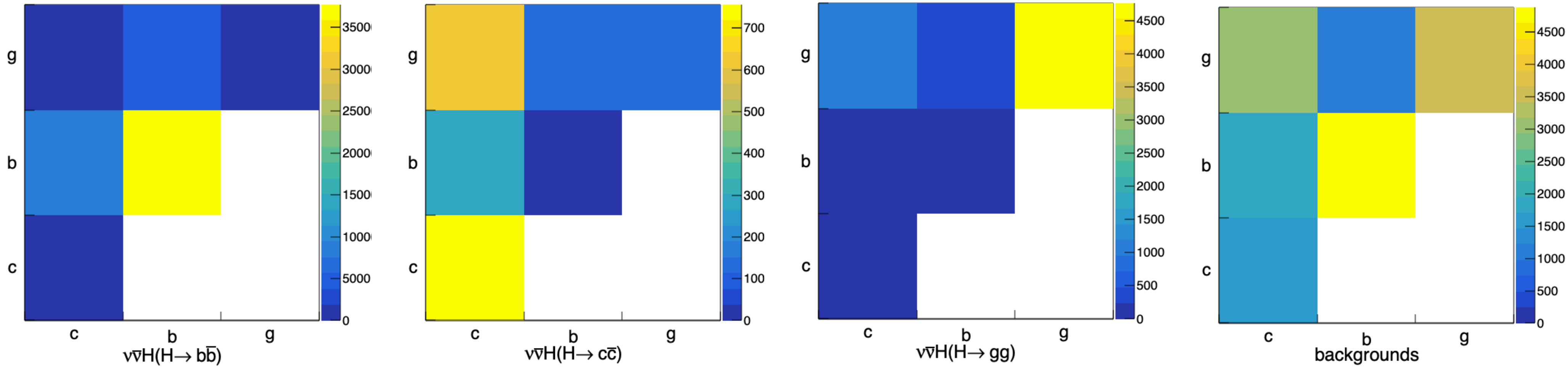
Full simulation reconstruction Chain with Arbor, iterating/validation with hardware studies

Flavor tagging @ vvH

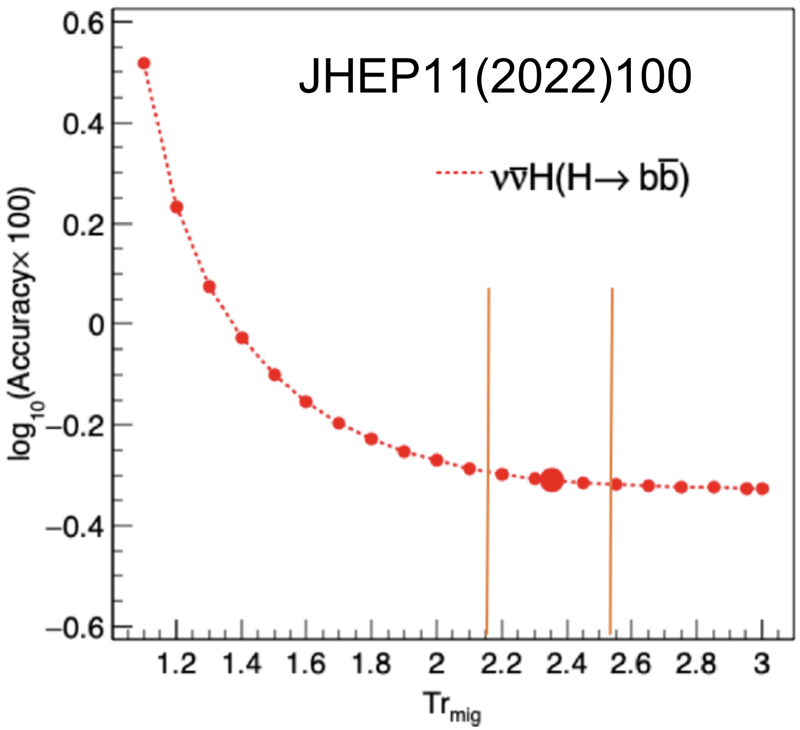


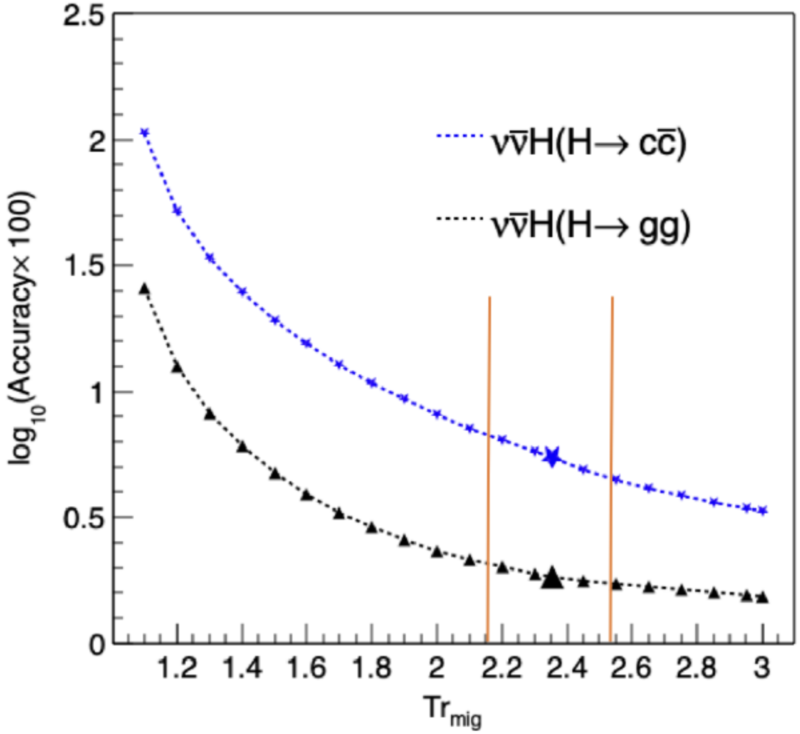


true \ identified as	b	c	g
b	0.8675	0.0887	0.0437
c	0.1136	0.6263	0.2601
g	0.0411	0.1007	0.8582



Vertex





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

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