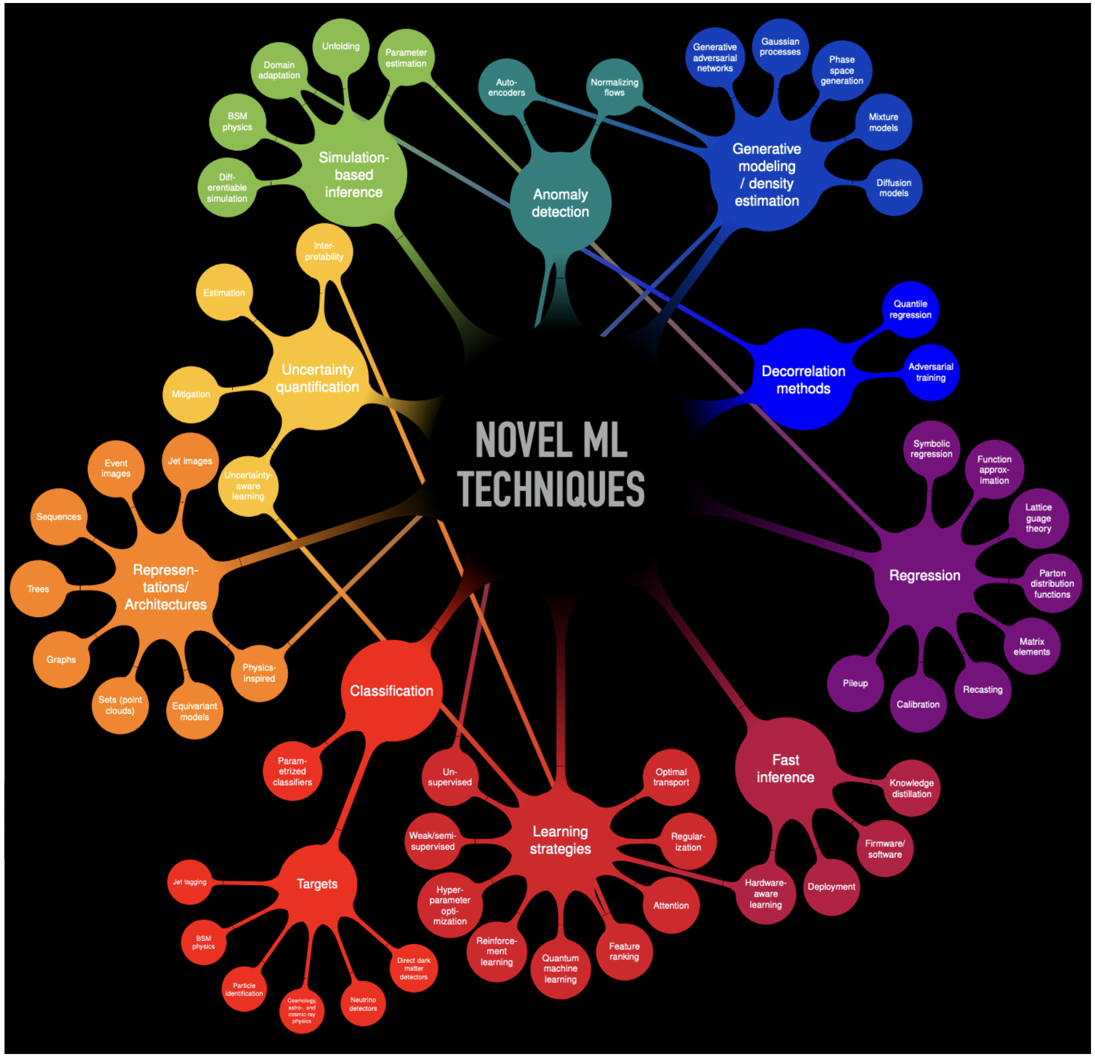
Recent highlight on usage of Machine Learning at Higgs factory

Manqi Ruan

AI & HEP

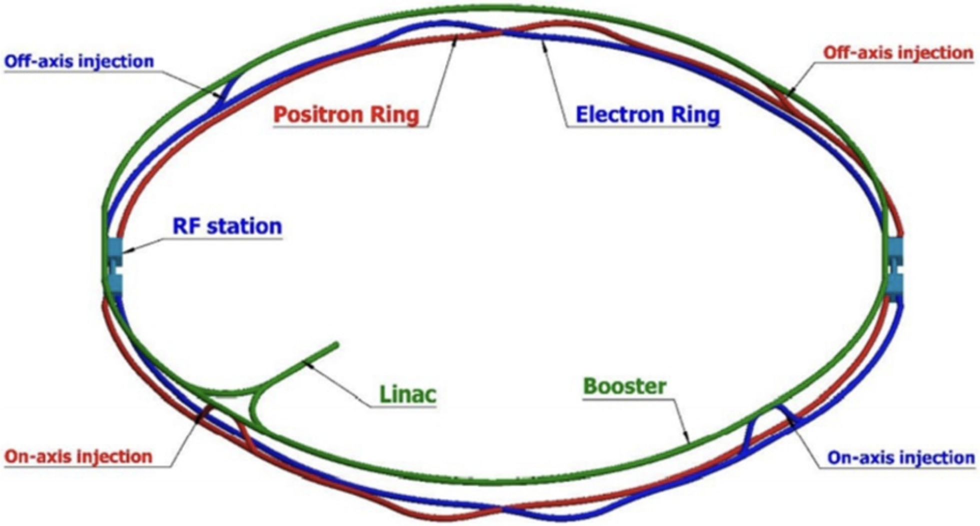


Plot from Javier Mauricio Duarte's talk at ICHEP 2024

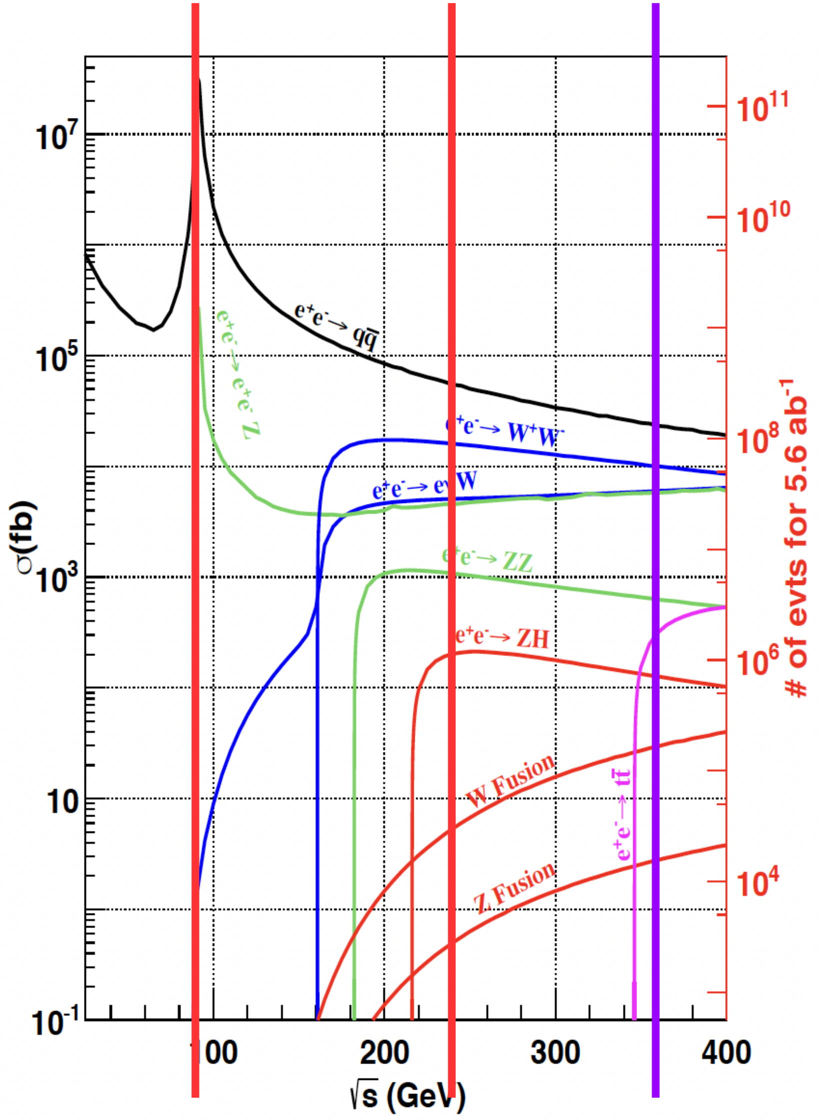
- HEP: data intensive + clear, meaningful & interpretable processing
  - Pioneering for neural network application, i.e., in tracking in 1980s
- An irresistible trend:
  - 17/48 parallel talks of Computing & Data handling session at ICHEP 2024 are relevant to AI: 11 machine learning, 3 deep learning, 3 neural network
  - Many domestic discussions & efforts
- CEPC is actively implementing AI to its data processing:
  - Trigger + DAQ...
  - Simulation: Fast sim.
  - Reconstruction: PFA, Jet Origin id, etc
  - Analysis

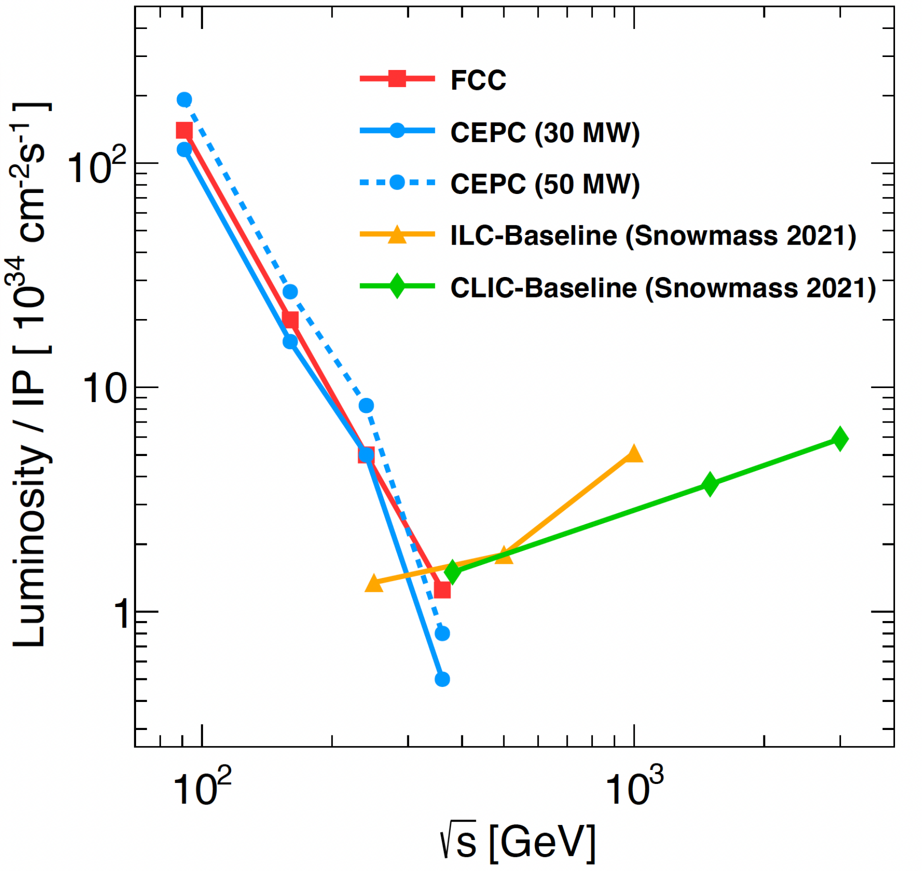
Outline

- *Disclaimer: this talk is not a review talk of all recent progress at Higgs factories, but only focus on a few recent progress that I've been involved*
- CEPC Physics & requirements
- Jet origin identification
- 1-1 correspondence reconstruction
- Progress with LLM and Color Singlet identification
- Discussion



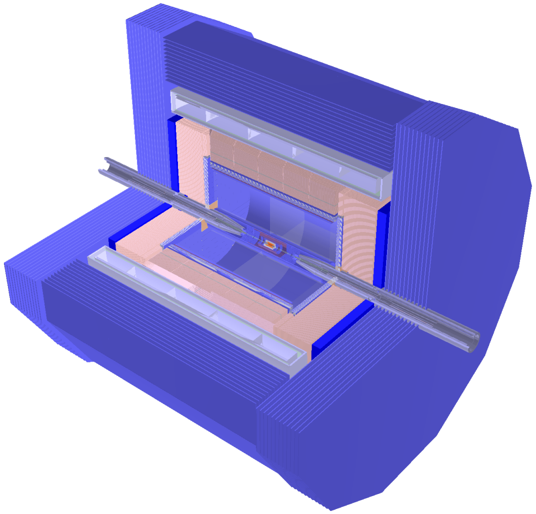
Yields ~ Xsec * Lumi * Time

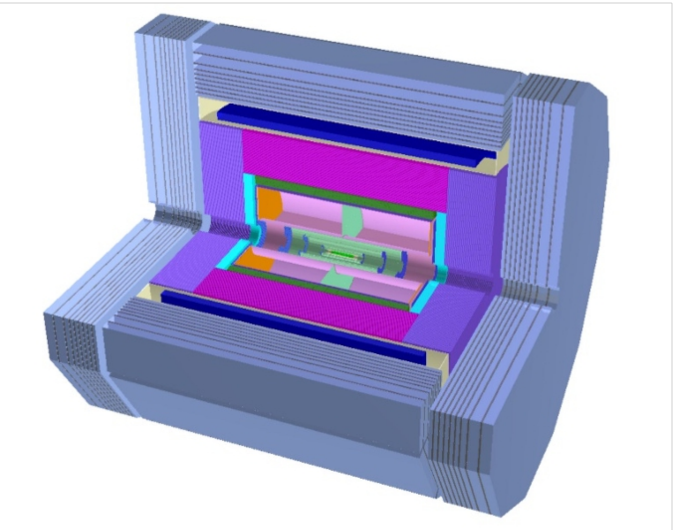


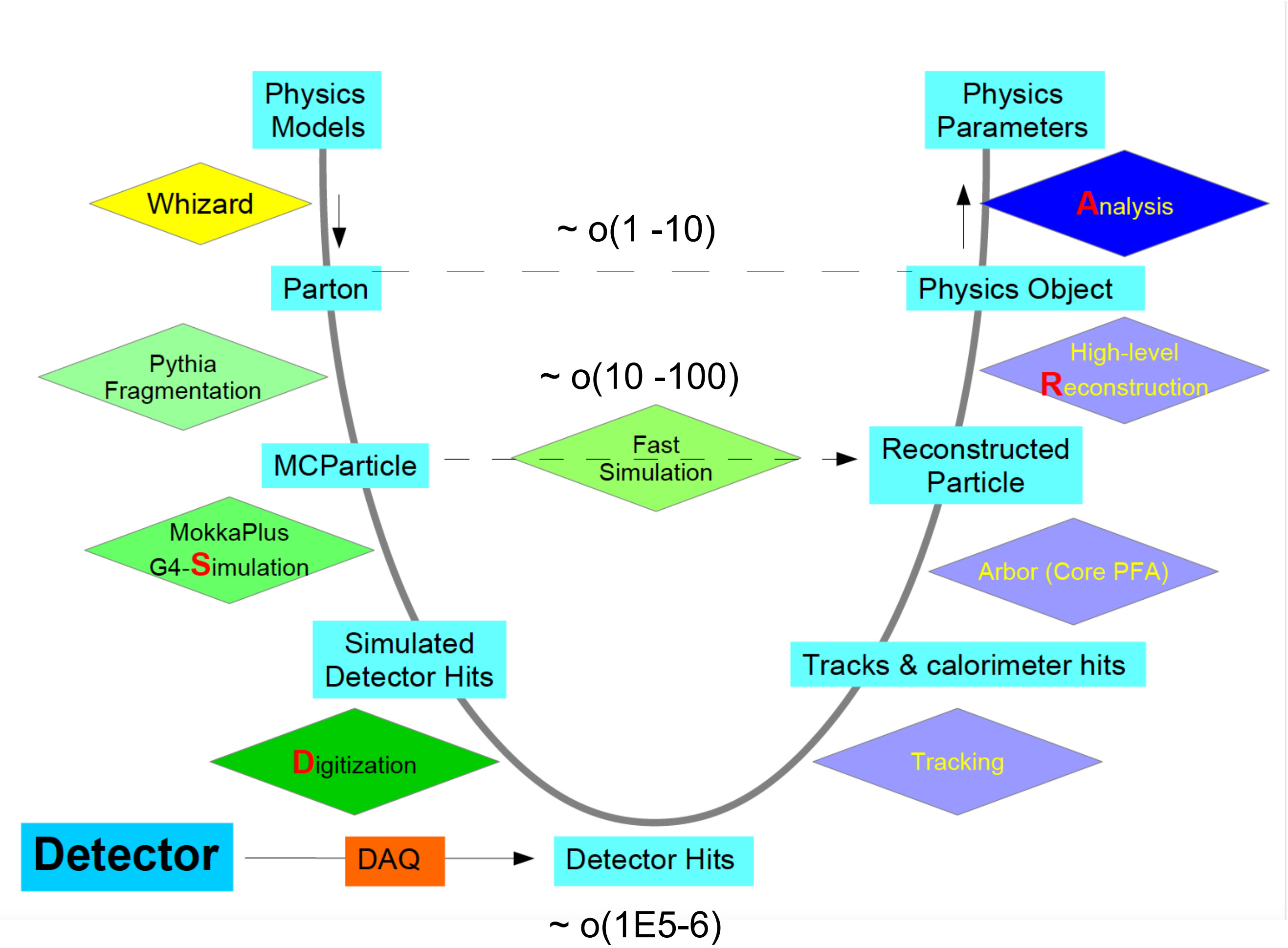


- 4 Million Higgs (10 years)
- ~ 1 Giga W (1 year) + 4 Tera Z (2 years)
- Upgradable: Top factory (500 k ttbar)

CEPC Detector & Reconstruction

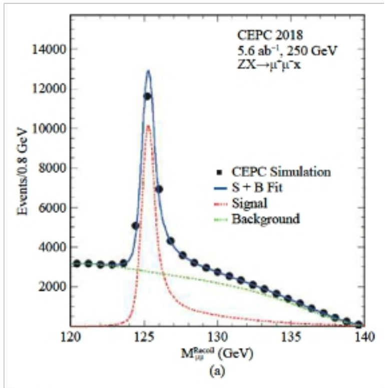




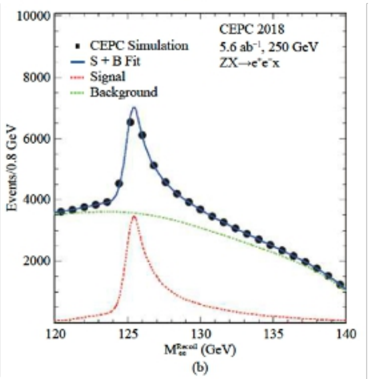


Full simulation reconstruction Chain with **Arbor, etc**

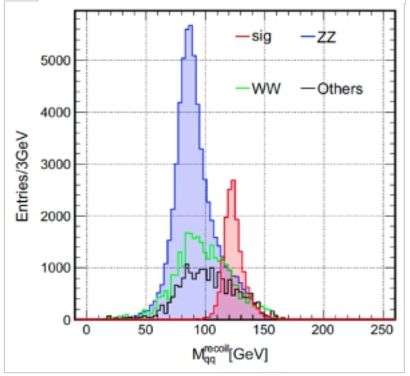
CEPC Physics study



(a)



(b)



Chinese Physics C Vol. 43, No. 4 (2019) 043002

Precision Higgs physics at the CEPC*

Fenfen An(安芬芬)[4,23] Yu Bai(白羽)[9] Chunhui Chen(陈春晖)[23] Xin Chen(陈新)[5] Zhenxing Chen(陈振兴)[3]
Joao Guimaraes da Costa[4] Zhenwei Cui(崔振崴)[3] Yaquan Fang(方亚泉)[4,6,34;1)] Chengdong Fu(付成栋)[4]
Jun Gao(高俊)[10] Yanyan Gao(高艳彦)[22] Yuanning Gao(高原宁)[3] Shaofeng Ge(葛韶锋)[15,29]
Jiayin Gu(顾嘉荫)[13;2)] Fangyi Guo(郭方毅)[1,4] Jun Guo(郭军)[10] Tao Han(韩涛)[5,31] Shuang Han(韩爽)[4]
Hongjian He(何红建)[11,10] Xianke He(何显柯)[10] Xiaogang He(何小刚)[11,10,20] Jifeng Hu(胡继峰)[10]
Shih-Chieh Hsu(徐士杰)[32] Shan Jin(金山)[8] Maoqiang Jing(荆茂强)[4,7] Susmita Jyotishmati[33] Ryuta Kiuchi[4]
Chia-Ming Kuo(郭家铭)[21] Peizhu Lai(赖培筑)[21] Boyang Li(李博扬)[5] Congqiao Li(李聪乔)[3] Gang Li(李刚)[4,34;3)]
Haifeng Li(李海峰)[12] Liang Li(李亮)[10] Shu Li(李数)[11,10] Tong Li(李通)[12] Qiang Li(李强)[3] Hao Liang(梁浩)[4,6]
Zhijun Liang(梁志均)[4] Libo Liao(廖立波)[4] Bo Liu(刘波)[4,23] Jianbei Liu(刘建北)[1] Tao Liu(刘涛)[14]
Zhen Liu(刘真)[26,30;4)] Xinchou Lou(娄辛丑)[4,6,33,34] Lianliang Ma(马连良)[12] Bruce Mellado[17,18] Xin Mo(莫欣)[4]
Mila Pandurovic[16] Jianming Qian(钱剑明)[24;5)] Zhuoni Qian(钱卓妮)[19] Nikolaos Rompotis[22]
Manqi Ruan(阮曼奇)[4;6)] Alex Schuy[32] Lianyou Shan(单连友)[4] Jingyuan Shi(史静远)[9] Xin Shi(史欣)[4]
Shufang Su(苏淑芳)[25] Dayong Wang(王大勇)[3] Jin Wang(王锦)[4] Liantao Wang(王连涛)[27;7)]
Yifang Wang(王贻芳)[4,6] Yuqian Wei(魏彧骞)[4] Yue Xu(许悦)[5] Haijun Yang(杨海军)[10,11] Ying Yang(杨迎)[4]
Weiming Yao(姚为民)[28] Dan Yu(于丹)[4] Kaili Zhang(张凯栗)[4,6;8)] Zhaoru Zhang(张照茹)[4]
Mingrui Zhao(赵明锐)[2] Xianghu Zhao(赵祥虎)[4] Ning Zhou(周宁)[10]

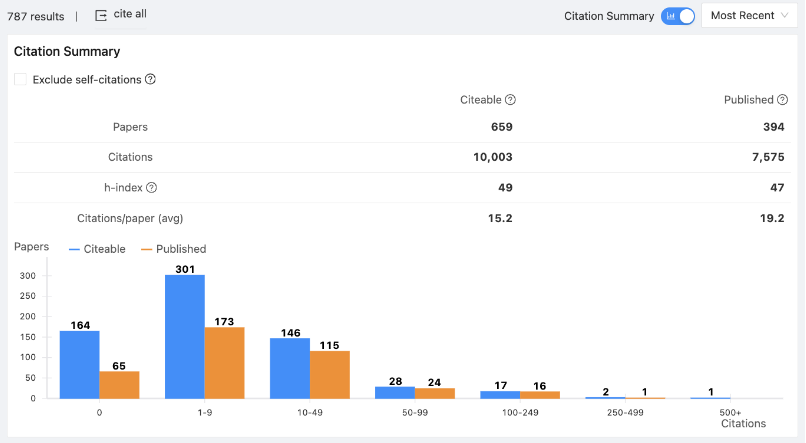
[1] Department of Modern Physics, University of Science and Technology of China, Anhui 230026, China
[2] China Institute of Atomic Energy, Beijing 102413, China
[3] School of Physics, Peking University, Beijing 100871, China
[4] Institute of High Energy Physics, Beijing 100049, China
[5] Department of Engineering Physics, Physics Department, Tsinghua University, Beijing 100084, China
[6] University of Chinese Academy of Science (UCAS), Beijing 100049, China
[7] School of Nuclear Science and Technology, University of South China, Hengyang 421001, China
[8] Department of Physics, Nanjing University, Nanjing 210093, China
[9] Department of Physics, Southeast University, Nanjing 210096, China
[10] School of Physics and Astronomy, Shanghai Jiao Tong University, KLPPAC-MoE, SKLPPC, Shanghai 200240, China
[11] Tsung-Dao Lee Institute, Shanghai 200240, China
[12] Institute of Frontier and Interdisciplinary Science and Key Laboratory of Particle Physics and Particle Irradiation (MOE), Shandong University, Qingdao 266237, China
[13] PRISMA Cluster of Excellence & Mainz Institute of Theoretical Physics, Johannes Gutenberg-Universität Mainz, Mainz 55128, Germany
[14] Department of Physics, Hong Kong University of Science and Technology, Hong Kong
[15] Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[16] Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade 11000, Serbia
[17] School of Physics and Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg 2050, South Africa

---

Received 9 November 2018, Revised 21 January 2019, Published online 4 March 2019

* Supported by the National Key Program for S&T Research and Development (2016YFA0400400); CAS Center for Excellence in Particle Physics; Yifang Wang's Science Studio of the Ten Thousand Talents Project; the CAS/SAFEA International Partnership Program for Creative Research Teams (H751S018S5); IHEP Innovation Grant (Y4545170Y2); Key Research Program of Frontier Sciences, CAS (XQYZDY-SSW-SLH002); Chinese Academy of Science Special Grant for Large Scientific Project (113111KYSB20170005); the National Natural Science Foundation of China(11675202); the Hundred Talent Programs of Chinese Academy of Science (Y3515540U1); the National 1000 Talents Program of China; Fermi Research Alliance, LLC (DE-AC02-07CH11359); the NSF(PHY1620074) ; by the Maryland Center for Fundamental Physics (MCFP); Tsinghua University Initiative Scientific Research Program; and the Beijing Municipal Science and Technology Commission project(Z181100004218003)

1) E-mail: fangyq@ihep.ac.cn
2) E-mail: jiagu@uni-mainz.de
3) E-mail: li.gang@mail.ihep.ac.cn
4) E-mail: zliuphys@umd.edu
5) E-mail: qianj@umich.edu
6) E-mail: manqi.ruan@ihep.ac.cn
7) E-mail: liantaow@uchicago.edu
8) E-mail: zhangkl@ihep.ac.cn



Citation Summary

☐ Exclude self-citations

	Citeable	Published
Papers	659	394
Citations	10,003	7,575
h-index	49	47
Citations/paper (avg)	15.2	19.2

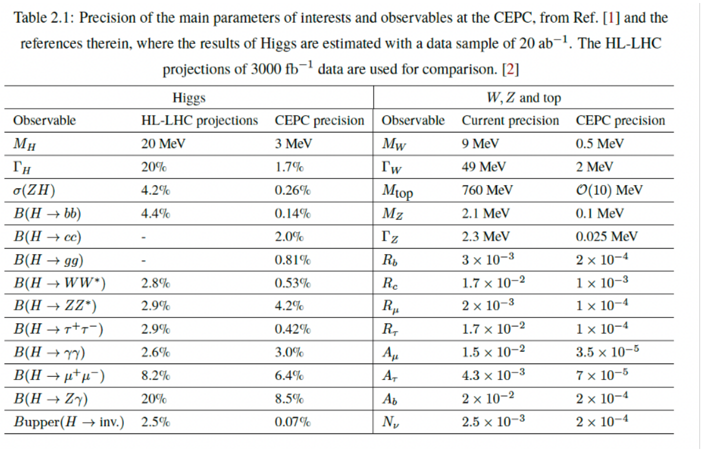


Table 2.1: Precision of the main parameters of interests and observables at the CEPC, from Ref. [1] and the references therein, where the results of Higgs are estimated with a data sample of 20 ab^{-1}. The HL-LHC projections of 3000 fb^{-1} data are used for comparison. [2]

Higgs			W, Z and top		
Observable	HL-LHC projections	CEPC precision	Observable	Current precision	CEPC precision
M_H	20 MeV	3 MeV	M_W	9 MeV	0.5 MeV
Γ_H	20%	1.7%	Γ_W	49 MeV	2 MeV
$\sigma(ZH)$	4.2%	0.26%	M_{top}	760 MeV	$\mathcal{O}(10)$ MeV
$B(H \to bb)$	4.4%	0.14%	M_Z	2.1 MeV	0.1 MeV
$B(H \to cc)$	-	2.0%	Γ_Z	2.3 MeV	0.025 MeV
$B(H \to gg)$	-	0.81%	R_b	3×10^{-3}	2×10^{-4}
$B(H \to WW^*)$	2.8%	0.53%	R_c	1.7×10^{-2}	1×10^{-3}
$B(H \to ZZ^*)$	2.9%	4.2%	R_μ	2×10^{-3}	1×10^{-4}
$B(H \to \tau^+\tau^-)$	2.9%	0.42%	R_τ	1.7×10^{-2}	1×10^{-4}
$B(H \to \gamma\gamma)$	2.6%	3.0%	A_μ	1.5×10^{-2}	3.5×10^{-5}
$B(H \to \mu^+\mu^-)$	8.2%	6.4%	A_τ	4.3×10^{-3}	7×10^{-5}
$B(H \to Z\gamma)$	20%	8.5%	A_b	2×10^{-2}	2×10^{-4}
Bupper$(H \to$ inv.)	2.5%	0.07%	N_ν	2.5×10^{-3}	2×10^{-4}

Scientific Significance quantified by **CEPC physics** studies, via full simulation/phenomenology studies:

- Higgs: Precisions exceed HL-LHC ~ 1 order of magnitude.
- EW: Precision improved from current limit by 1-2 orders.
- Flavor Physics, sensitive to NP of 10 TeV or even higher.
- Sensitive to varies of NP signal.
- …

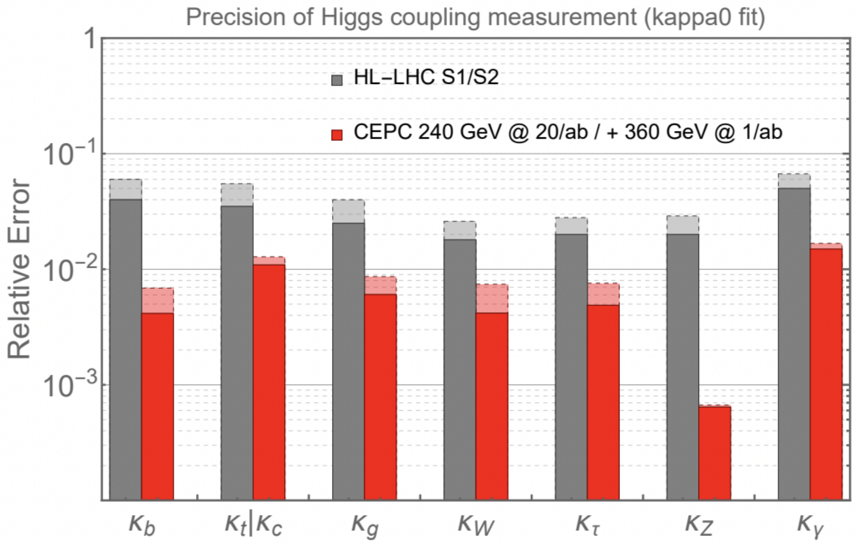
White papers +

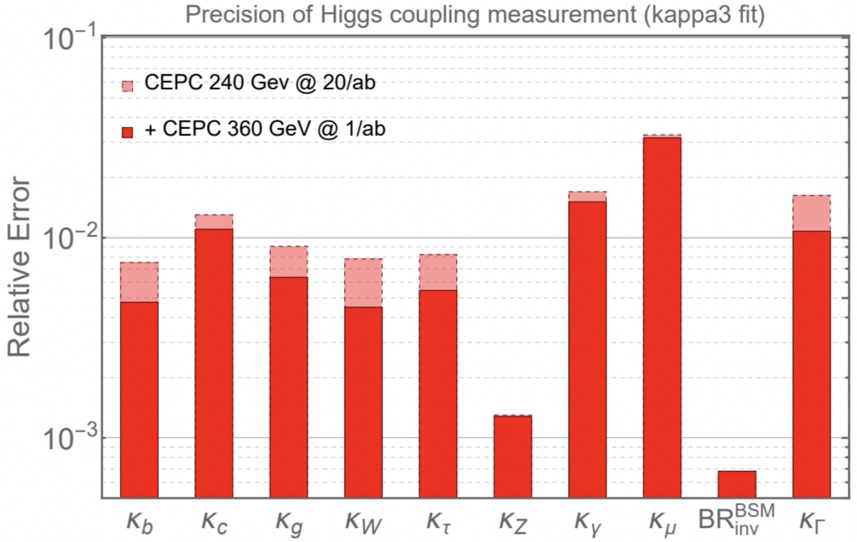
~300 Journal/AxXiv citables

18/11/2024

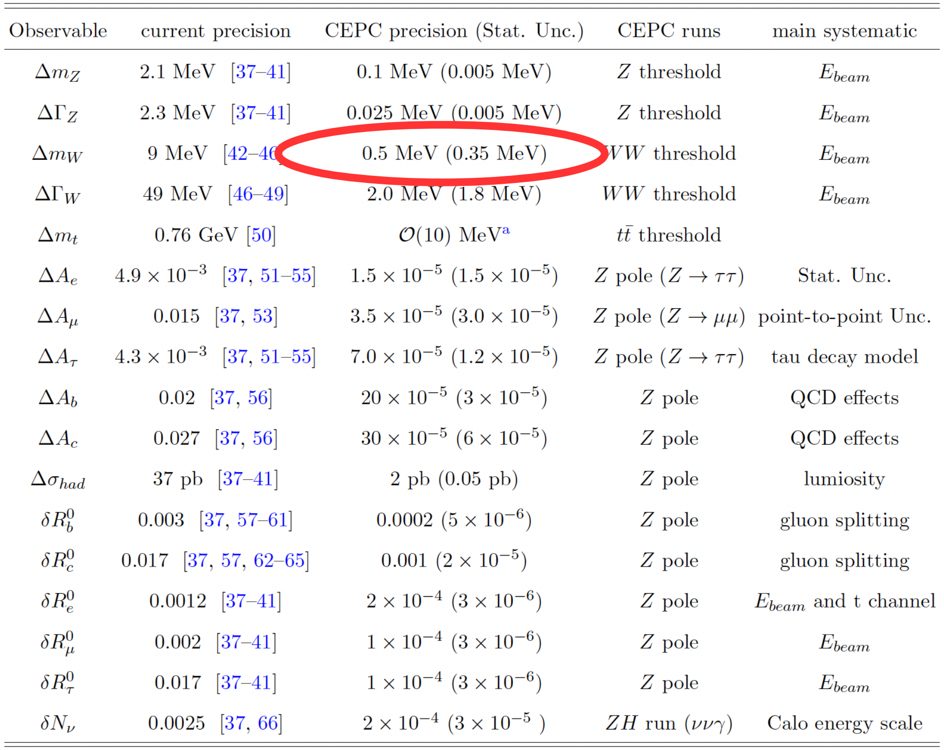
Higgs & Snowmass White Paper

	$240\,\text{GeV},\ 20\ \text{ab}^{-1}$		$360\,\text{GeV},\ 1\ \text{ab}^{-1}$		
	ZH	**vvH**	ZH	**vvH**	**eeH**
inclusive	**0.26%**		**1.40%**	\	\
$H\to bb$	**0.14%**	**1.59%**	**0.90%**	**1.10%**	**4.30%**
$H\to cc$	**2.02%**		**8.80%**	**16%**	**20%**
$H\to gg$	**0.81%**		**3.40%**	**4.50%**	**12%**
$H\to WW$	**0.53%**		**2.80%**	**4.40%**	**6.50%**
$H\to ZZ$	**4.17%**		**20%**	**21%**	
$H\to\tau\tau$	**0.42%**		**2.10%**	**4.20%**	**7.50%**
$H\to\gamma\gamma$	**3.02%**		**11%**	**16%**	
$H\to\mu\mu$	**6.36%**		**41%**	**57%**	
$H\to Z\gamma$	**8.50%**		**35%**		
$\text{Br}_{upper}(H\to inv.)$	**0.07%**				
Γ_H	**1.65%**		**1.10%**		

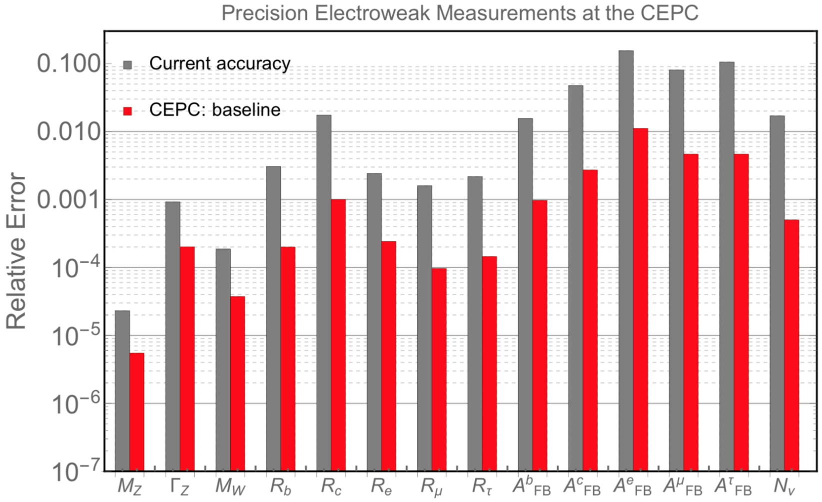


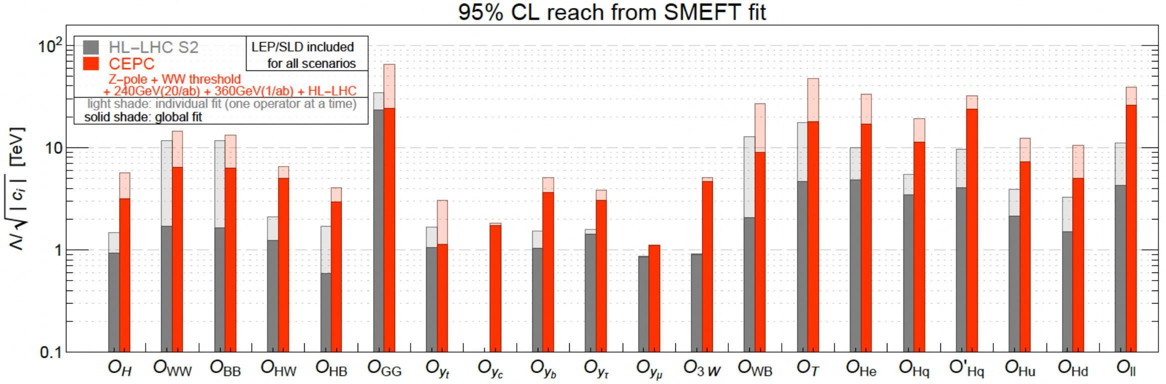


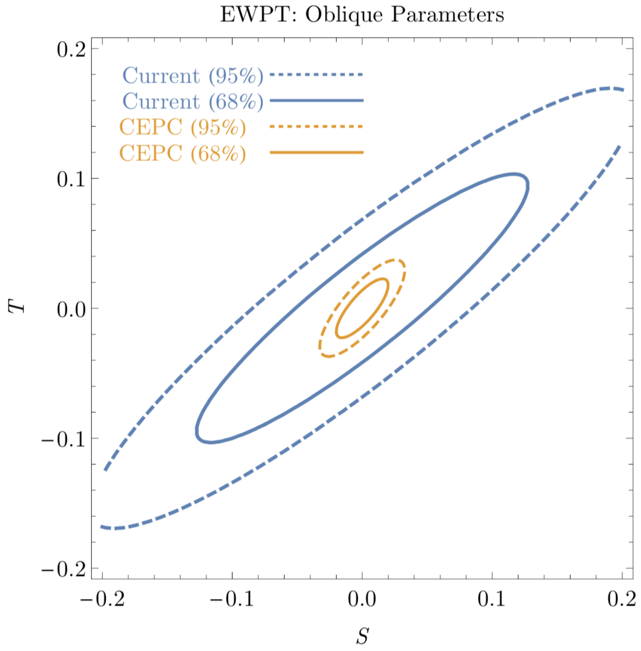
EW measurements & SMEFT



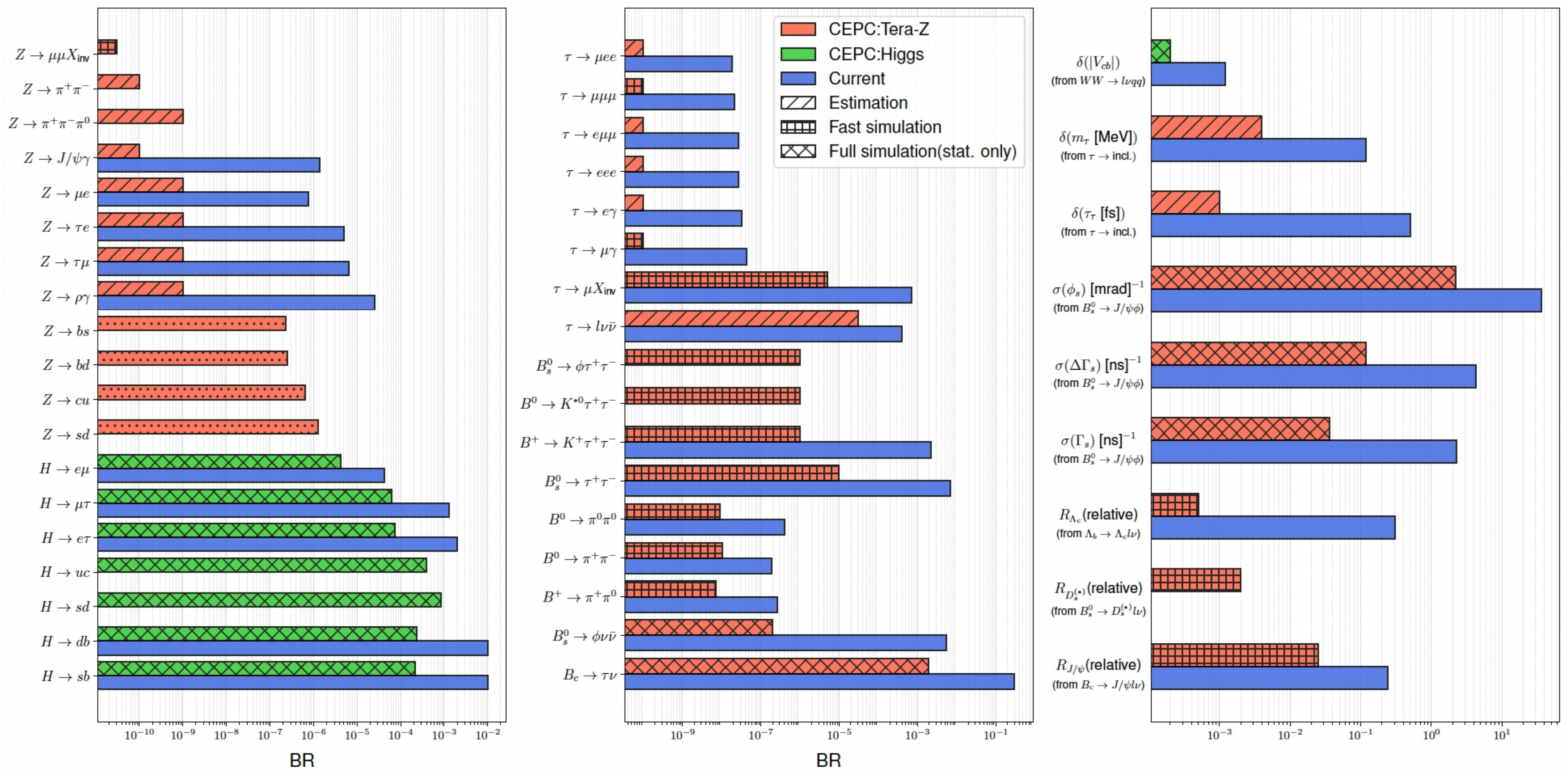
Observable	current precision	CEPC precision (Stat. Unc.)	CEPC runs	main systematic
Δm_Z	2.1 MeV [37–41]	0.1 MeV (0.005 MeV)	Z threshold	E_{beam}
$\Delta\Gamma_Z$	2.3 MeV [37–41]	0.025 MeV (0.005 MeV)	Z threshold	E_{beam}
Δm_W	9 MeV [42–46	0.5 MeV (0.35 MeV)	WW threshold	E_{beam}
$\Delta\Gamma_W$	49 MeV [46–49]	2.0 MeV (1.8 MeV)	WW threshold	E_{beam}
Δm_t	0.76 GeV [50]	$\mathcal{O}(10)$ MeV[a]	$t\bar{t}$ threshold	
ΔA_e	4.9×10^{-3} [37, 51–55]	1.5×10^{-5} (1.5×10^{-5})	Z pole ($Z \to \tau\tau$)	Stat. Unc.
ΔA_μ	0.015 [37, 53]	3.5×10^{-5} (3.0×10^{-5})	Z pole ($Z \to \mu\mu$)	point-to-point Unc.
ΔA_τ	4.3×10^{-3} [37, 51–55]	7.0×10^{-5} (1.2×10^{-5})	Z pole ($Z \to \tau\tau$)	tau decay model
ΔA_b	0.02 [37, 56]	20×10^{-5} (3×10^{-5})	Z pole	QCD effects
ΔA_c	0.027 [37, 56]	30×10^{-5} (6×10^{-5})	Z pole	QCD effects
$\Delta\sigma_{had}$	37 pb [37–41]	2 pb (0.05 pb)	Z pole	lumiosity
δR_b^0	0.003 [37, 57–61]	0.0002 (5×10^{-6})	Z pole	gluon splitting
δR_c^0	0.017 [37, 57, 62–65]	0.001 (2×10^{-5})	Z pole	gluon splitting
δR_e^0	0.0012 [37–41]	2×10^{-4} (3×10^{-6})	Z pole	E_{beam} and t channel
δR_μ^0	0.002 [37–41]	1×10^{-4} (3×10^{-6})	Z pole	E_{beam}
δR_τ^0	0.017 [37–41]	1×10^{-4} (3×10^{-6})	Z pole	E_{beam}
δN_ν	0.0025 [37, 66]	2×10^{-4} (3×10^{-5})	ZH run ($\nu\nu\gamma$)	Calo energy scale







Flavor Physics

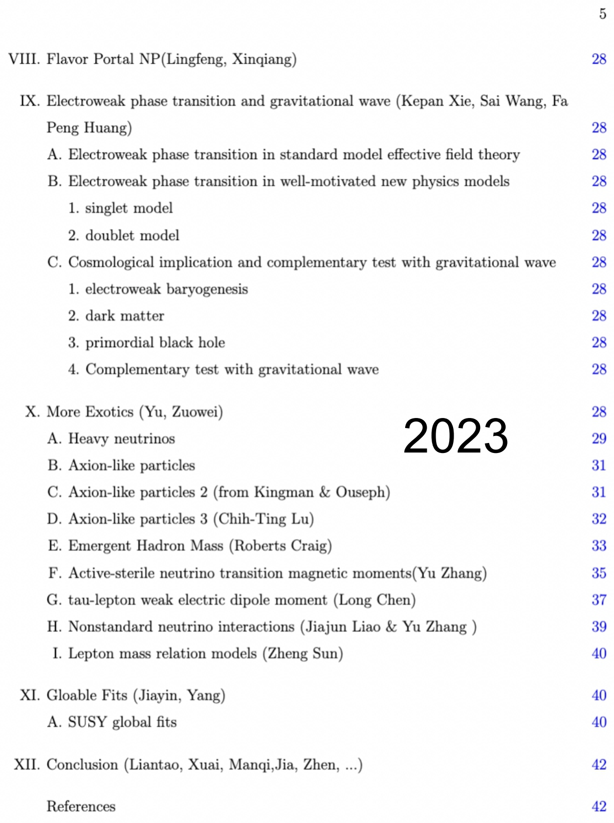


See the non-seen: i.e, Bc→tauv, Bs→Phivv

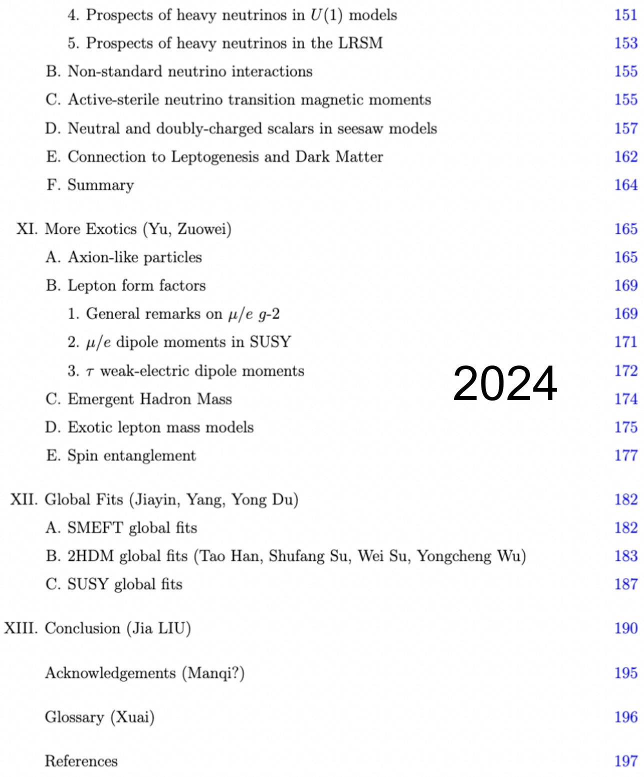
Orders of magnitudes improvements (1 – 2.5 orders...).

Access New Physics with energy scale of 10 TeV, or even above

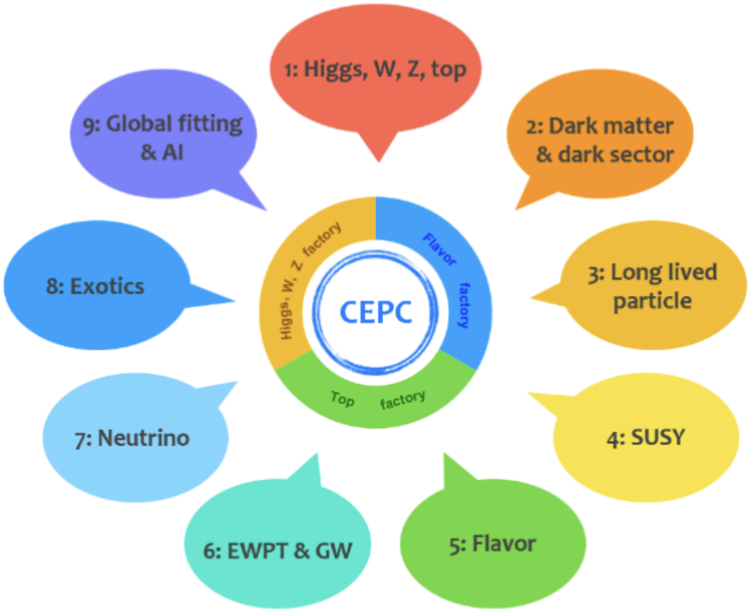
New Physics white paper



2023



2024

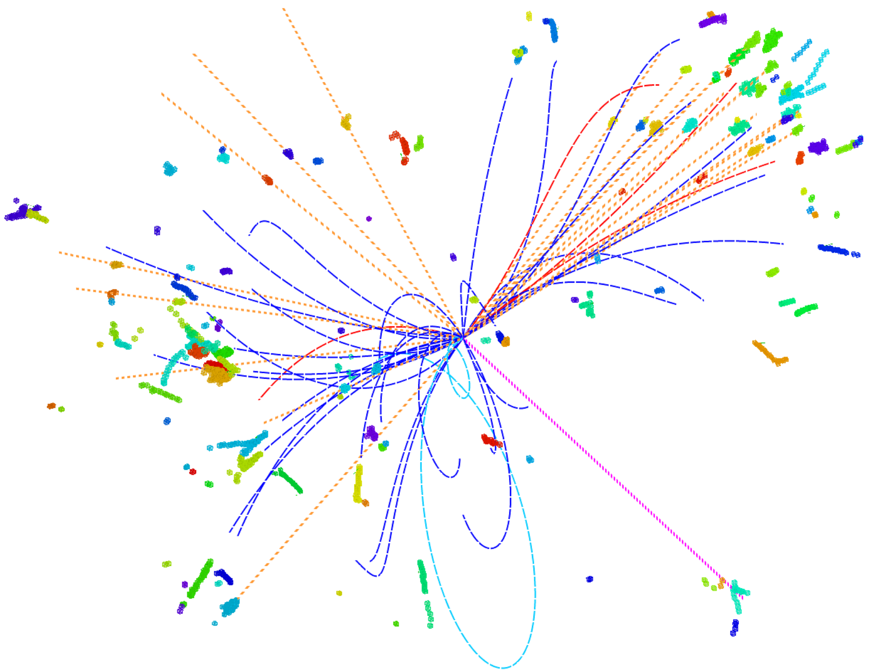


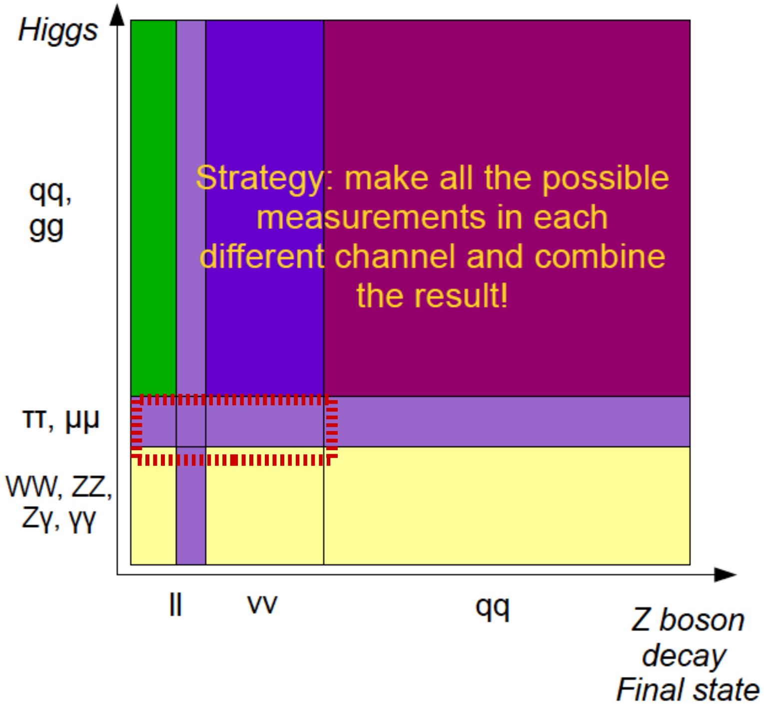
- Credit: hanhua Cui, Yu Gao, Xuai Zhuang

Contents extends from 40 pages → 200 pages...

Performance requirements

- To well reconstruct all Physics Object, especially Jets
    - Z & W: ~ 70% goes to a pair of jets
    - Higgs: ~90% final state with jets (ZH events)
    - Top: t → W + b

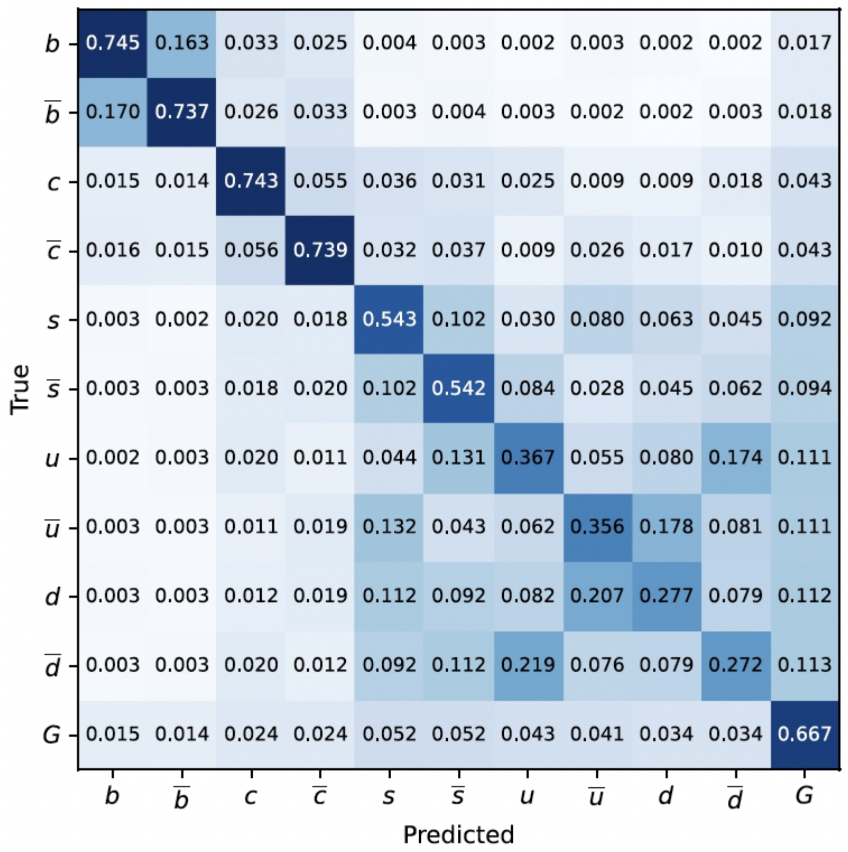




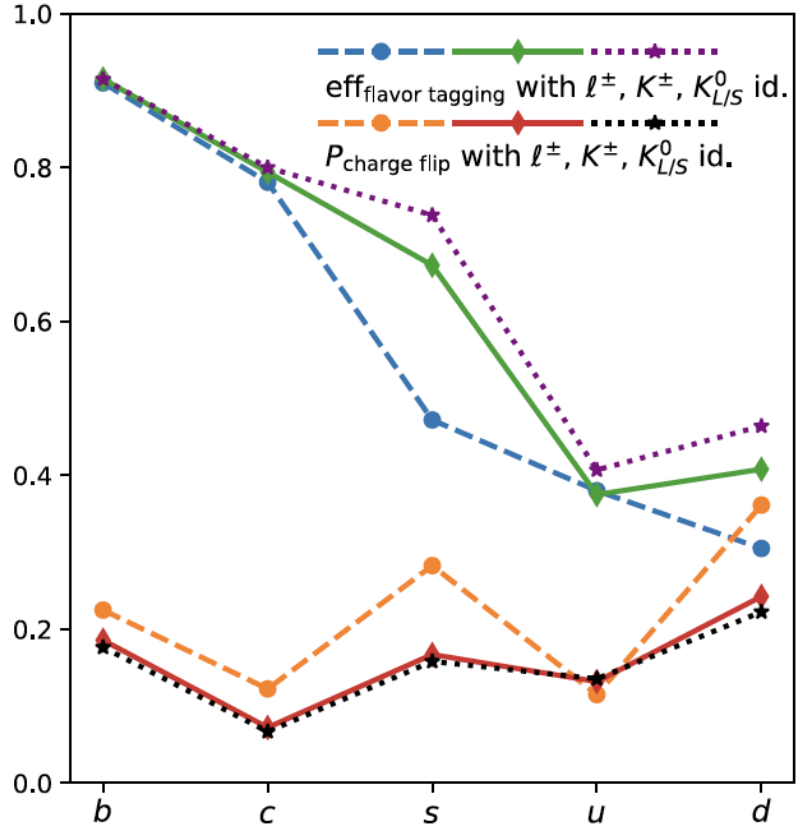
Strategy: make all the possible measurements in each different channel and combine the result!

- Look inside the jet: **1-1 correspondence reco.**
    - Larger acceptance...
    - Excellent intrinsic resolutions
    - Extremely stable...
- Be addressed by state-of-art detector design, technology, and reconstruction algorithm!

Jet origin id



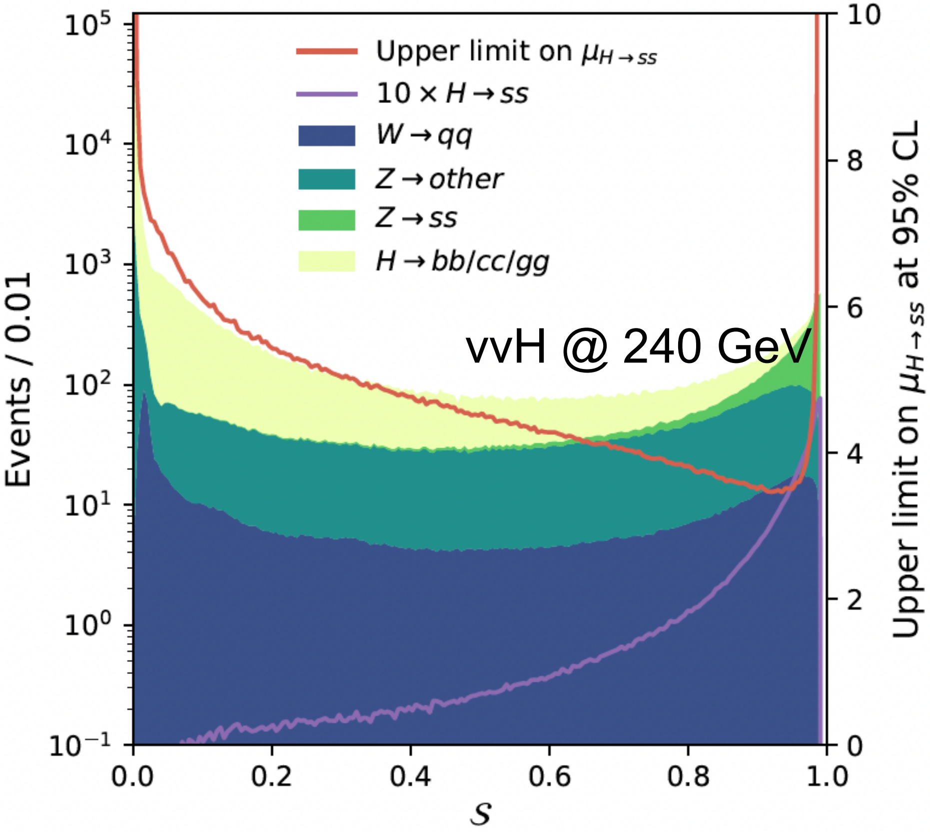
	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.745	0.163	0.033	0.025	0.004	0.003	0.002	0.003	0.002	0.002	0.017
$\bar{b}$	0.170	0.737	0.026	0.033	0.003	0.004	0.003	0.002	0.002	0.003	0.018
c	0.015	0.014	0.743	0.055	0.036	0.031	0.025	0.009	0.009	0.018	0.043
$\bar{c}$	0.016	0.015	0.056	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

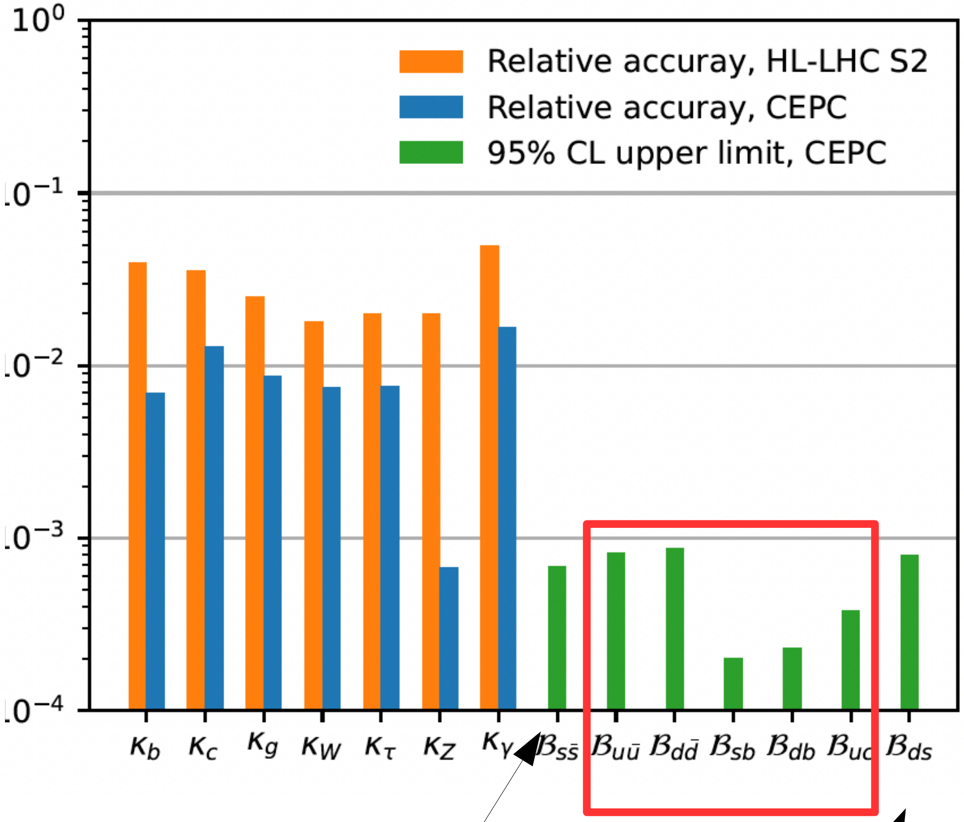


- 11 categories (5 quarks + 5 anti quarks + gluon) identification, realized at Full Simulated di-jet events at CEPC CDR baseline with Arbor + ParticleNet
- Published in PRL *132, 221802 (2024)*. Comment from the referee: *"demonstrate the world-leading performance of tagger", "a "game changer" and opens new horizons for precision flavor studies at all future experiments."*

https://arxiv.org/abs/2310.03440
https://arxiv.org/abs/2309.13231

Benchmark analyses: Higgs rare/FCNC



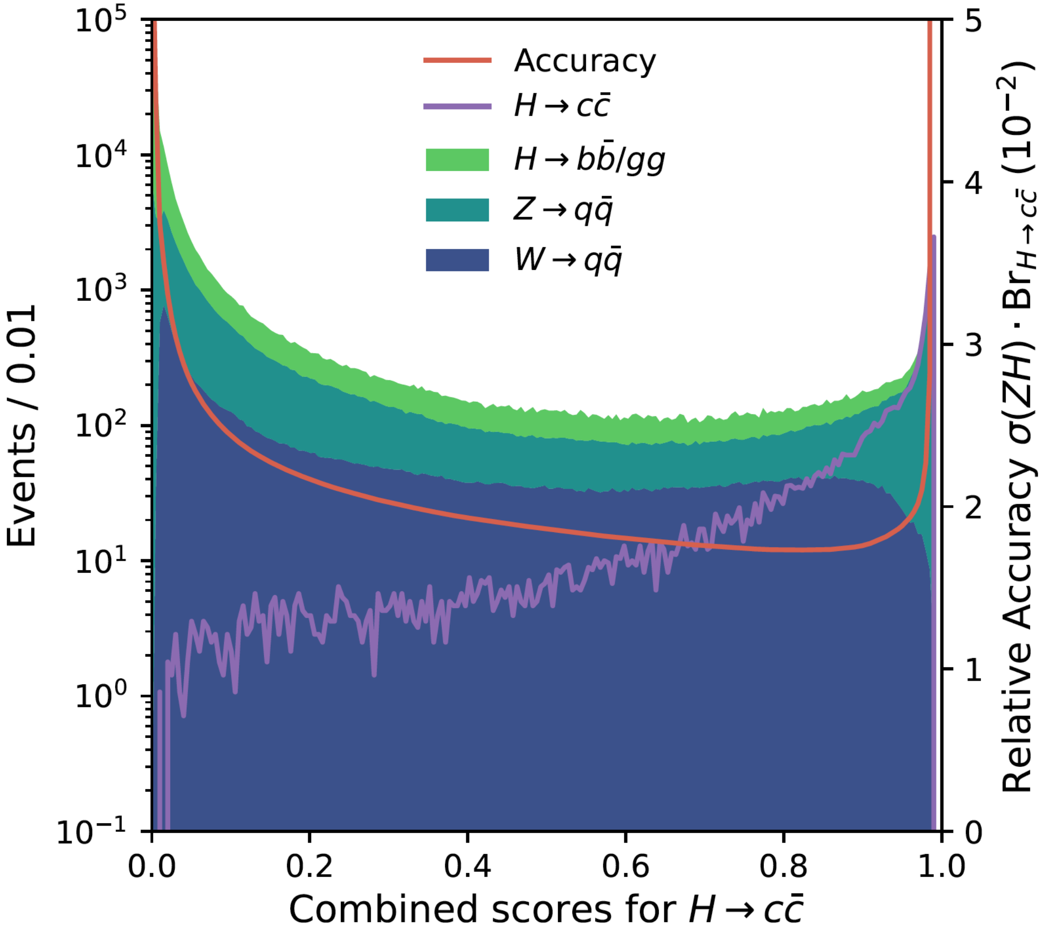


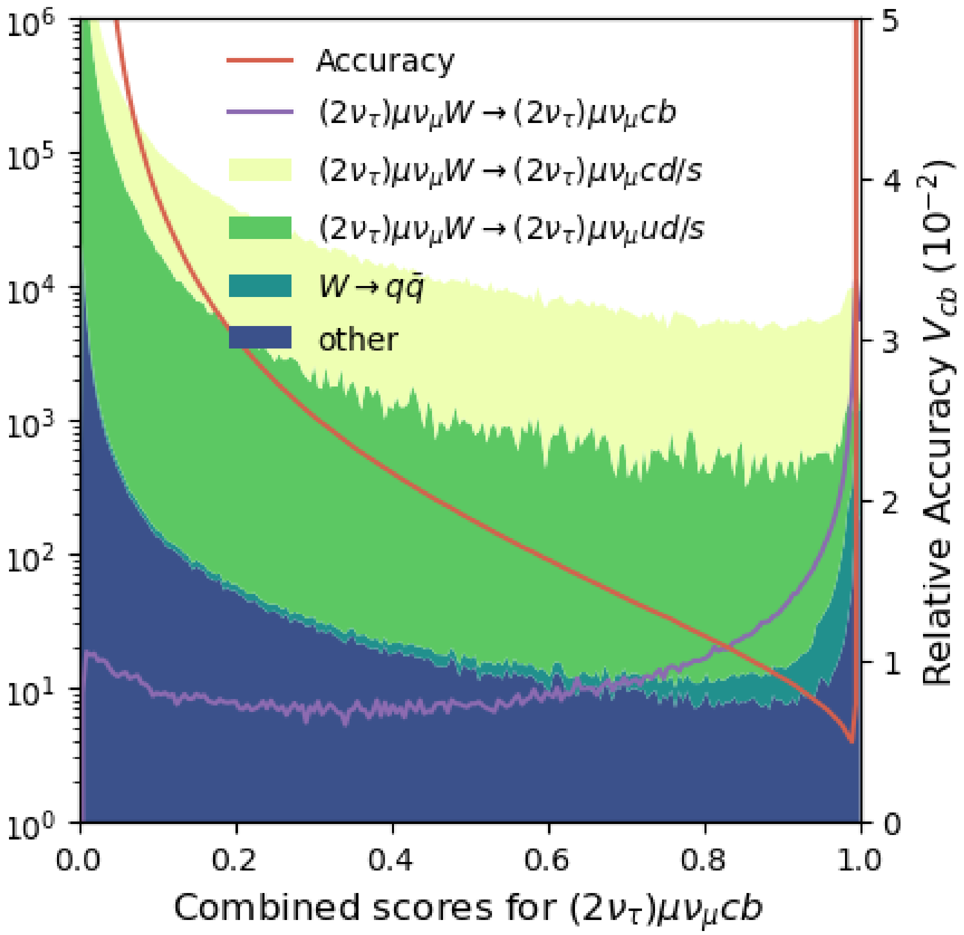
Improved by ~3 times

Improved by 1-2 orders of magnitudes

Presumably... firstly quantified

More benchmarks





- From Jet Flavor Tagging to Jet Origin ID:
  - vvH, H→cc: 3% → 1.7%
  - Vcb: 0.75% → 0.5% (other CKM elements on the target list)

Updated result on $\sin^2\theta_{eff}^{l}$ measurement

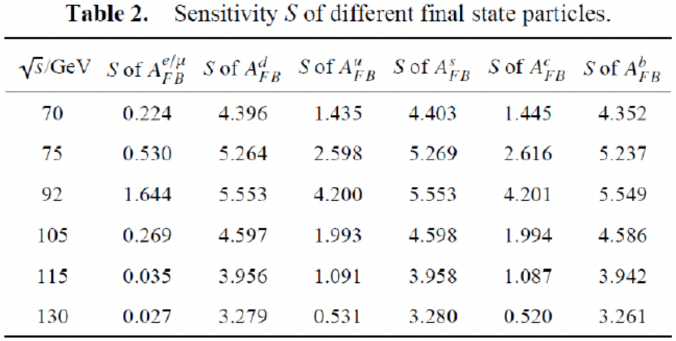


Table 2. Sensitivity S of different final state particles.

$\sqrt{s}$/GeV	S of $A_{FB}^{e/\mu}$	S of A_{FB}^{d}	S of A_{FB}^{u}	S of A_{FB}^{s}	S of A_{FB}^{c}	S of A_{FB}^{b}
70	0.224	4.396	1.435	4.403	1.445	4.352
75	0.530	5.264	2.598	5.269	2.616	5.237
92	1.644	5.553	4.200	5.553	4.201	5.549
105	0.269	4.597	1.993	4.598	1.994	4.586
115	0.035	3.956	1.091	3.958	1.087	3.942
130	0.027	3.279	0.531	3.280	0.520	3.261

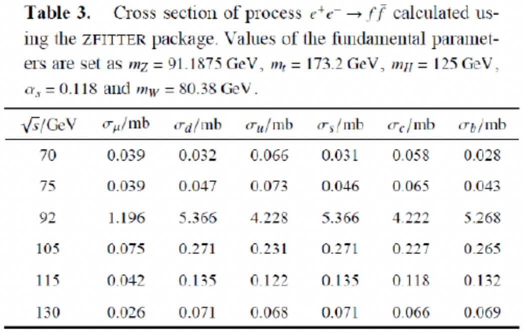


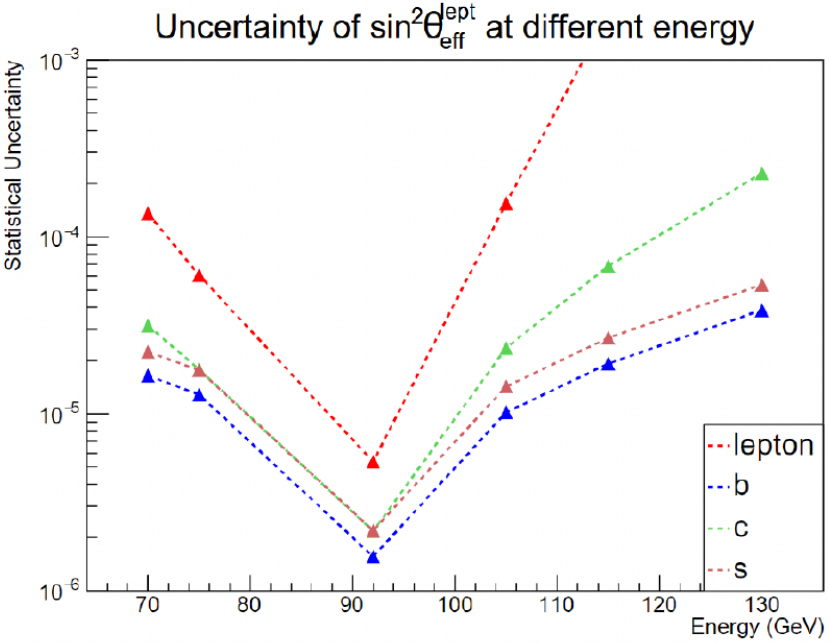
Table 3. Cross section of process $e^+e^- \to f\bar{f}$ calculated using the ZFITTER package. Values of the fundamental parameters are set as $m_Z = 91.1875$ GeV, $m_t = 173.2$ GeV, $m_H = 125$ GeV, $\alpha_s = 0.118$ and $m_W = 80.38$ GeV.

$\sqrt{s}$/GeV	σ_μ/mb	σ_d/mb	σ_u/mb	σ_s/mb	σ_c/mb	σ_b/mb
70	0.039	0.032	0.066	0.031	0.058	0.028
75	0.039	0.047	0.073	0.046	0.065	0.043
92	1.196	5.366	4.228	5.366	4.222	5.268
105	0.075	0.271	0.231	0.271	0.227	0.265
115	0.042	0.135	0.122	0.135	0.118	0.132
130	0.026	0.071	0.068	0.071	0.066	0.069

Verify the RG behavior… using ~1 month of data taking

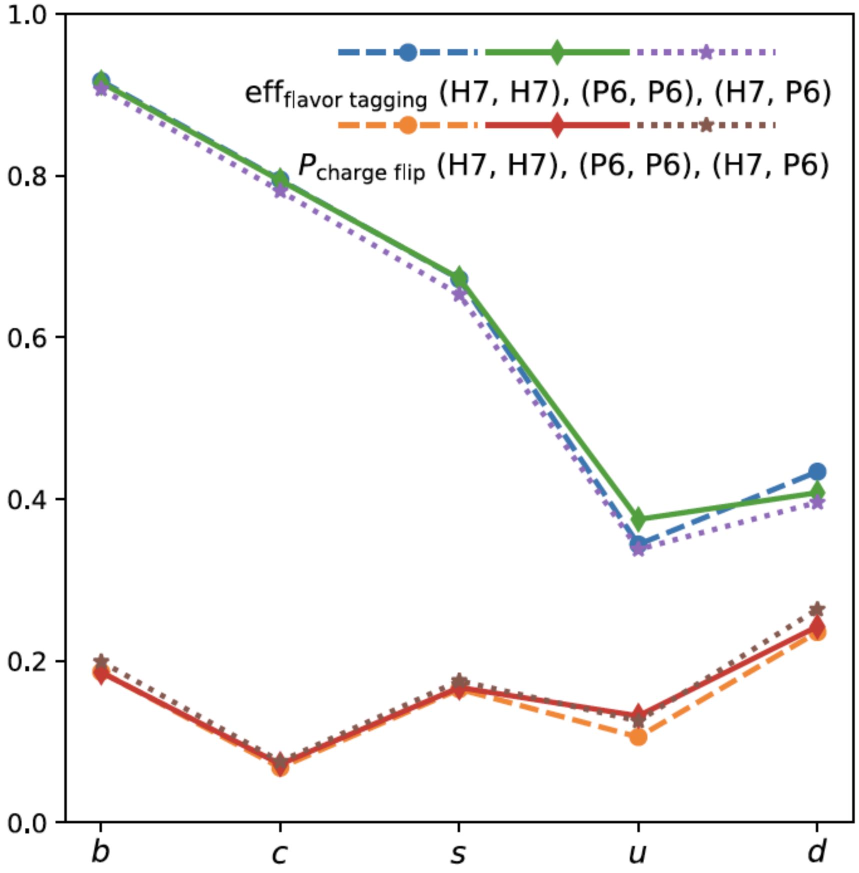
Expected statistical uncertainties on $\sin^2\theta_{eff}^{l}$ measurement.

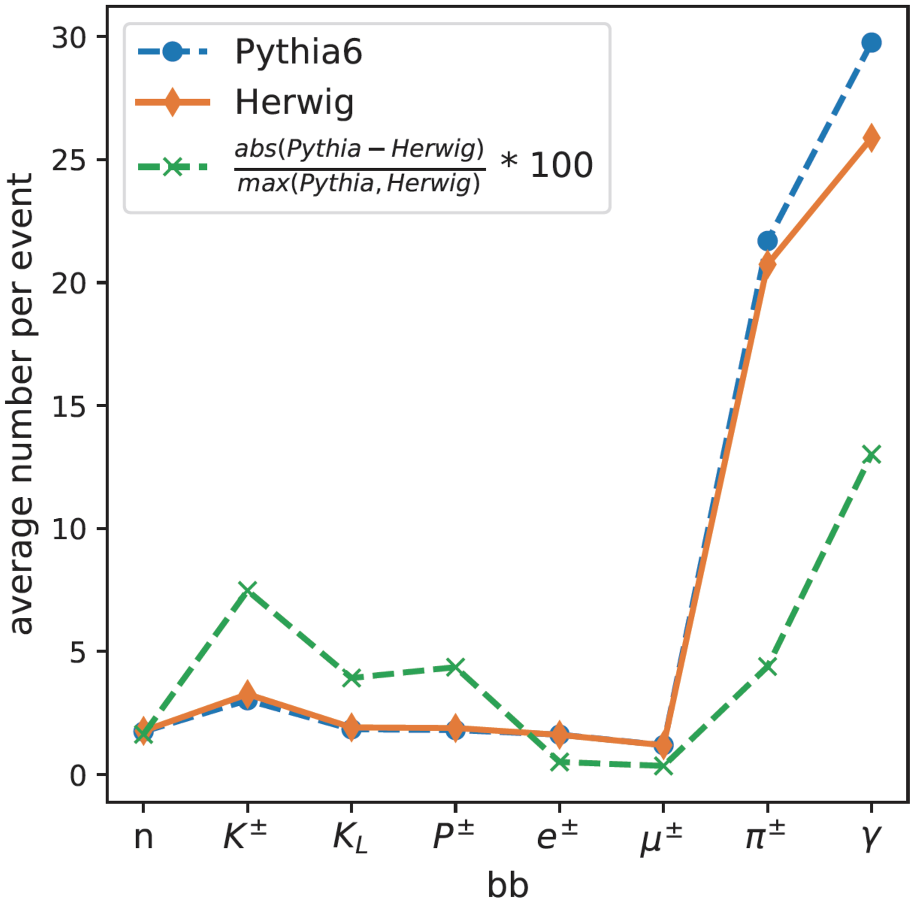
(Using one-month data collection, ~ *4e12/24 Z events* at Z pole)



$\sqrt{s}$	b	c	s
70	1.6×10^{-5}	3.2×10^{-5}	2.2×10^{-5}
75	1.3×10^{-5}	1.8×10^{-5}	1.8×10^{-5}
92	1.6×10^{-6}	2.2×10^{-6}	2.2×10^{-6}
105	1.0×10^{-5}	2.4×10^{-5}	1.4×10^{-5}
115	1.9×10^{-5}	6.8×10^{-5}	2.7×10^{-5}
130	3.9×10^{-5}	2.3×10^{-4}	5.4×10^{-5}

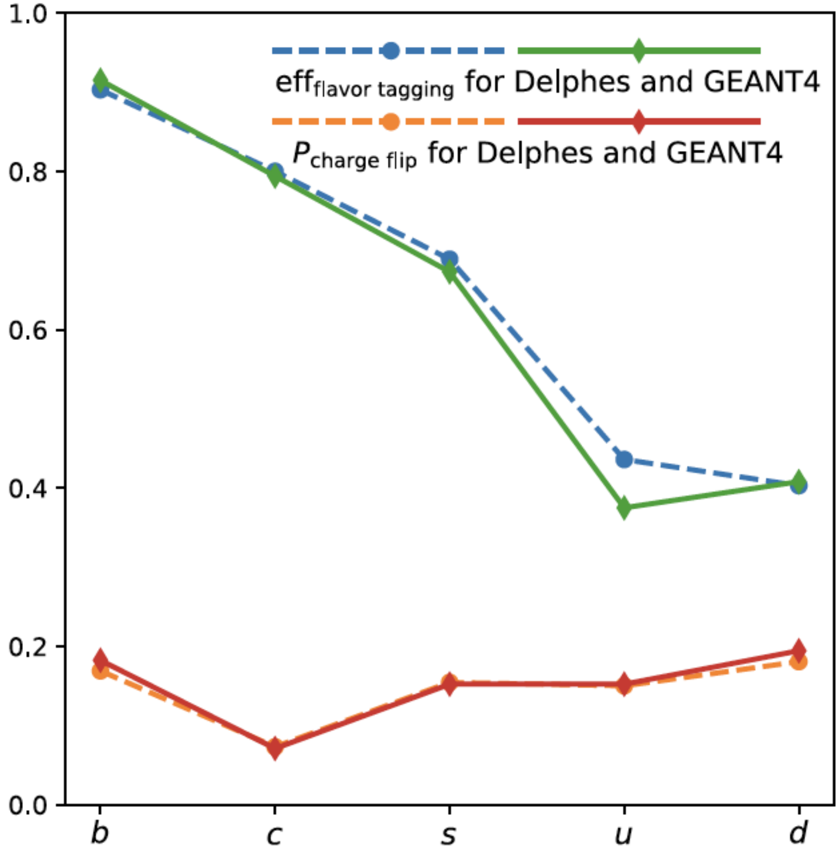
V.S. Hadronization models



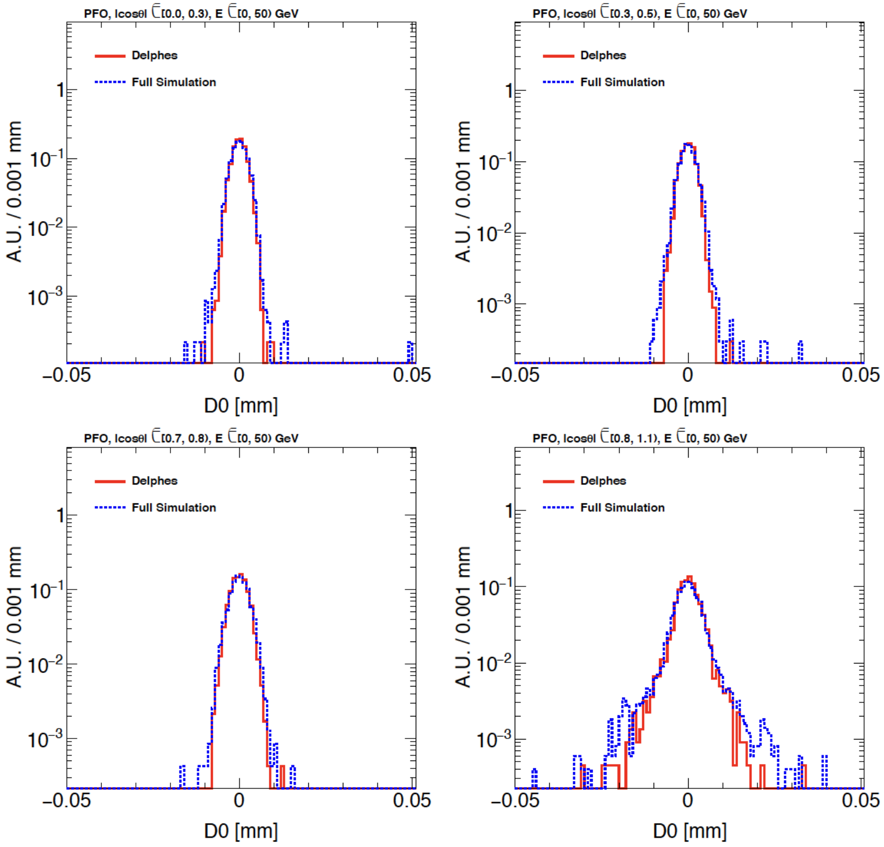


- Different hadronization model have significantly different predictions...

Fast/Full Simulation

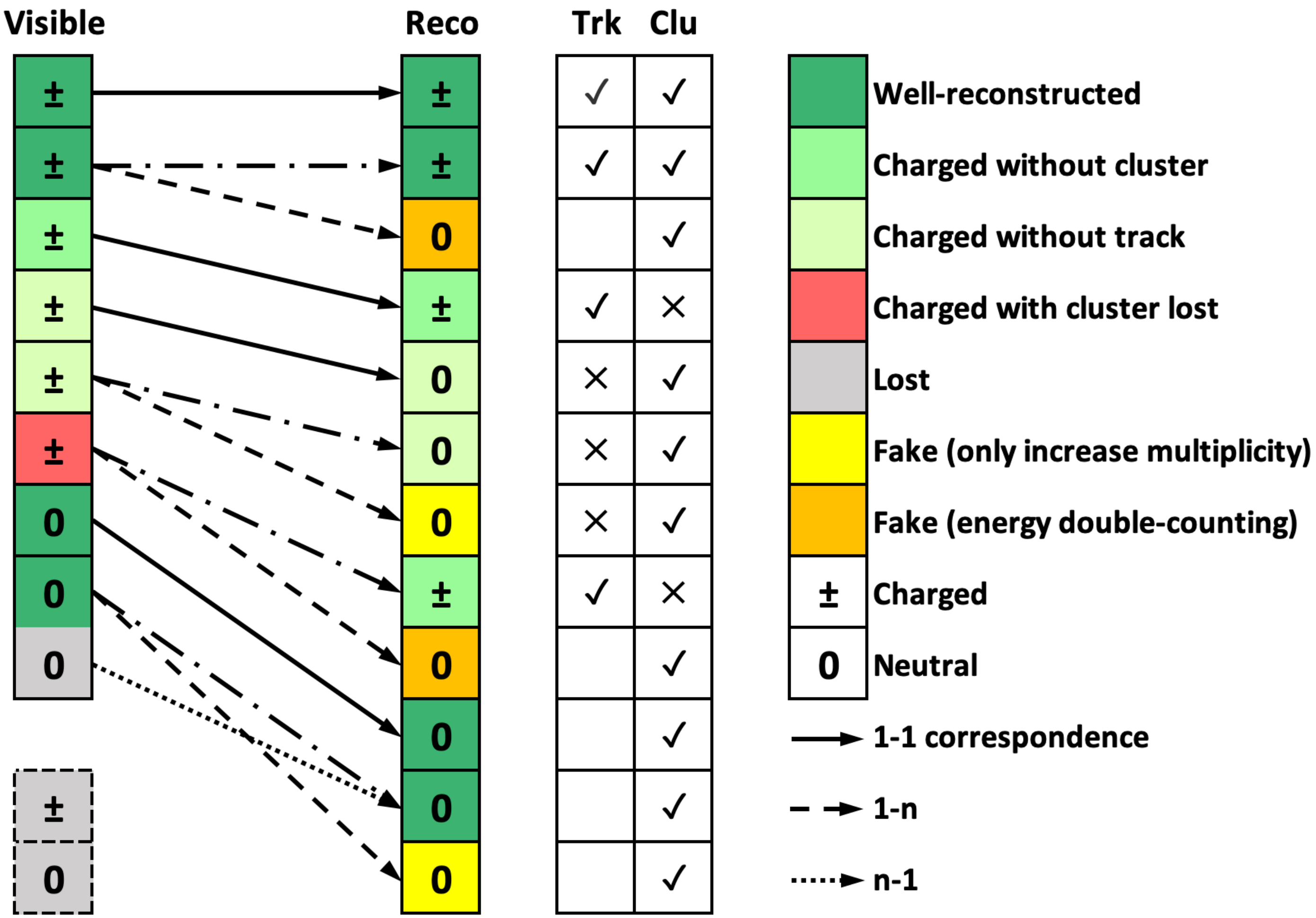


Z->μμ (91.2 GeV)



- Delphes ~ Perfect PFA (1 – 1 correspondence..)

1-1 correspondence: ultimate Mapping between visible & reco



Visible

- ±
- ±
- ±
- ±
- ±
- ±
- 0
- 0
- 0

- ±
- 0

Reco	Trk	Clu
±	✓	✓
±	✓	✓
0		✓
±	✓	×
0	×	✓
0	×	✓
0	×	✓
±	✓	×
0		✓
0		✓
0		✓
0		✓

Well-reconstructed

Charged without cluster

Charged without track

Charged with cluster lost

Lost

Fake (only increase multiplicity)

Fake (energy double-counting)

± **Charged**

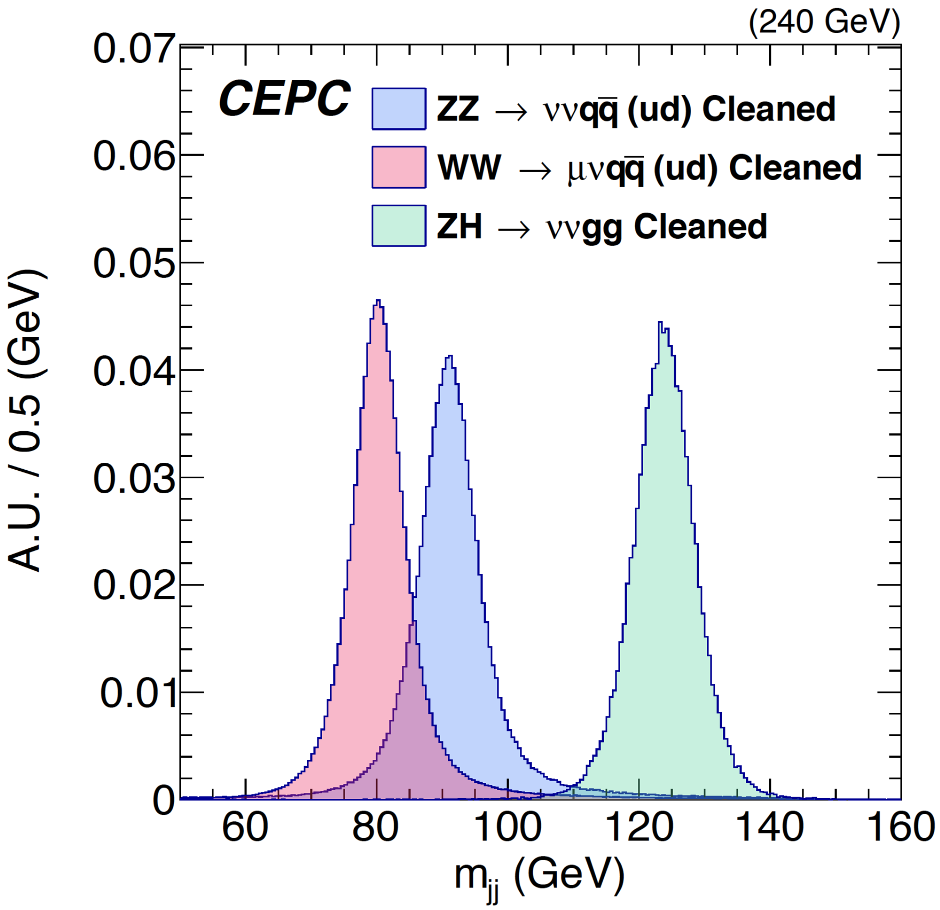
0 **Neutral**

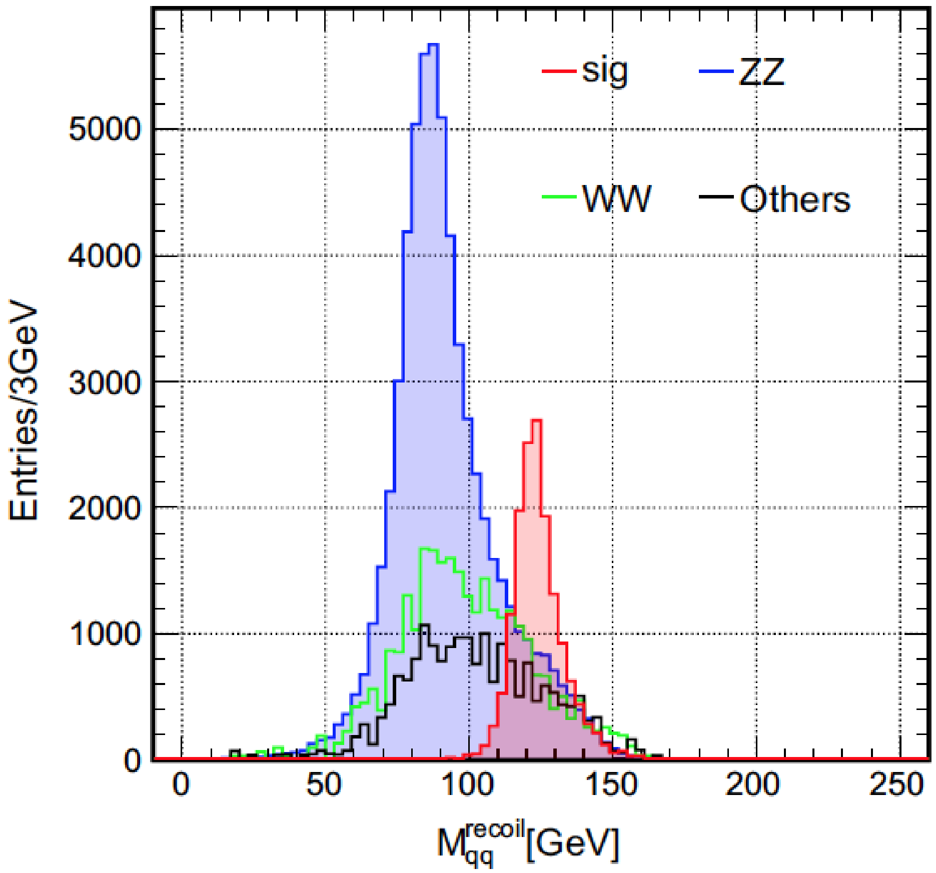
→ **1-1 correspondence**

- - → **1-n**

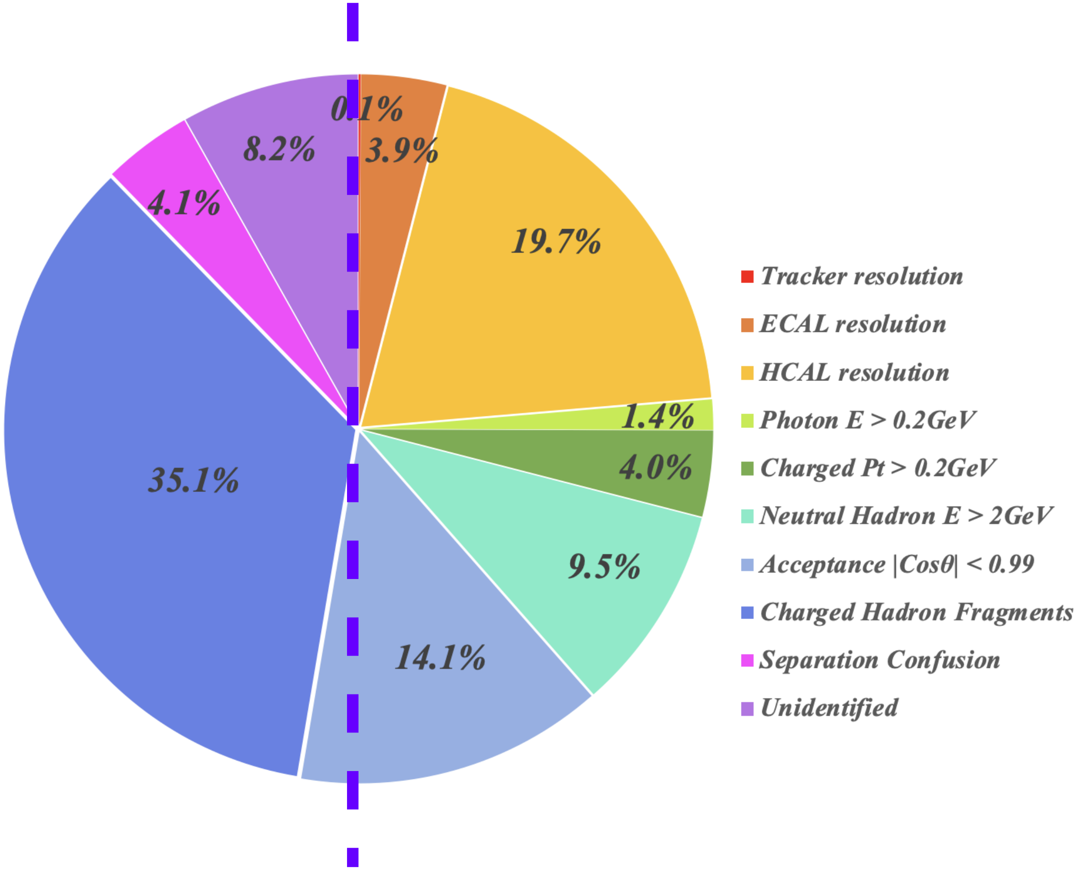
····→ **n-1**

Boson Mass Resolution: Key Per. Para



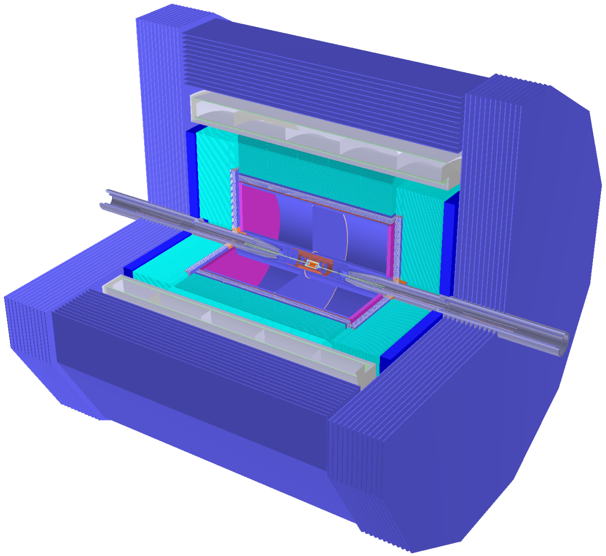


BMR decomposition @ CDR baseline

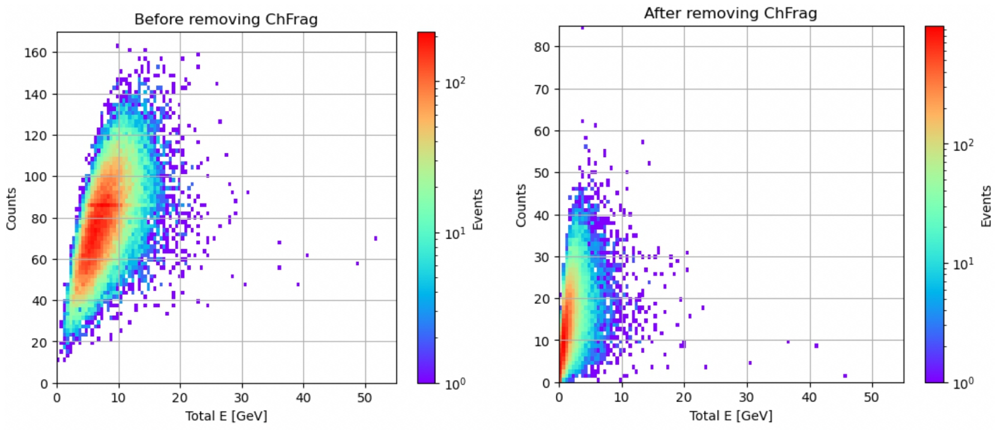


- *1st HCAL resolution dominant the uncertainties from intrinsic detector resolution: need better HCAL → usage of GSHCAL*
- *2nd Leading contribution: Confusion from shower Fragments (fake particles), need better Pattern Reco.*

- *CDR baseline - GRPC HCAL*

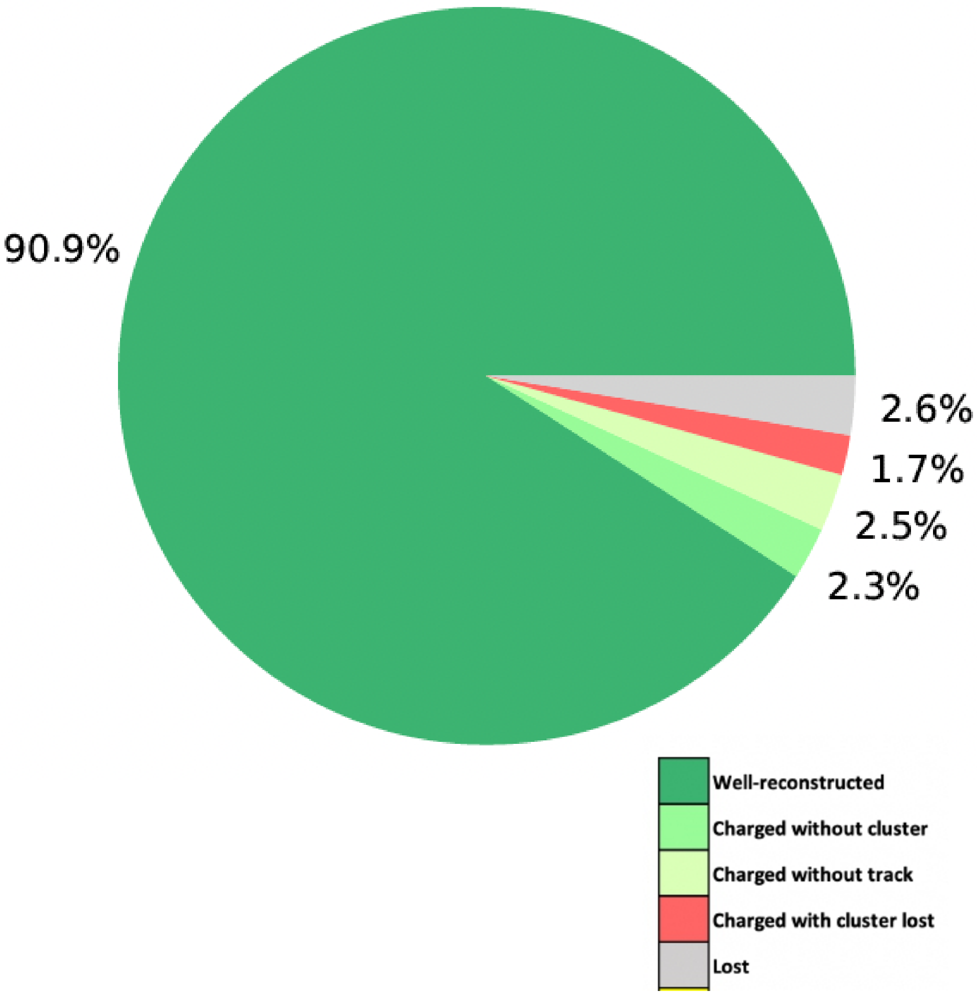


Confusion identification & treatment: frag. veto

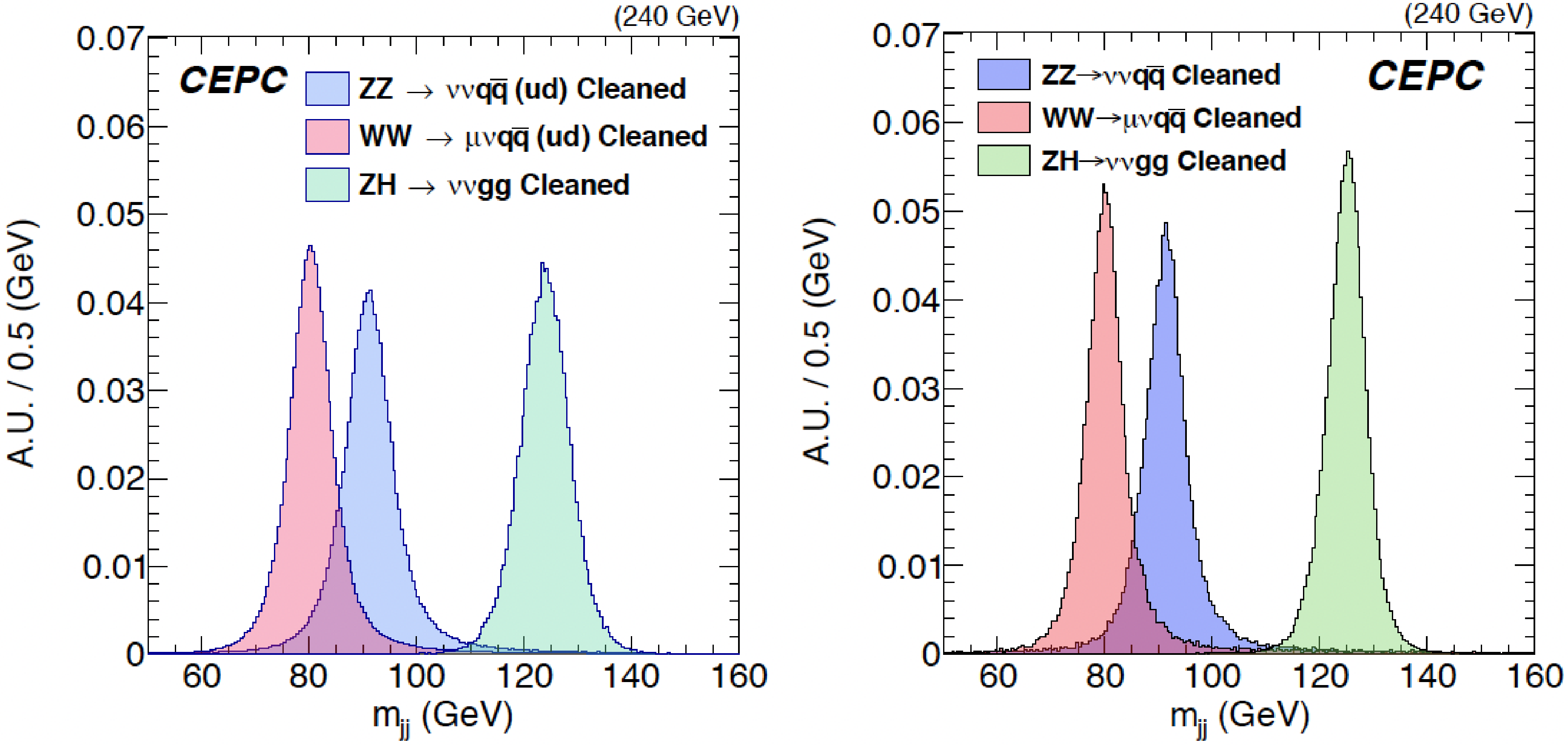


Fake particle originated Confusion reduced by 1 order of magnitude, at nominal vvH, H→gg event

Ignoring the remaining fragments with total E < 1 GeV, more than 95% of the visible energy preserves 1-1 correspondence



BMR of 2.75% reached

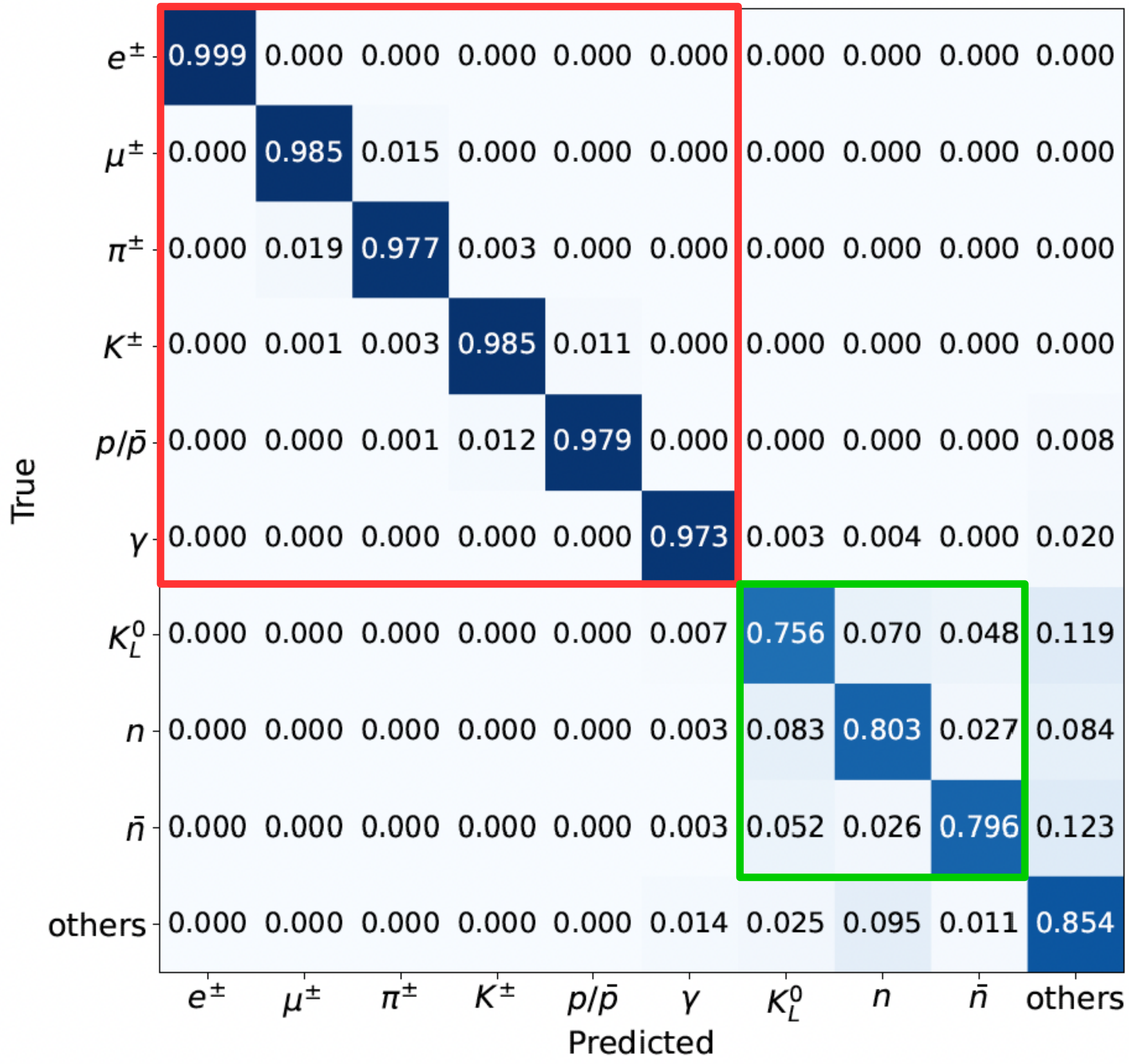


Detector change (GRPC HCAL (5 λ)→GSHCAL (6 λ)): BMR 3.7% → 3.4%;

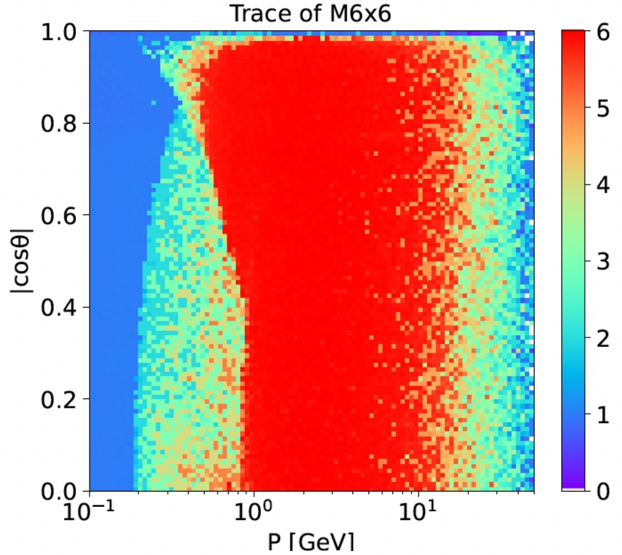
AI enhanced reconstruction: 3.4% → 2.8%.

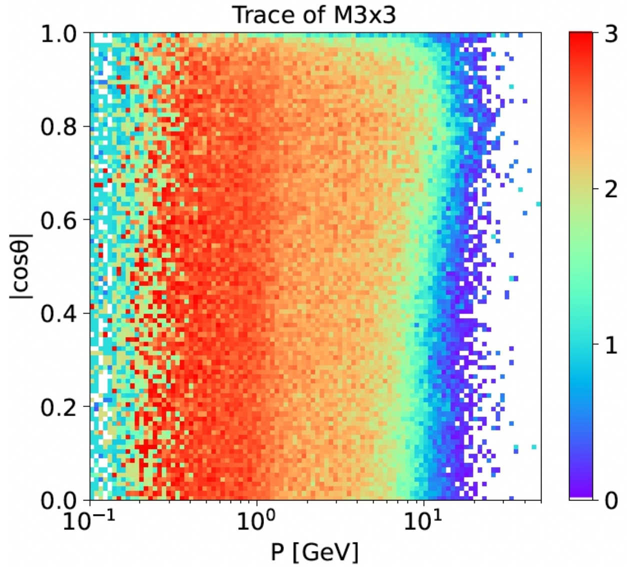
Impact from Beam induced background + impact on objects inside jet reco: to be evaluated.

Pid: differential performance



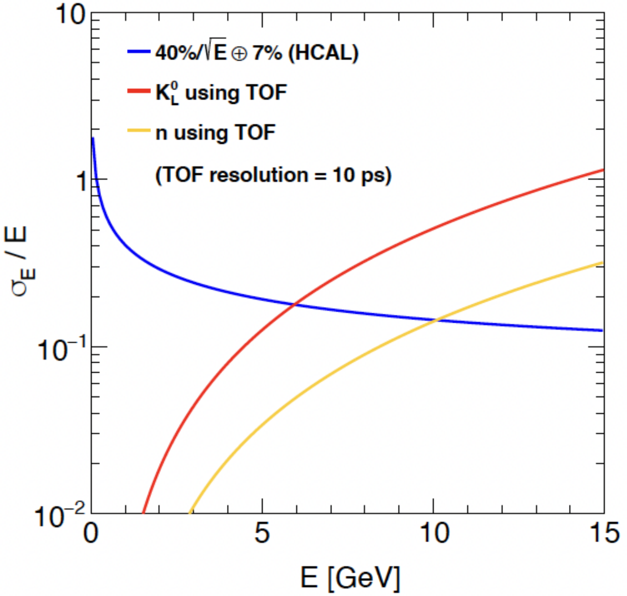
True \ Predicted	$e^{\pm}$	$\mu^{\pm}$	$\pi^{\pm}$	$K^{\pm}$	$p/\bar{p}$	γ	K_L^0	n	$\bar{n}$	others
$e^{\pm}$	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\mu^{\pm}$	0.000	0.985	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\pi^{\pm}$	0.000	0.019	0.977	0.003	0.000	0.000	0.000	0.000	0.000	0.000
$K^{\pm}$	0.000	0.001	0.003	0.985	0.011	0.000	0.000	0.000	0.000	0.000
$p/\bar{p}$	0.000	0.000	0.001	0.012	0.979	0.000	0.000	0.000	0.000	0.008
γ	0.000	0.000	0.000	0.000	0.000	0.973	0.003	0.004	0.000	0.020
K_L^0	0.000	0.000	0.000	0.000	0.000	0.007	0.756	0.070	0.048	0.119
n	0.000	0.000	0.000	0.000	0.000	0.003	0.083	0.803	0.027	0.084
$\bar{n}$	0.000	0.000	0.000	0.000	0.000	0.003	0.052	0.026	0.796	0.123
others	0.000	0.000	0.000	0.000	0.000	0.014	0.025	0.095	0.011	0.854

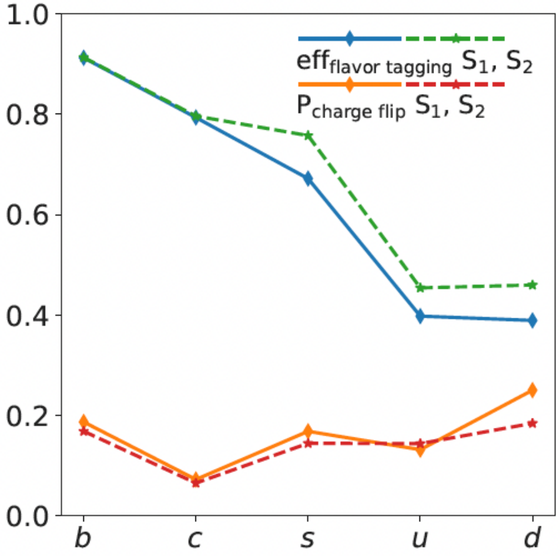


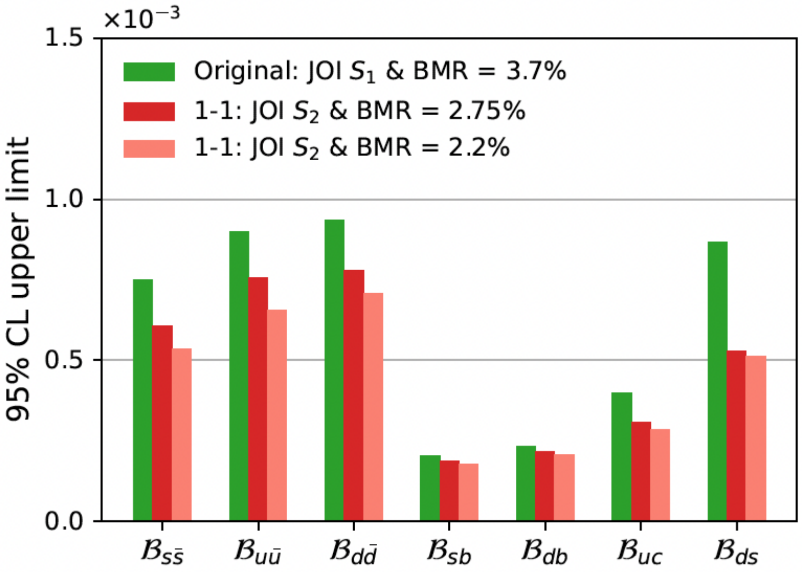


Neutral Hadron ID: excellent Calorimetry with ToF capability (δt~100 ps/hit)

Perspectives with 1-1 correspondence







- ToF enhanced energy measurement: BMR: 2.8% → 2.2-2.4%
  - Need excellent CALO + ToF ~ o(10 ps)
  - Assume Low energy neutrons & secondary particles can be tamed... still very challenge...
- Strongly Boost the light quark ID.
- Benchmark precision improved... up to nearly two times.

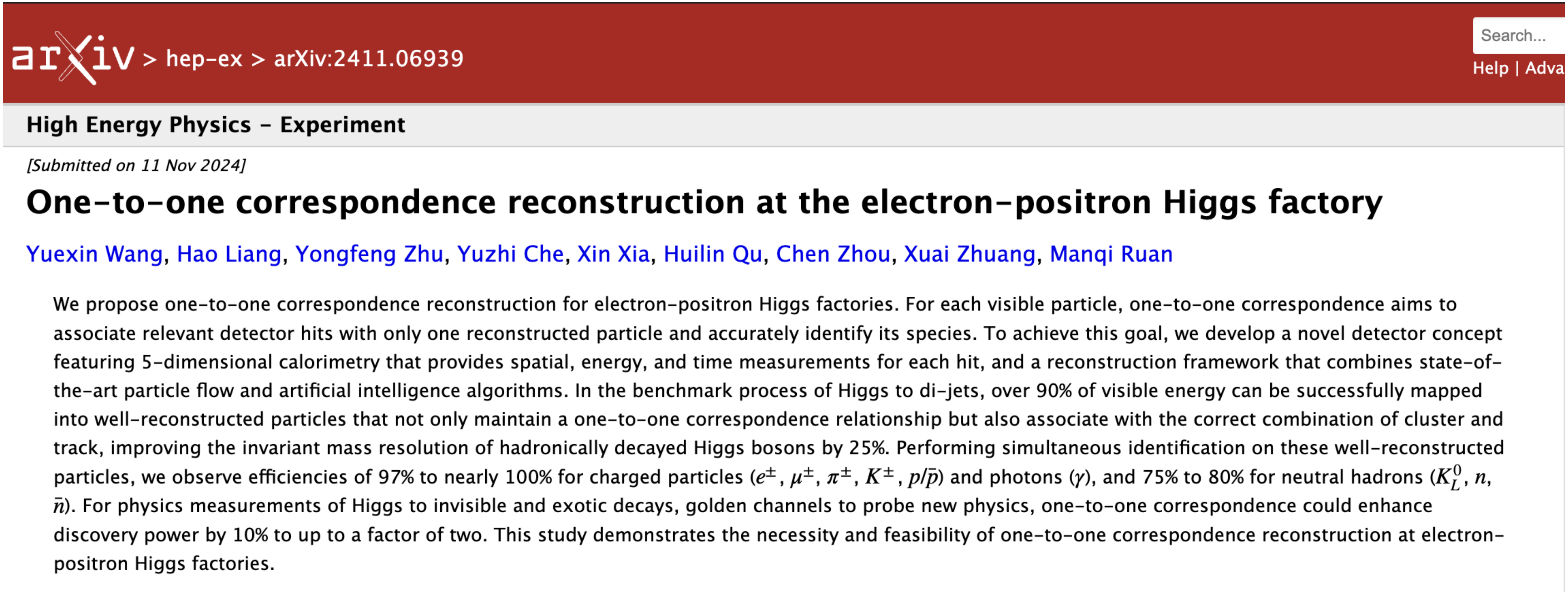
1-1 Correspondence

Holistic description of physics events

Efficient & interpretable information compression: (o(1E5) Hits → o(100) reco particles)

~ Confusion Free PFA + Excellent Particle identification

~ New method for the detector monitoring & measurements



High Energy Physics – Experiment

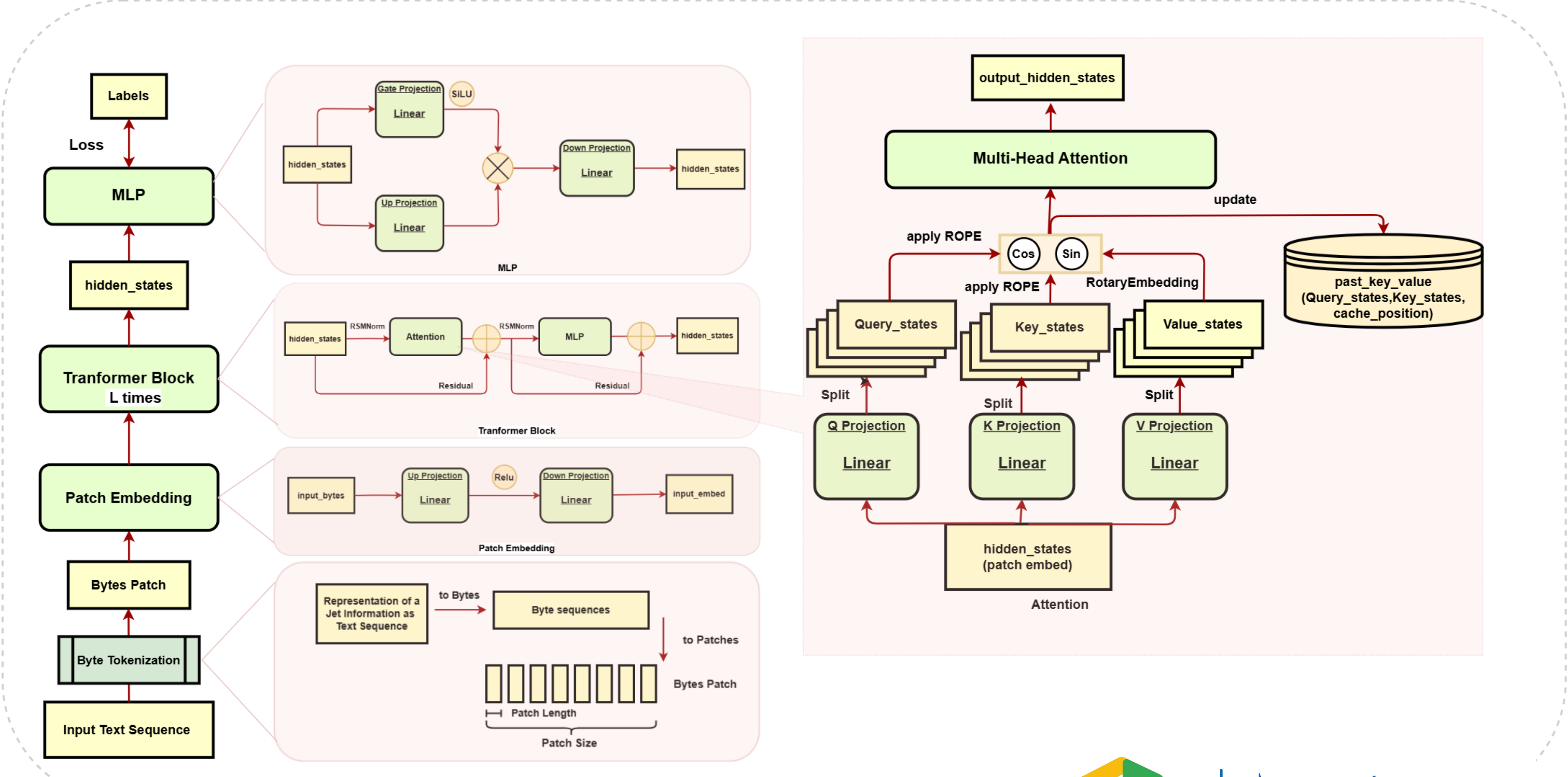
[Submitted on 11 Nov 2024]

One-to-one correspondence reconstruction at the electron-positron Higgs factory

Yuexin Wang, Hao Liang, Yongfeng Zhu, Yuzhi Che, Xin Xia, Huilin Qu, Chen Zhou, Xuai Zhuang, Manqi Ruan

We propose one-to-one correspondence reconstruction for electron-positron Higgs factories. For each visible particle, one-to-one correspondence aims to associate relevant detector hits with only one reconstructed particle and accurately identify its species. To achieve this goal, we develop a novel detector concept featuring 5-dimensional calorimetry that provides spatial, energy, and time measurements for each hit, and a reconstruction framework that combines state-of-the-art particle flow and artificial intelligence algorithms. In the benchmark process of Higgs to di-jets, over 90% of visible energy can be successfully mapped into well-reconstructed particles that not only maintain a one-to-one correspondence relationship but also associate with the correct combination of cluster and track, improving the invariant mass resolution of hadronically decayed Higgs bosons by 25%. Performing simultaneous identification on these well-reconstructed particles, we observe efficiencies of 97% to nearly 100% for charged particles ($e^\pm$, $\mu^\pm$, $\pi^\pm$, $K^\pm$, $p/\bar{p}$) and photons (γ), and 75% to 80% for neutral hadrons (K_L^0, n, $\bar{n}$). For physics measurements of Higgs to invisible and exotic decays, golden channels to probe new physics, one-to-one correspondence could enhance discovery power by 10% to up to a factor of two. This study demonstrates the necessity and feasibility of one-to-one correspondence reconstruction at electron-positron Higgs factories.

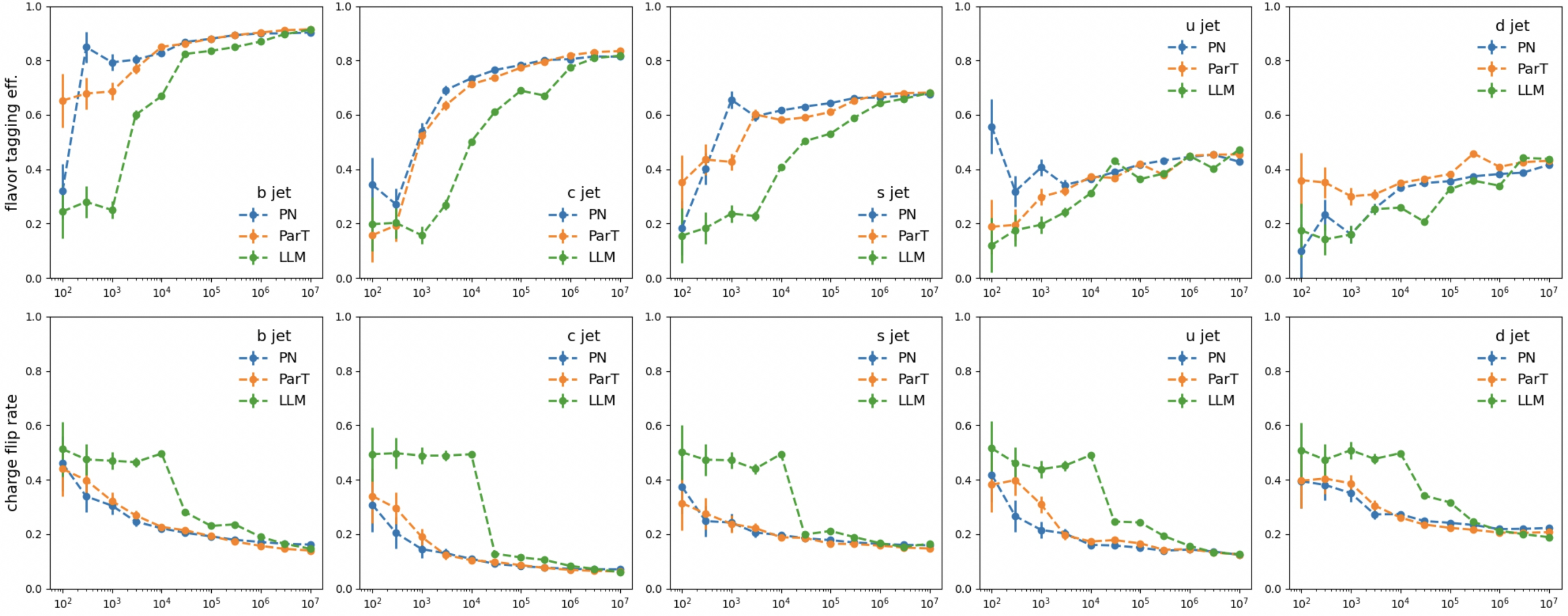
Ongoing study: from specialized Models to LLM



BIN-BBT Model

- New tokenization method to address numeric problems at LLM

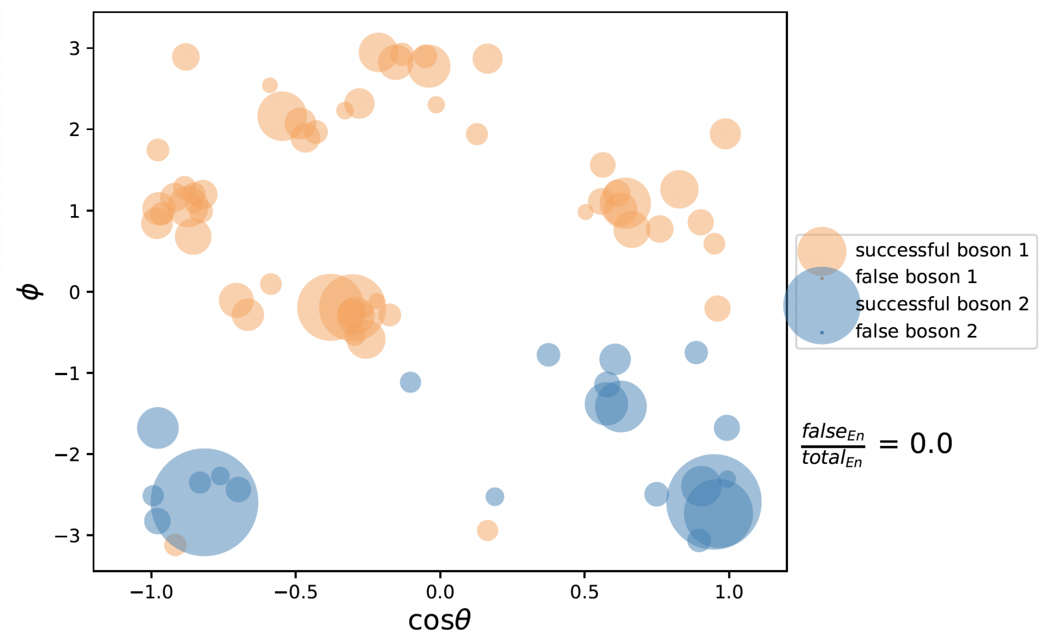
Ongoing study: from specialized Models to LLM

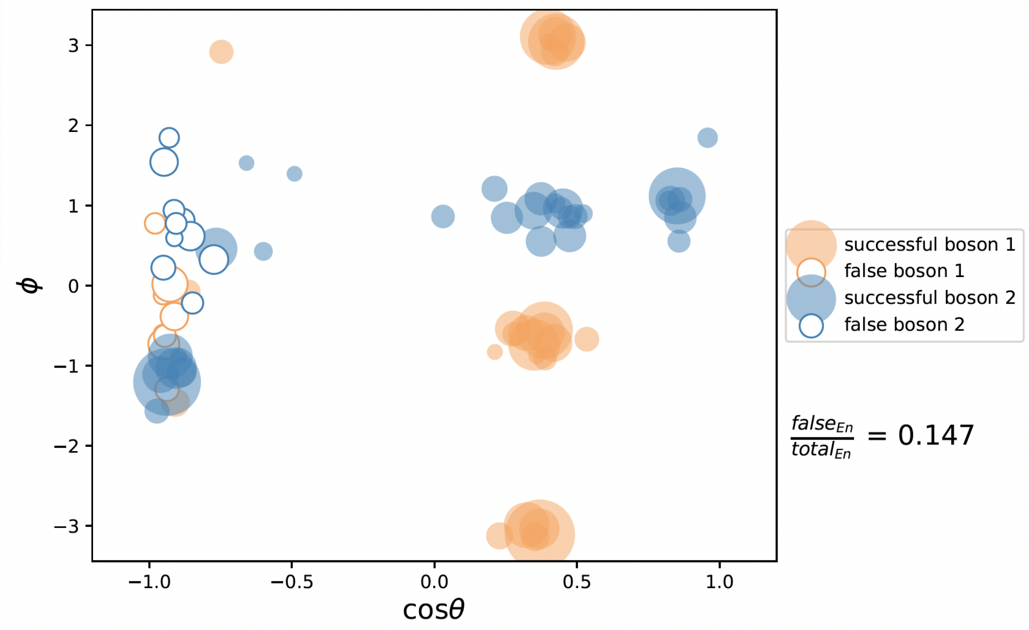


- Comparable result with different scaling behavior
- Para. Numbers: PN 360k, ParT 2.4M, BINBBT 150 M
- Be submit to arXiv soon

超对称
Super Symmetry
Technologies

Color Singlet Identification





- CSI: identify the color single origin of each final state particle
- Grouping problem: essential for the physics measurements with multi-jet events, i.e., measurements with full hadronic ZH events
- AI might well strongly enhance its performance: compared to conventional jet clustering & matching

Meta questions

- Problem categorization
  - Identification problem: JoI, Pid, 1-1 correspondence (from Arbor)
  - Grouping problem: Color singlet id, tracking, clustering, ...
  - Assessment/regression problem: such as energy/momentum/time estimation, fitting
  - What's the most suited corresponding AI architecture, or general AI, and Why?

- AI for HEP, and HEP for AI (HEP → Science)
  - HEP, as a mature & vivid field, has the potential to impact the AI development, i.e., interpretability analysis

- Be relax, and have fun!...

Summary

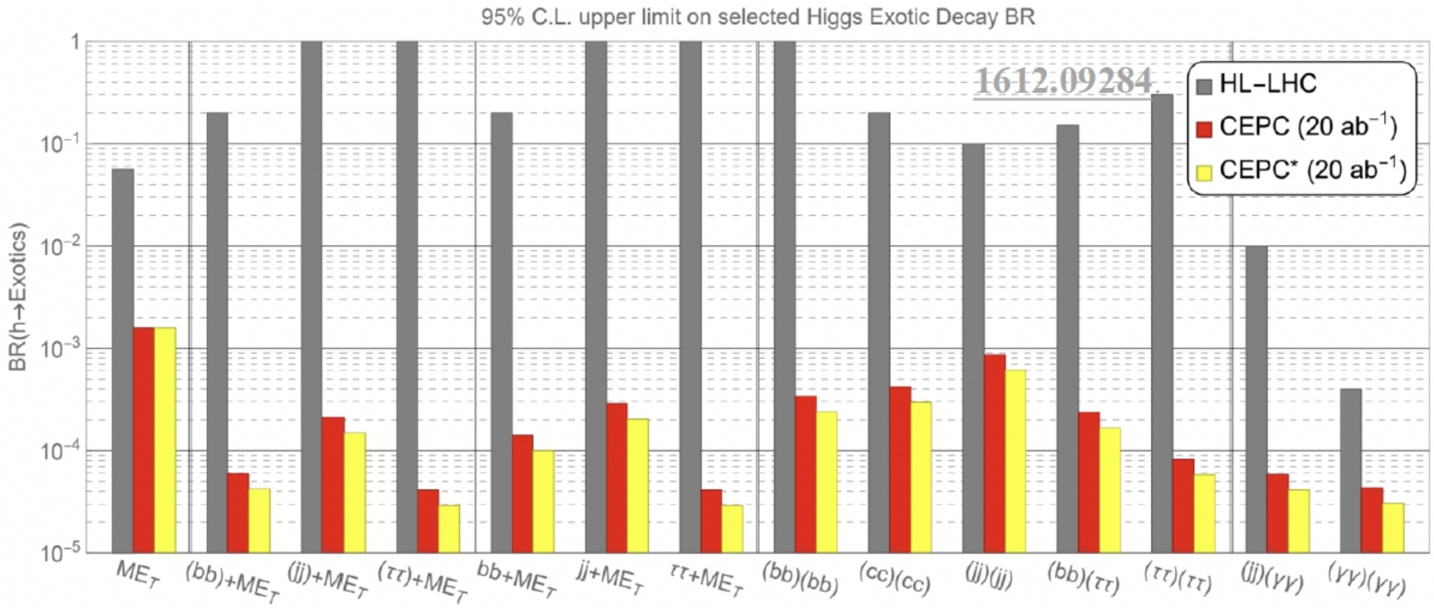
- Higgs factory: extremely rich physics requires excellent performance
    - Excellent Pattern, reco → high eff/purity & precision reco. of all physics objects
    - Large acceptance, Extremely stable & excellent intrinsic det performance

- AI: the trends & indispensable tool towards this requirements
    - Significantly enhance the physics reach & alters the detector design/optimization
    - Jet Origin ID: 'see' quark & gluon as lepton & photon
        - *...A "game changer" and opens new horizon for precise flavor studies at all future experiments...*
    - 1-1 correspondence, at least at Higgs factory:
        - Should be pursued because of its impact & elegancy
        - Could be achieved using state-of-art AI + detector design & technology

- Lots to be explored: Large Language Model, CSI, Meta questions, General AI...

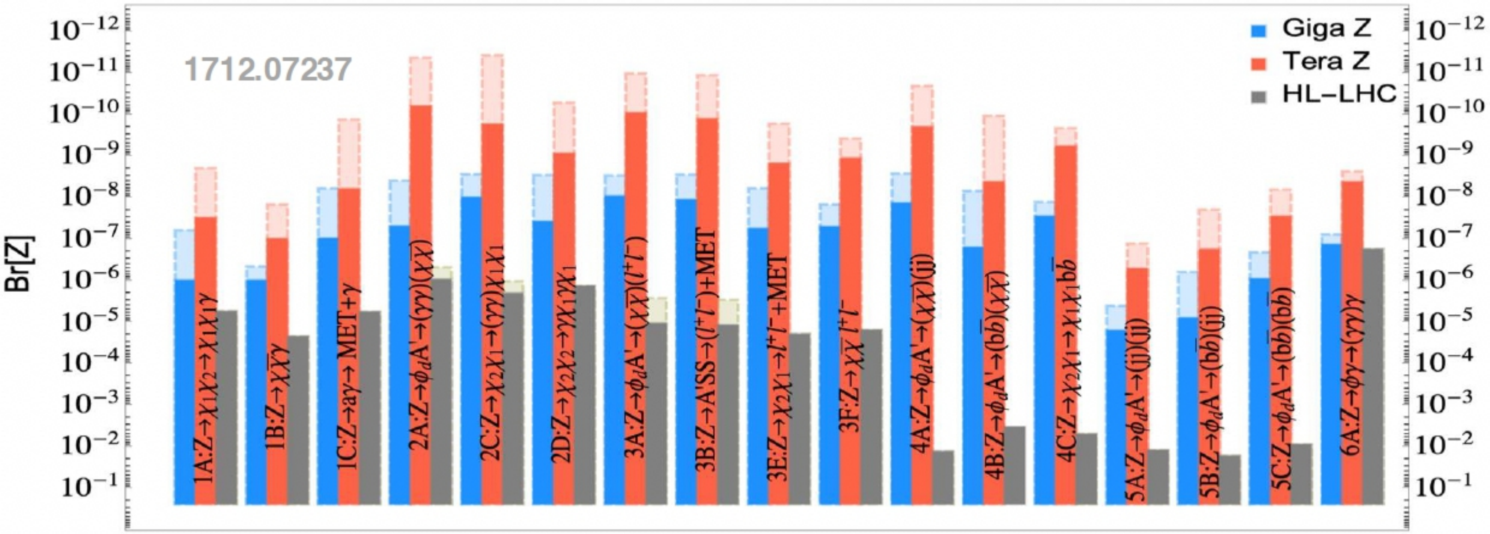
Back up

Exotic decays



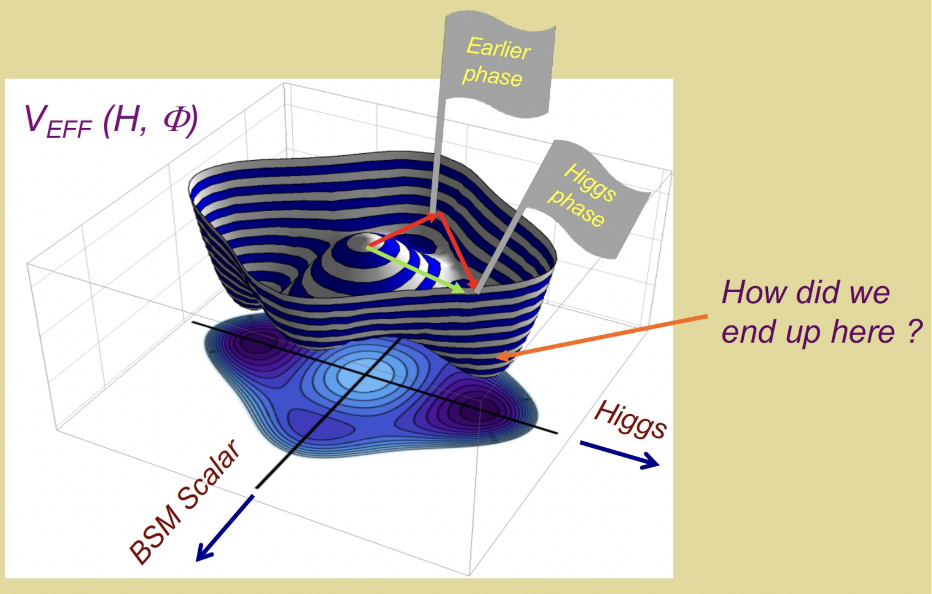
The 95% C.L. upper limit on selected Higgs exotic decay BR

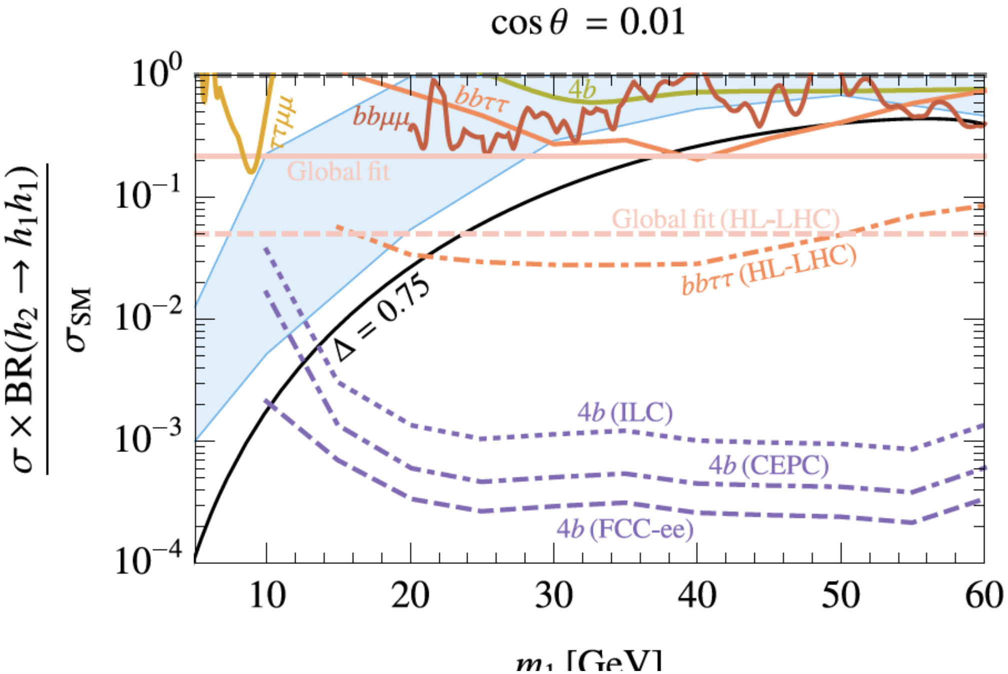
- Credit: Zhen Liu, Jia Liu, Xuai Zhuang, etc

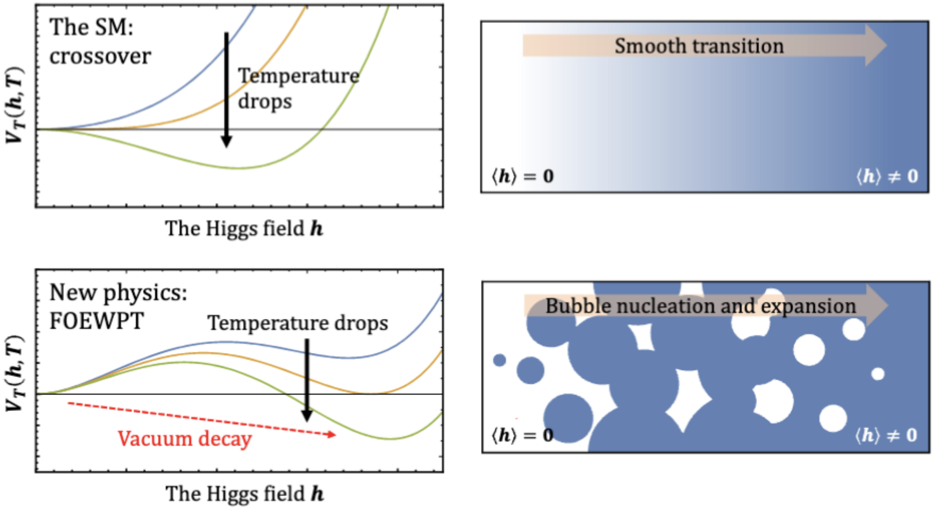


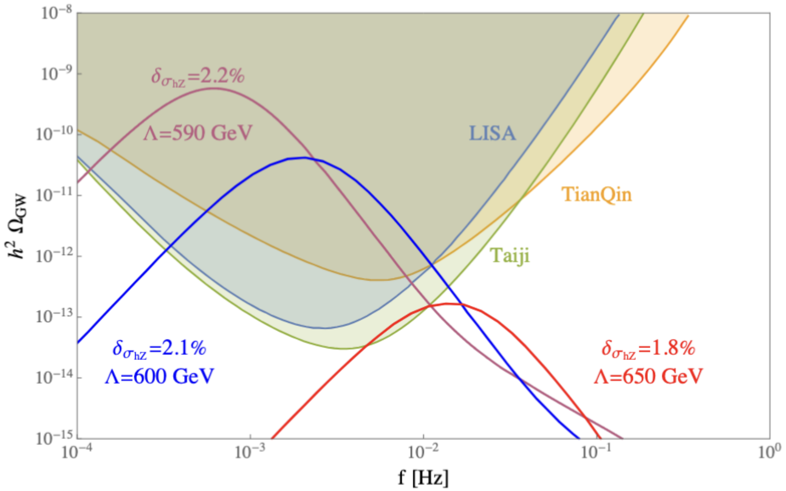
The reach for the branching ratio of various exotic Z decay modes

Phase Transition in early Universe





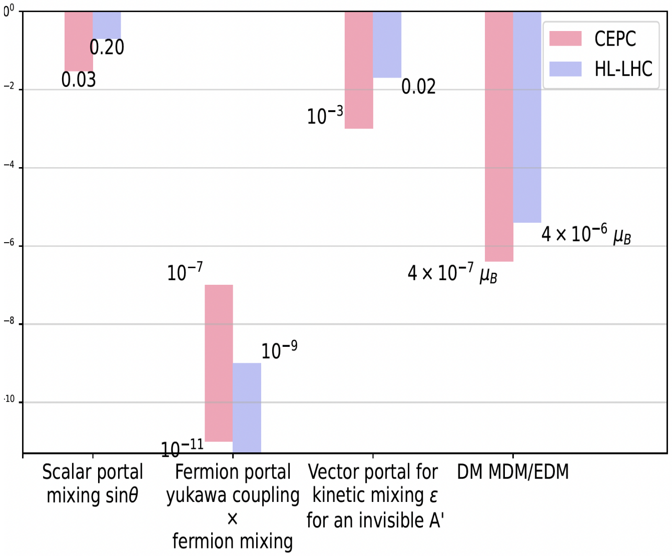


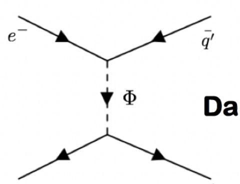


- Credit: Michael Ramse Musolfy, Kepan Xie, etc

Dark sector

Portal	Effective operator	$\sqrt{s}$ [GeV]	$\mathcal{L}[ab^{-1}]$	Sensitivity of CEPC (HL-LHC)	Figs.	Ref.
Scalar	$\lambda_{HP}\lvert H\rvert^2 S^2 \to$ scalar mixing $\sin\theta$	250	5	invisible S, $\sin\theta \approx 0.03$ (0.20 global-fits)	22	[108]
Fermion	$y_\ell \bar{\chi}_L S^\dagger \ell_R$ + H.c.	250	5	covering $100\,\text{GeV} < m_S < 170\,\text{GeV}$	23	[56]
	$\kappa\Phi\overline{q'_L}\ell_R$ + H.c. (dark QCD)	250	5	$m_\Phi \sim 10$ TeV for $c\tau_{\text{darkpion}} \in [1, 10^3]$ cm (Null)	25	[109]
	$y\Phi\bar{F}_L\ell_R$ + H.c.	240	5.6	$y\theta_L \in [10^{-11},\ 10^{-7}]$ ($\lesssim 10^{-8} - 10^{-9}$)	26	[110]
Vector	$A'_\mu\left(e\epsilon J^\mu_{\text{em}} + g_D\bar{\chi}\gamma^\mu\chi\right)$	250	5	$\epsilon \sim 10^{-3}$ for $g_D = e$ and $m_{A'} < 125$ GeV ($\epsilon \sim 0.02$)	27, 28	[108]
	$\varepsilon A_\mu\bar{\chi}\gamma^\mu\chi$, (millicharge DM)	250	5	$\epsilon \sim 0.1$ for $m_\chi \sim 50$ GeV	29	[111]
		91.2	2.6	$\epsilon \sim 0.02$ for $m_\chi \sim 5$ GeV		
		160	16	$\epsilon \sim 0.5$ for $m_\chi \sim 10$ GeV		
	$\frac{1}{2}\mu_\chi\bar{\chi}\sigma^{\mu\nu}\chi F_{\mu\nu} + \frac{i}{2}d_\chi\bar{\chi}\sigma^{\mu\nu}\gamma^5\chi F_{\mu\nu}$ $-a_\chi\bar{\chi}\gamma^\mu\gamma^5\chi\partial^\nu F_{\mu\nu} + b_\chi\bar{\chi}\gamma^\mu\chi\partial^\nu F_{\mu\nu}$	91.2	100	$\mu_\chi, d_\chi \sim 4\times10^{-7}$ $(4\times10^{-6})\mu_B$ for $m_\chi < 25\,\text{GeV}$	30	[112]
		240	20	$a_\chi, b_\chi \sim 10^{-6}$ $(2\times10^{-6})\,\text{GeV}^{-2}$ for $\text{m}_\chi < 80$ GeV		
EFT	$\frac{1}{\Lambda^2}\sum_i\left(\bar{\chi}\gamma_\mu(1-\gamma_5)\chi\right)\left(\bar{\ell}\gamma^\mu(1-\gamma_5)\ell\right)$	250	5	$\Lambda_i \sim 2$ TeV $(m_\chi = 0)$ (Null)	31	[113]
	$\frac{1}{\Lambda_A^2}\bar{\chi}\gamma_\mu\gamma_5\chi\bar{\ell}\gamma^\mu\gamma_5\ell$	250	5	$\Lambda_A \sim 1.5$ TeV (Null)	32	[111]
	$\sum_i \frac{1}{\Lambda_i^2}\left(\bar{e}\Gamma_\mu e\right)\left(\bar{\nu}_L\Gamma^\mu\chi_L\right)$ + H.c. $\Gamma_\mu = 1, \gamma_5, \gamma_\mu, \gamma_\mu\gamma_5, \sigma_{\mu\nu}$	240	20	$\Lambda_i \sim 1$ TeV $(m_\chi = 0)$ (Null)	33	[114]



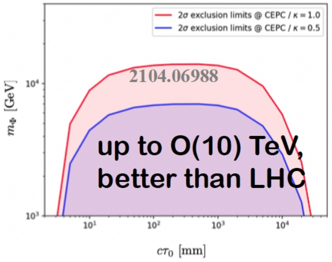


Decay to leptons

Dark quark q'

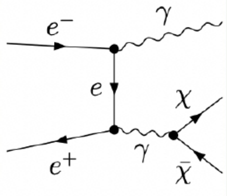
Jet-like

Signature: Lepton + jet + MET



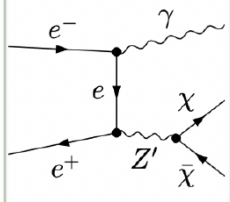
up to O(10) TeV, better than LHC

millicharged DM



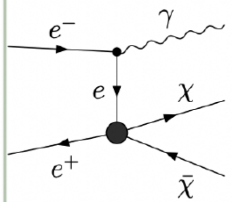
photon mediator

vector-portal DM

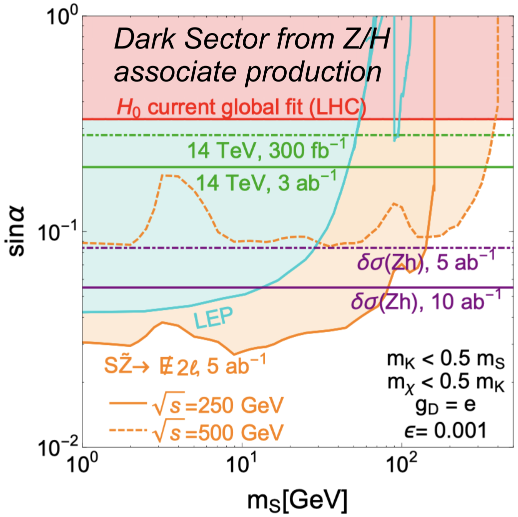


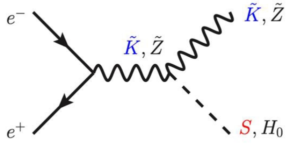
Z′ mediator

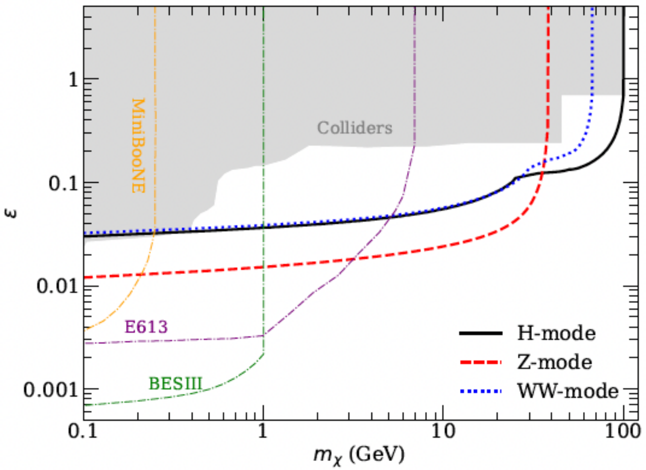
EFT DM



4-F interaction



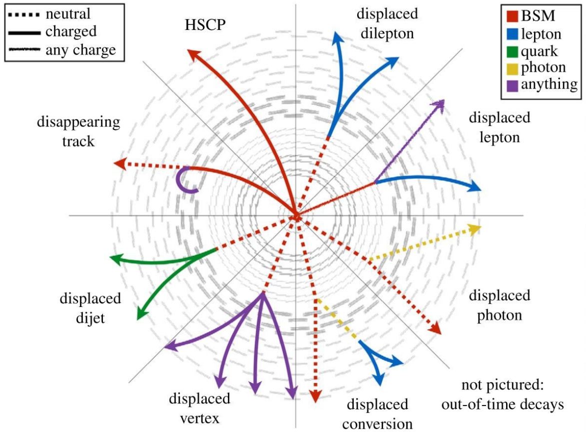


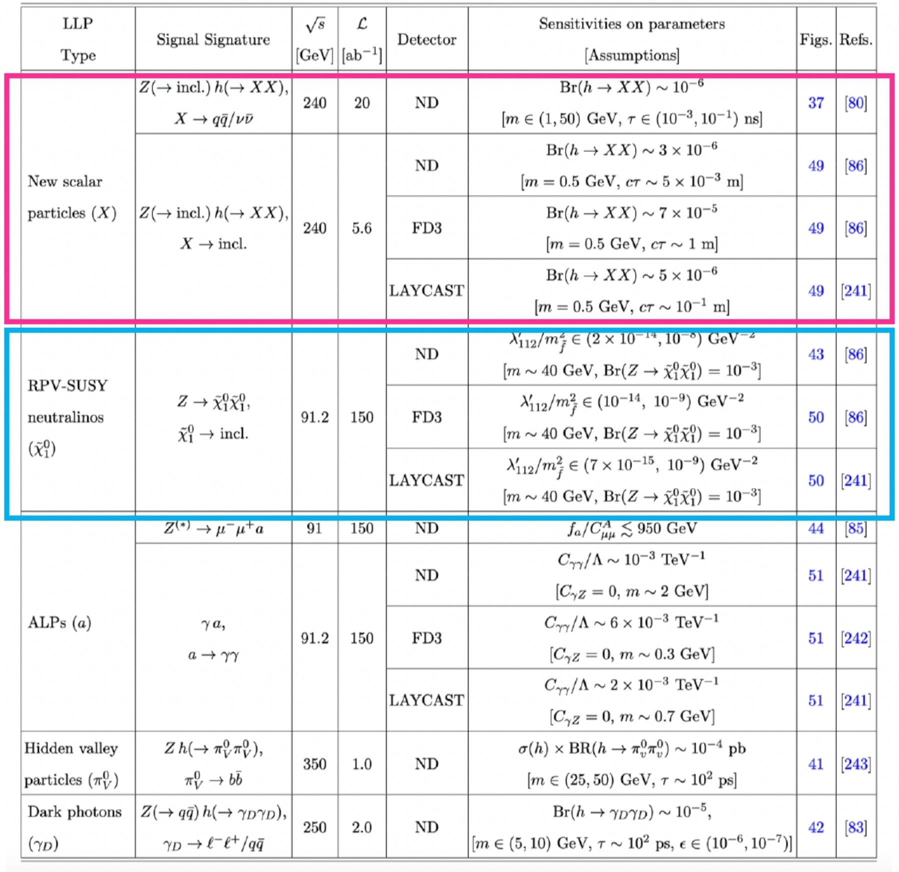


Vector portal DM

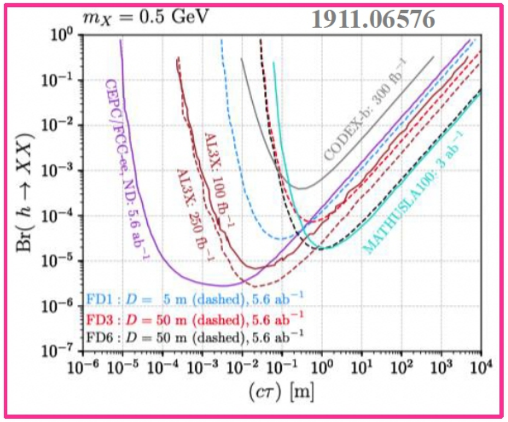
- Credit: Jia Liu, etc

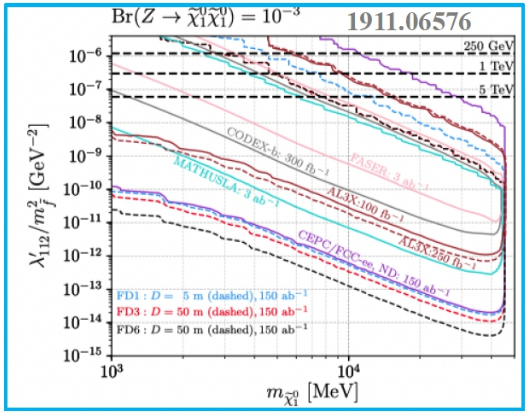
LLP, especially with Far detector

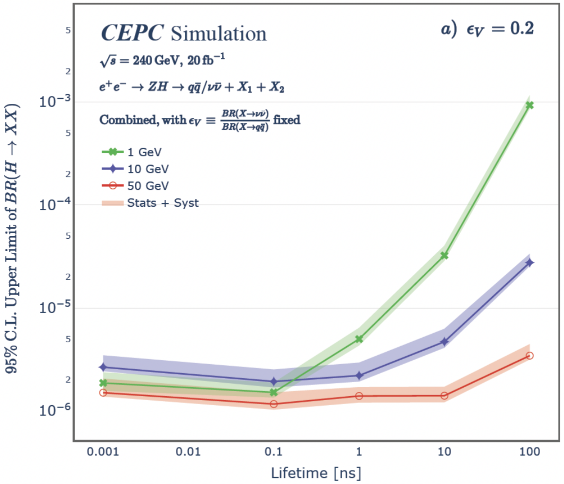




LLP Type	Signal Signature	$\sqrt{s}$ [GeV]	$\mathcal{L}$ [ab^{-1}]	Detector	Sensitivities on parameters [Assumptions]	Figs.	Refs.
New scalar particles (X)	$Z(\to \text{incl.})\, h(\to XX)$, $X \to q\bar{q}/\nu\bar{\nu}$	240	20	ND	$\text{Br}(h \to XX) \sim 10^{-6}$ $[m \in (1,50)\ \text{GeV},\ \tau \in (10^{-3}, 10^{-1})\ \text{ns}]$	37	[80]
	$Z(\to \text{incl.})\, h(\to XX)$, $X \to \text{incl.}$	240	5.6	ND	$\text{Br}(h \to XX) \sim 3 \times 10^{-6}$ $[m = 0.5\ \text{GeV},\ c\tau \sim 5 \times 10^{-3}\ \text{m}]$	49	[86]
				FD3	$\text{Br}(h \to XX) \sim 7 \times 10^{-5}$ $[m = 0.5\ \text{GeV},\ c\tau \sim 1\ \text{m}]$	49	[86]
				LAYCAST	$\text{Br}(h \to XX) \sim 5 \times 10^{-6}$ $[m = 0.5\ \text{GeV},\ c\tau \sim 10^{-1}\ \text{m}]$	49	[241]
RPV-SUSY neutralinos ($\tilde{\chi}_1^0$)	$Z \to \tilde{\chi}_1^0\tilde{\chi}_1^0$, $\tilde{\chi}_1^0 \to \text{incl.}$	91.2	150	ND	$\lambda'_{112}/m_{\tilde{f}}^2 \in (2 \times 10^{-14}, 10^{-8})\ \text{GeV}^{-2}$ $[m \sim 40\ \text{GeV},\ \text{Br}(Z \to \tilde{\chi}_1^0\tilde{\chi}_1^0) = 10^{-3}]$	43	[86]
				FD3	$\lambda'_{112}/m_{\tilde{f}}^2 \in (10^{-14},\ 10^{-9})\ \text{GeV}^{-2}$ $[m \sim 40\ \text{GeV},\ \text{Br}(Z \to \tilde{\chi}_1^0\tilde{\chi}_1^0) = 10^{-3}]$	50	[86]
				LAYCAST	$\lambda'_{112}/m_{\tilde{f}}^2 \in (7 \times 10^{-15},\ 10^{-9})\ \text{GeV}^{-2}$ $[m \sim 40\ \text{GeV},\ \text{Br}(Z \to \tilde{\chi}_1^0\tilde{\chi}_1^0) = 10^{-3}]$	50	[241]
ALPs (a)	$Z^{(*)} \to \mu^-\mu^+ a$	91	150	ND	$f_a/C_{\mu\mu}^A \lesssim 950\ \text{GeV}$	44	[85]
	$\gamma\, a$, $a \to \gamma\gamma$	91.2	150	ND	$C_{\gamma\gamma}/\Lambda \sim 10^{-3}\ \text{TeV}^{-1}$ $[C_{\gamma Z} = 0,\ m \sim 2\ \text{GeV}]$	51	[241]
				FD3	$C_{\gamma\gamma}/\Lambda \sim 6 \times 10^{-3}\ \text{TeV}^{-1}$ $[C_{\gamma Z} = 0,\ m \sim 0.3\ \text{GeV}]$	51	[242]
				LAYCAST	$C_{\gamma\gamma}/\Lambda \sim 2 \times 10^{-3}\ \text{TeV}^{-1}$ $[C_{\gamma Z} = 0,\ m \sim 0.7\ \text{GeV}]$	51	[241]
Hidden valley particles (π_V^0)	$Z\, h(\to \pi_V^0\pi_V^0)$, $\pi_V^0 \to b\bar{b}$	350	1.0	ND	$\sigma(h) \times \text{BR}(h \to \pi_v^0\pi_v^0) \sim 10^{-4}\ \text{pb}$ $[m \in (25, 50)\ \text{GeV},\ \tau \sim 10^2\ \text{ps}]$	41	[243]
Dark photons (γ_D)	$Z(\to q\bar{q})\, h(\to \gamma_D\gamma_D)$, $\gamma_D \to \ell^-\ell^+/q\bar{q}$	250	2.0	ND	$\text{Br}(h \to \gamma_D\gamma_D) \sim 10^{-5}$, $[m \in (5, 10)\ \text{GeV},\ \tau \sim 10^2\ \text{ps},\ \epsilon \in (10^{-6}, 10^{-7})]$	42	[83]



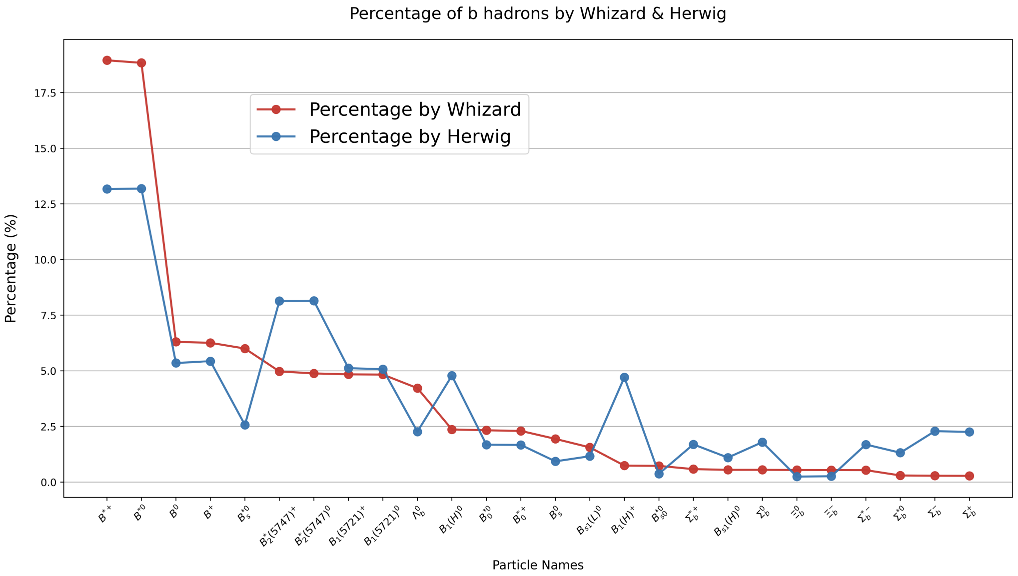


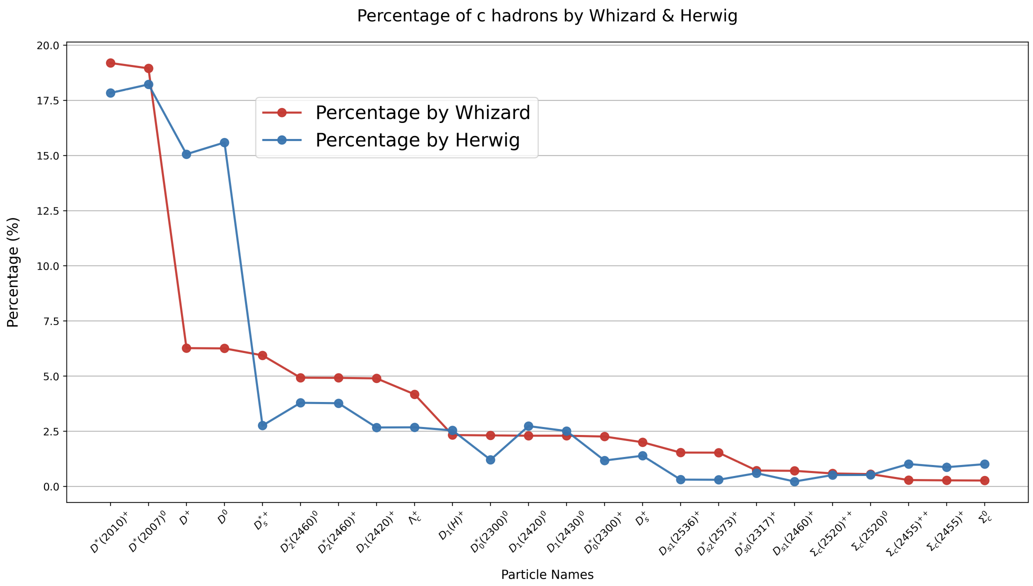


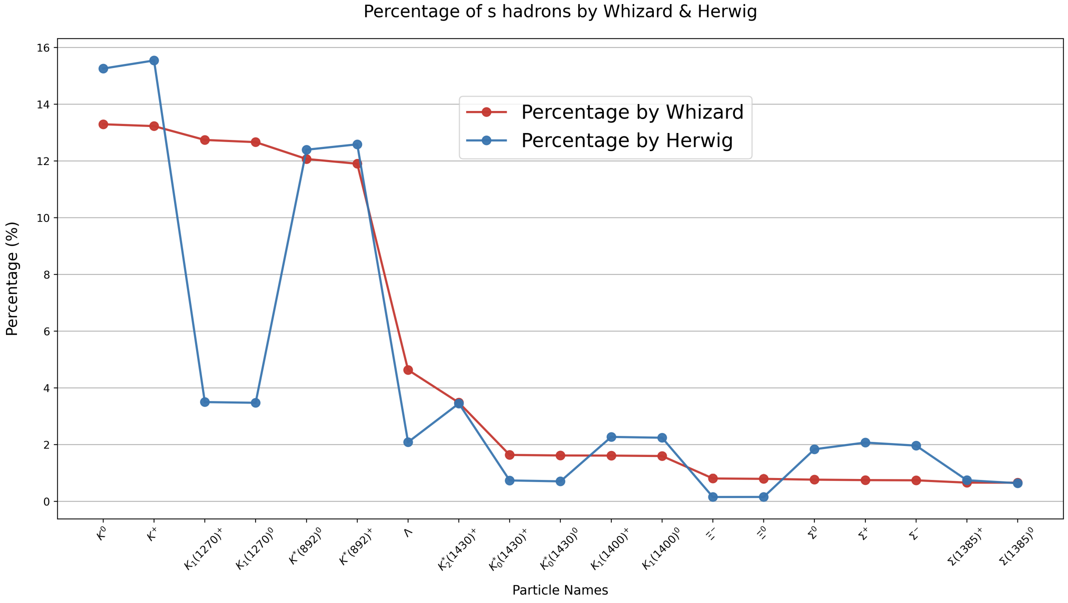
Far detector could enhance & complement the near detector (main detector) sensitivities;
While the understanding of background is the key issue.

- Credit: Kechen Wang, Yongchao Zhang, etc

Fragmentation comparison

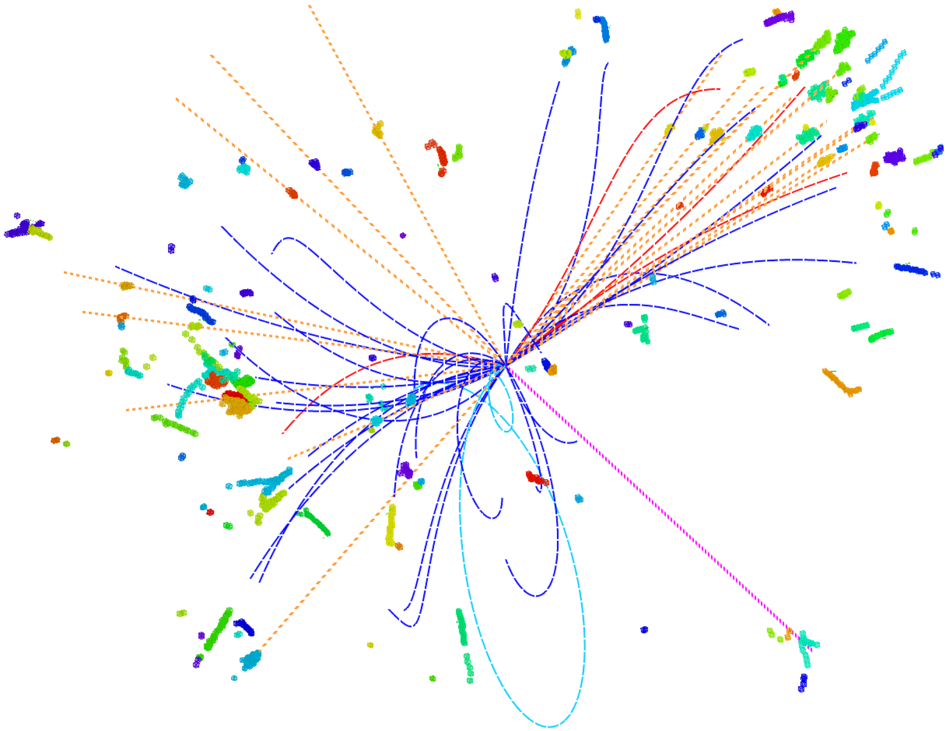

Percentage of b hadrons by Whizard & Herwig

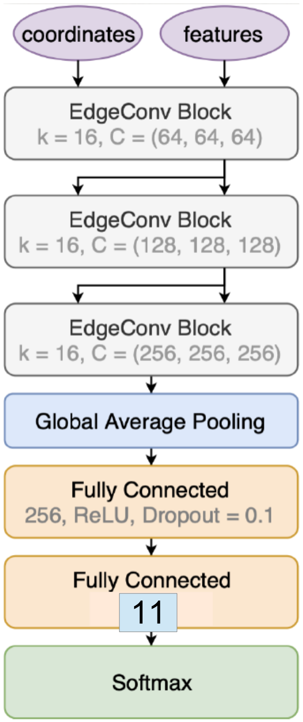

Percentage of c hadrons by Whizard & Herwig


Percentage of s hadrons by Whizard & Herwig

Lots to be studied...

Geo. & Tools





coordinates

features

EdgeConv Block
k = 16, C = (64, 64, 64)

EdgeConv Block
k = 16, C = (128, 128, 128)

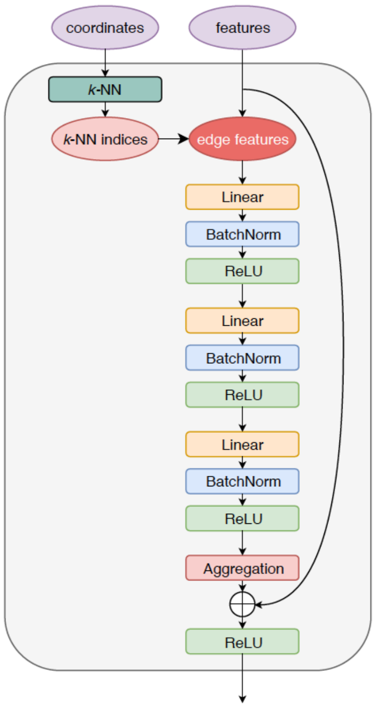
EdgeConv Block
k = 16, C = (256, 256, 256)

Global Average Pooling

Fully Connected
256, ReLU, Dropout = 0.1

Fully Connected
11

Softmax



coordinates

features

k-NN

k-NN indices

edge features

Linear

BatchNorm

ReLU

Linear

BatchNorm

ReLU

Linear

BatchNorm

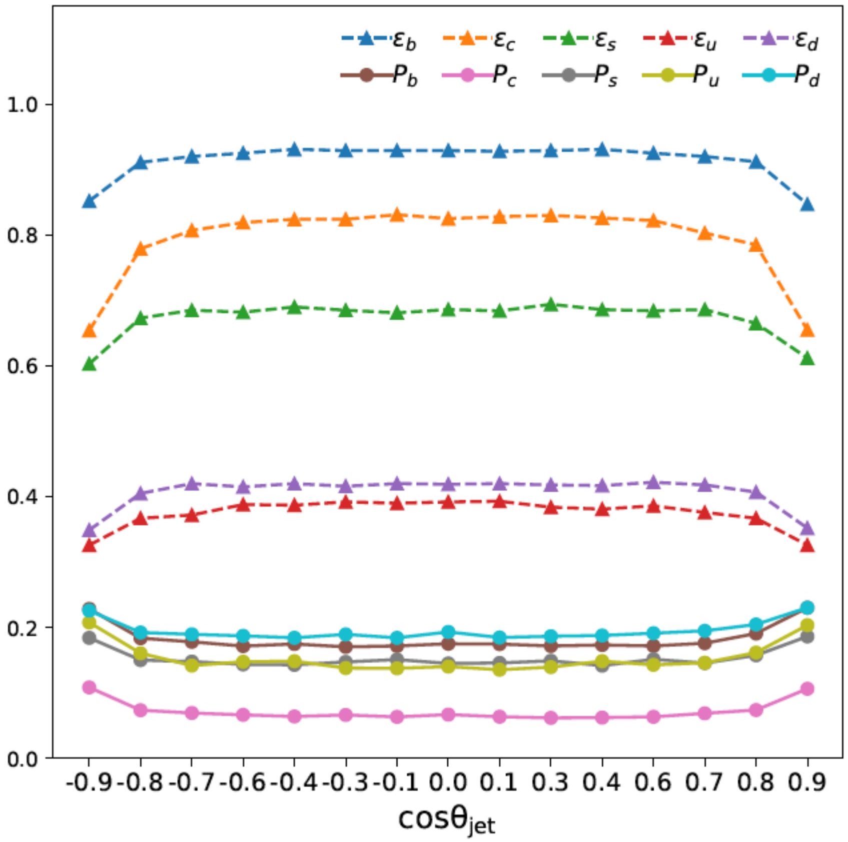
ReLU

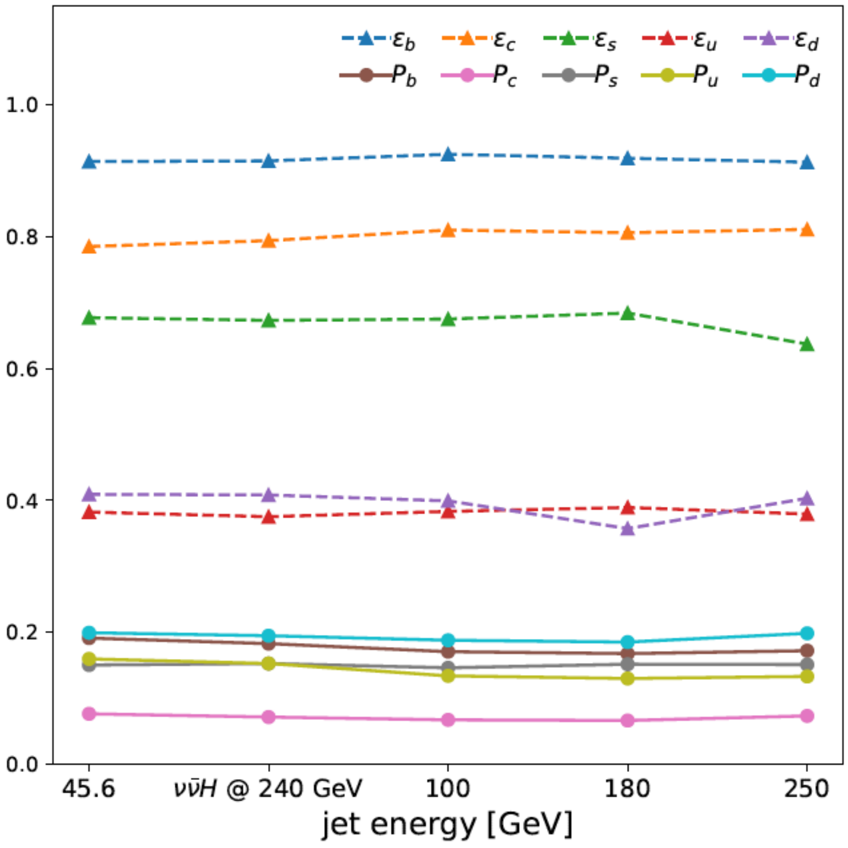
Aggregation

ReLU

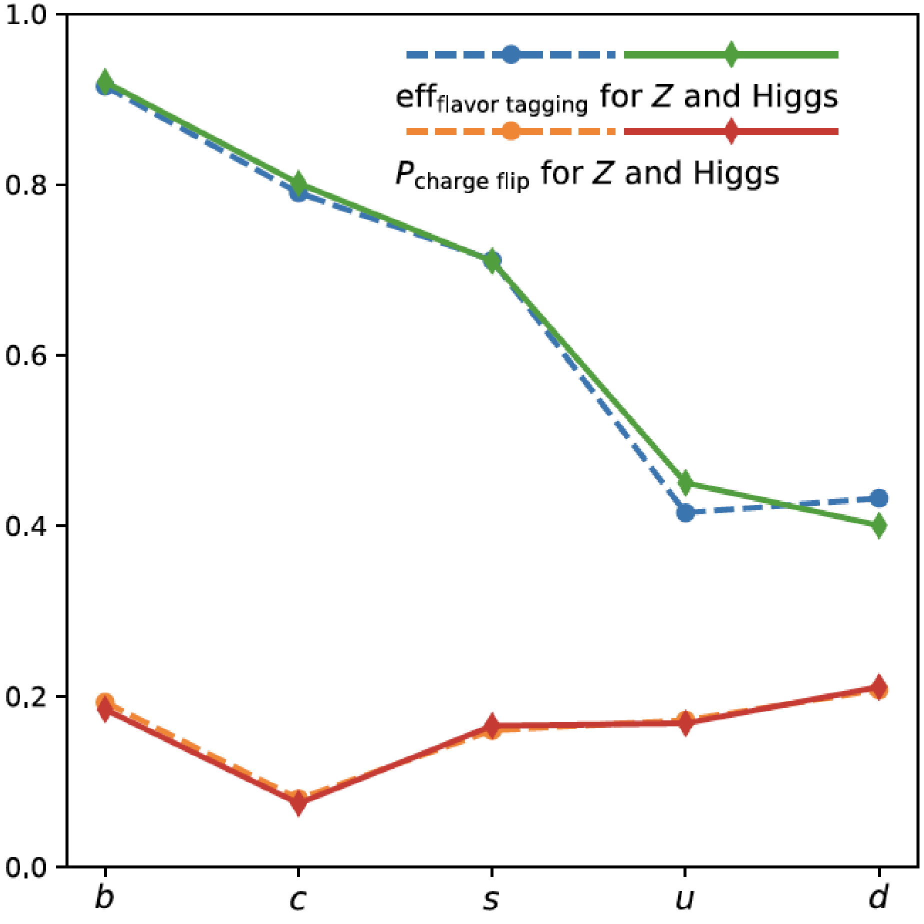
- **Jet origin identification: 11 categories (5 quarks + 5 anti quarks + gluon)**
- Full Simulated vvH, Higgs to two jets sample at CEPC baseline configuration: CEPC-v4 detector, reconstructed with **Arbor + ParticleNet (Deep Learning Tech.)**
  - Input: measurable information of all reconstructed jet particles (~ 10 float)
  - Output: 10(11)-likelihoods to different categories
- 1 Million samples each, 60/20/20% for training, validation & test

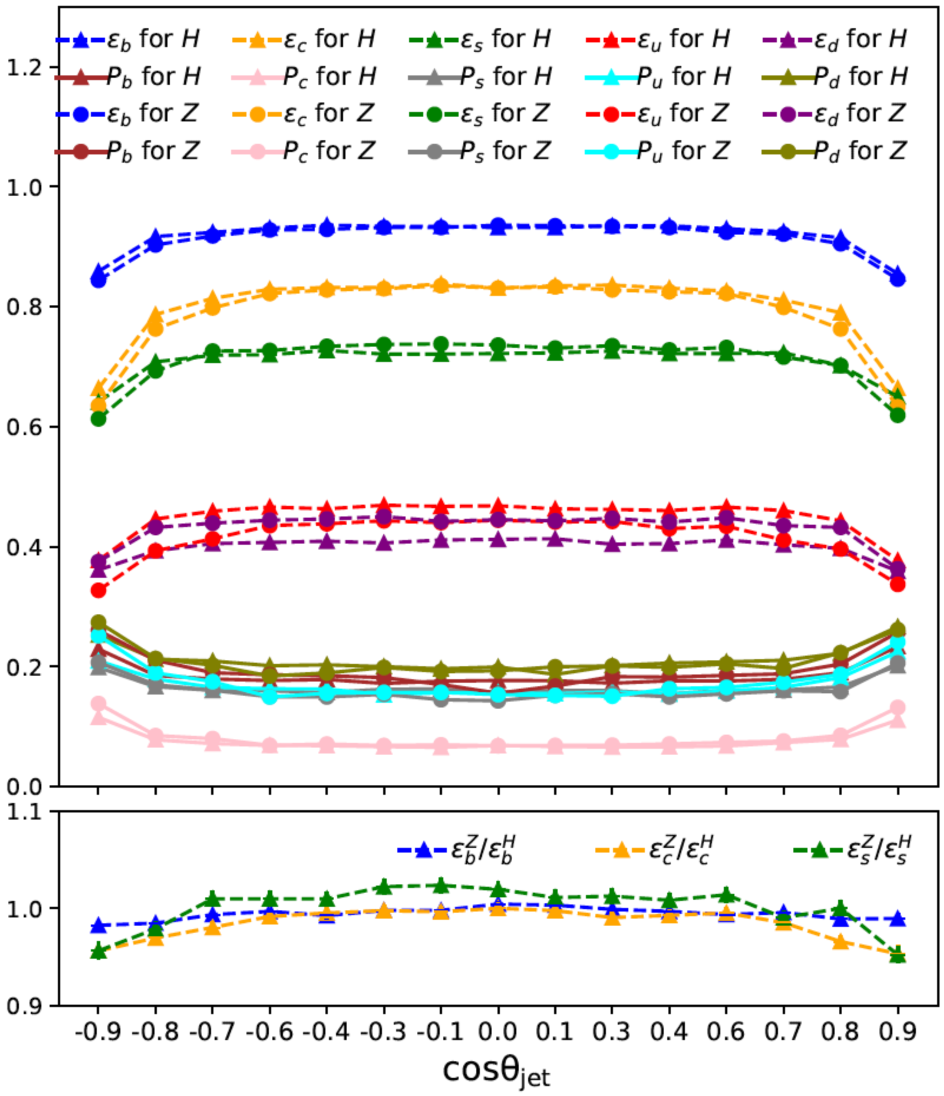
Performance V.S. Jet Kinematics





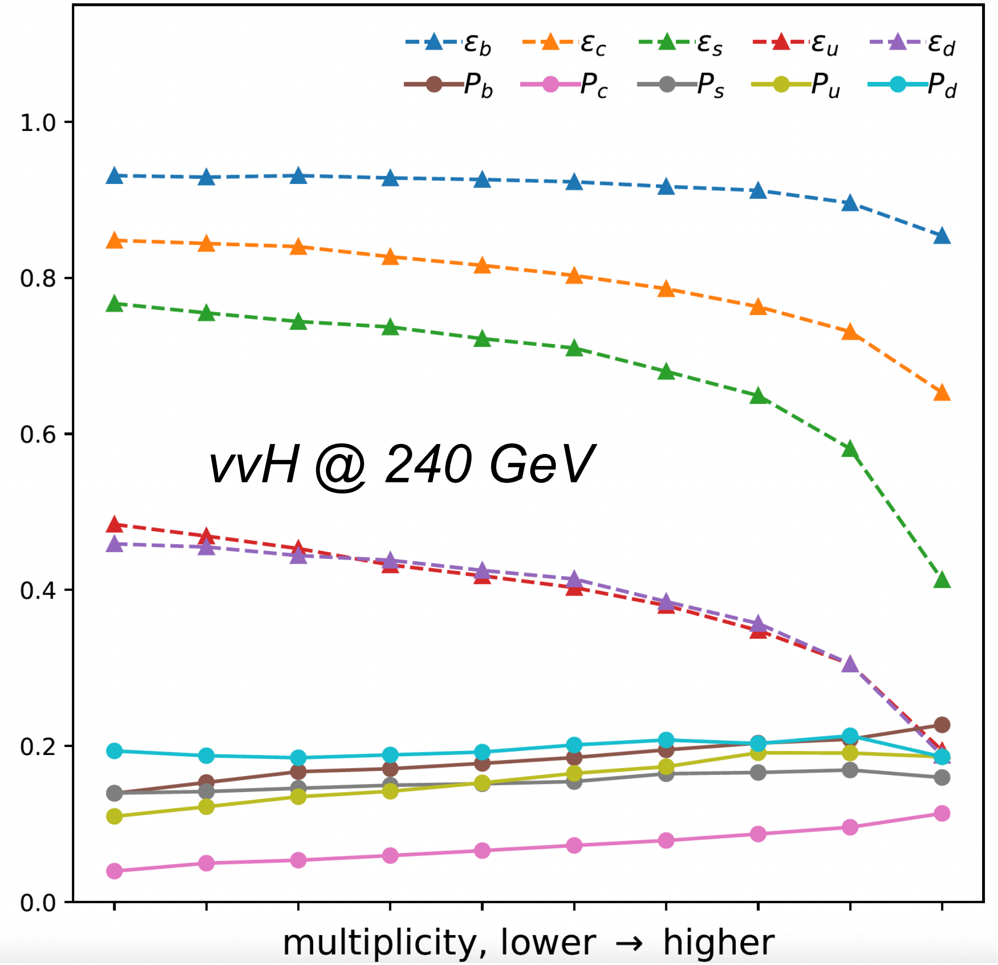
Performance @ Z and Higgs

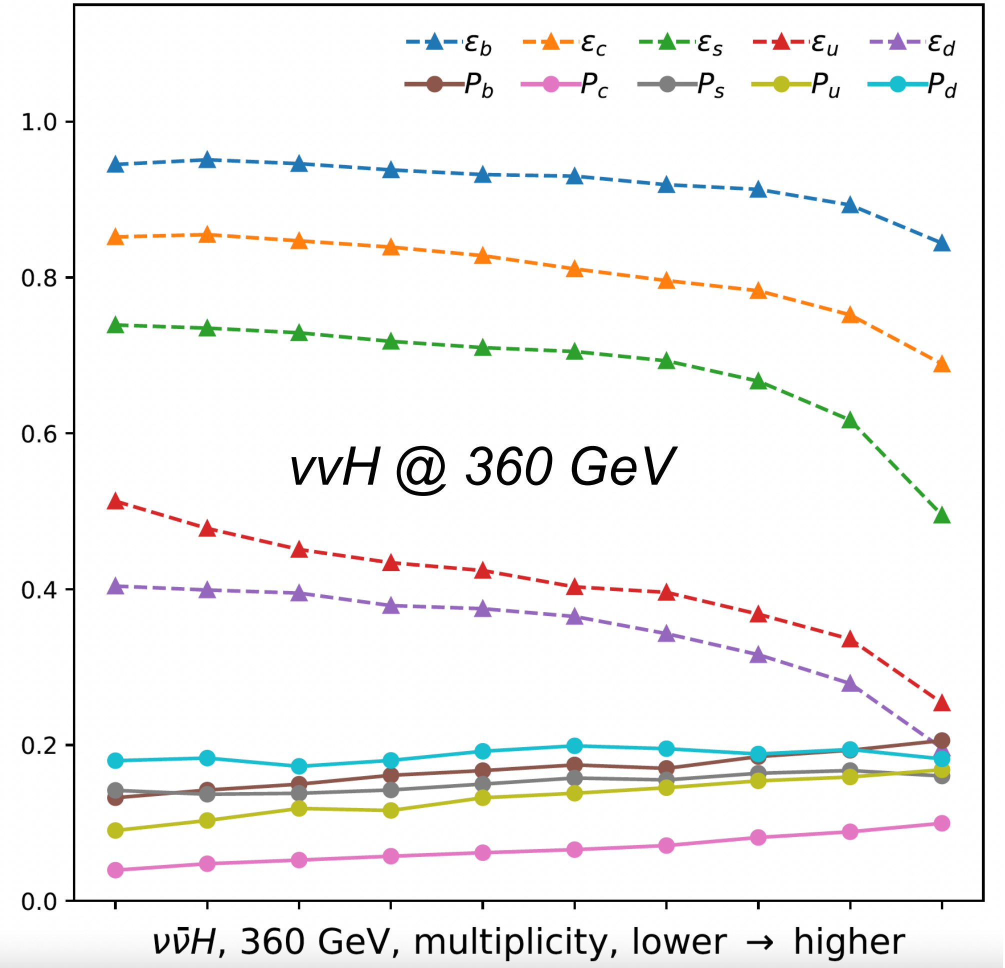




- *M10 instead of M11*

V.S. Multiplicity

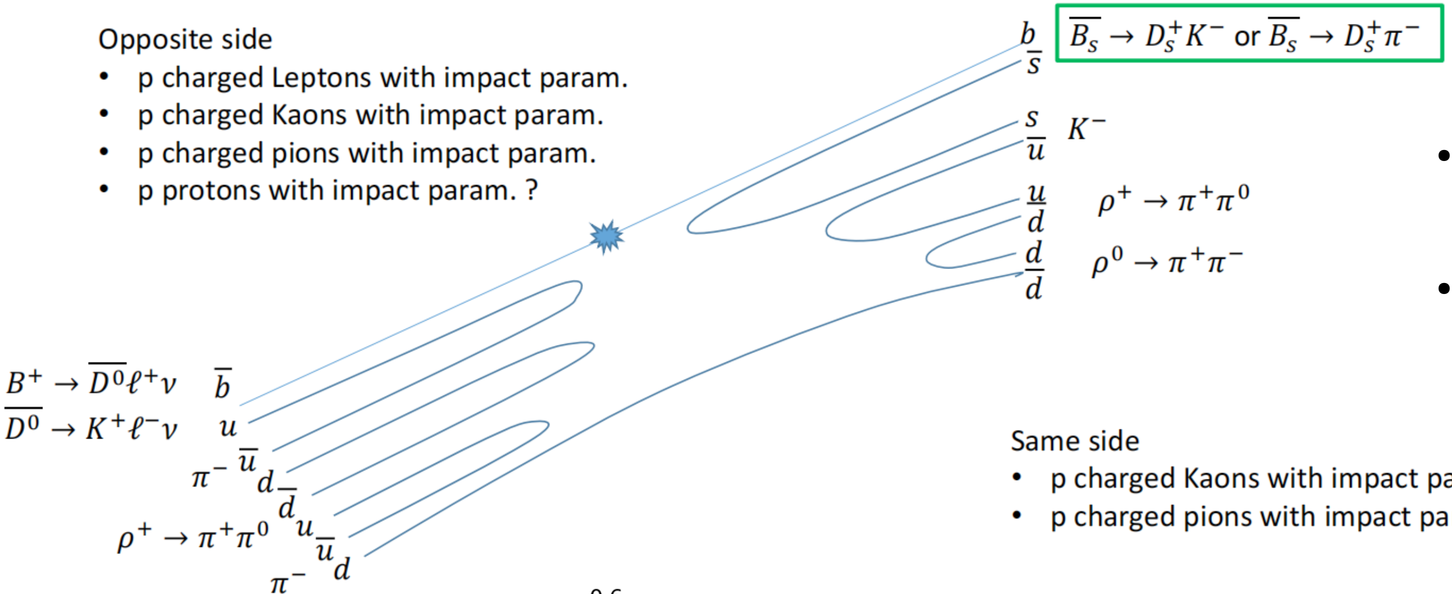




- *...many patterns need further understanding & towards further optimization...*

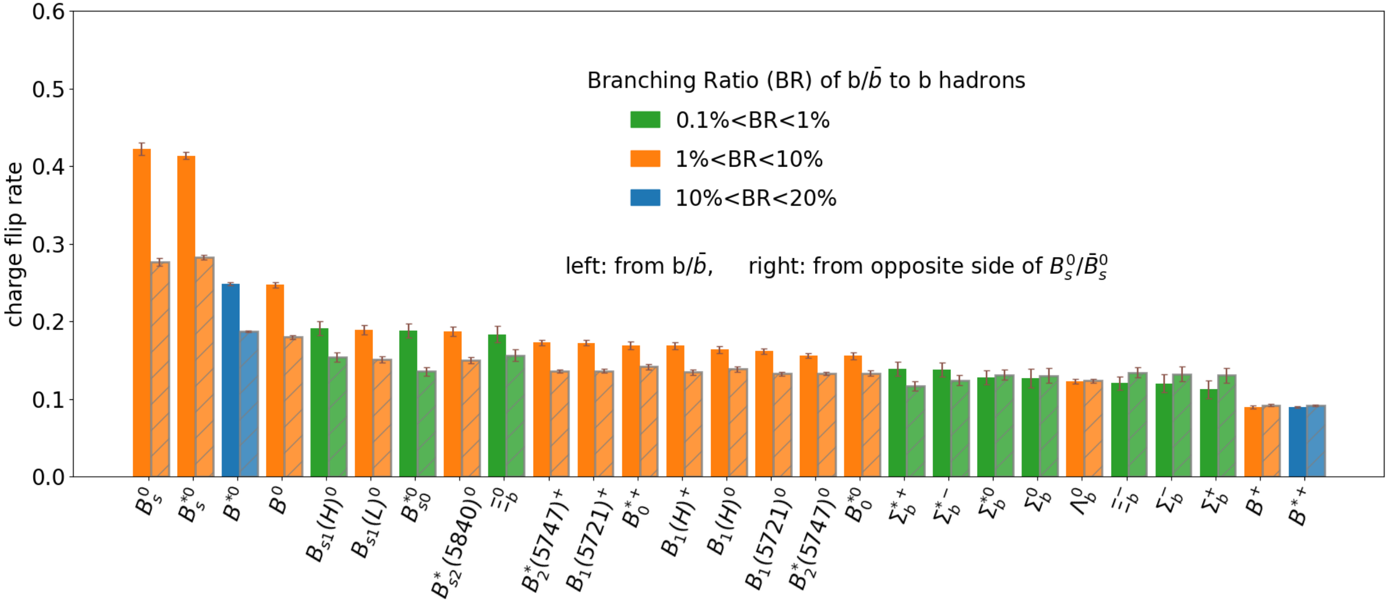
B-charge flip rate: Bs oscillations

Opposite side
- p charged Leptons with impact param.
- p charged Kaons with impact param.
- p charged pions with impact param.
- p protons with impact param. ?

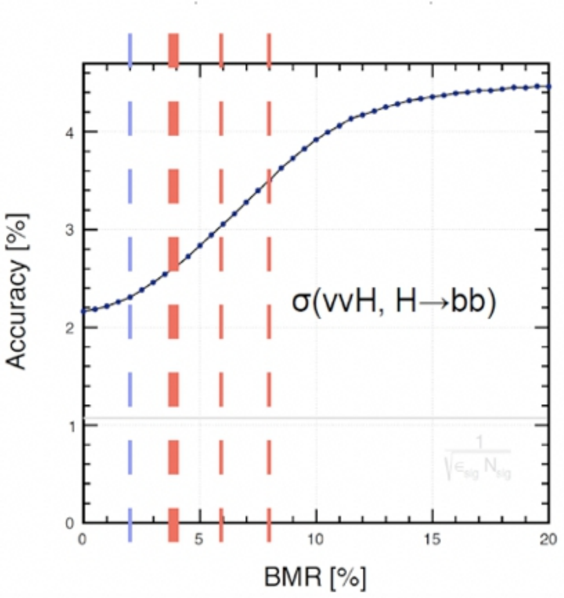


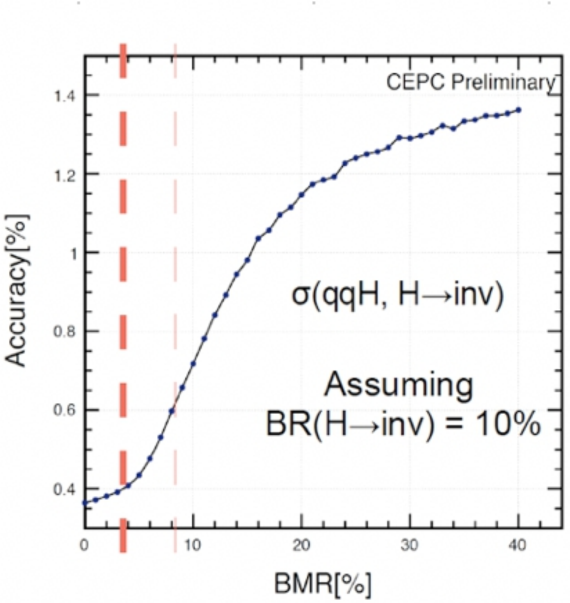
- Using all reco P (exc. Bs decay final state):
- Flip rate ~ 15%, Eff. Tagging power > 40%

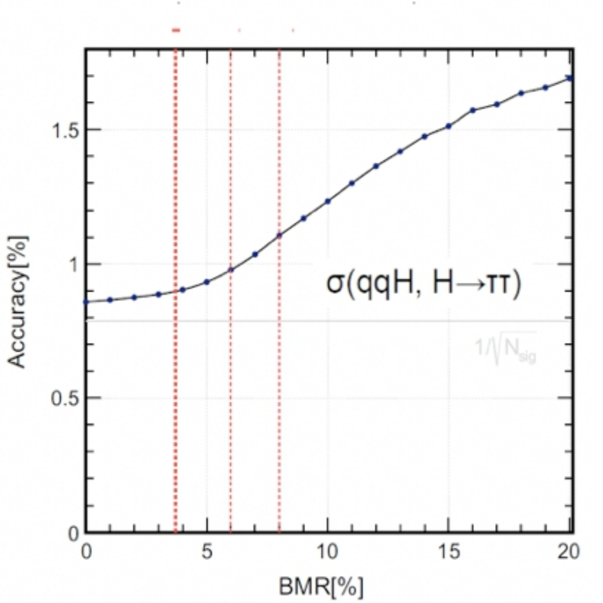
Same side
- p charged Kaons with impact param.
- p charged pions with impact param.

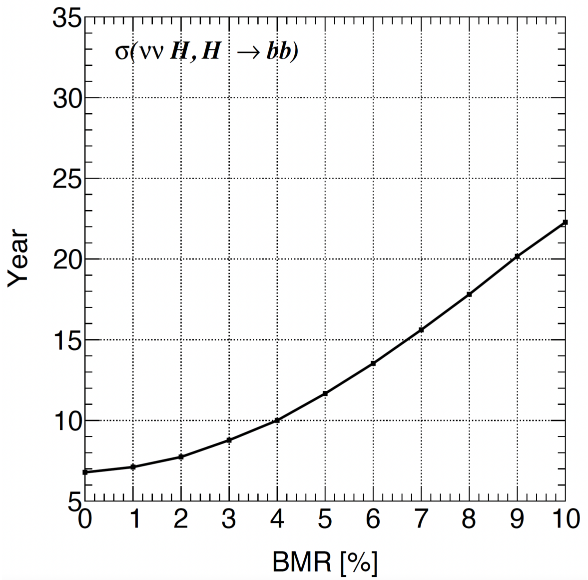


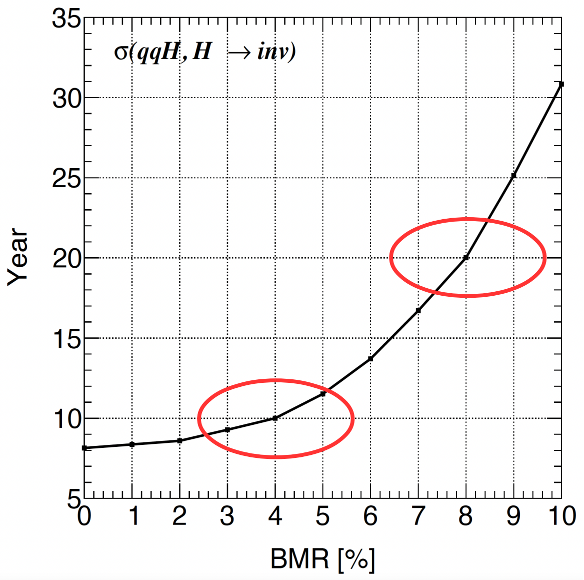
BMR: impact on critical measurements

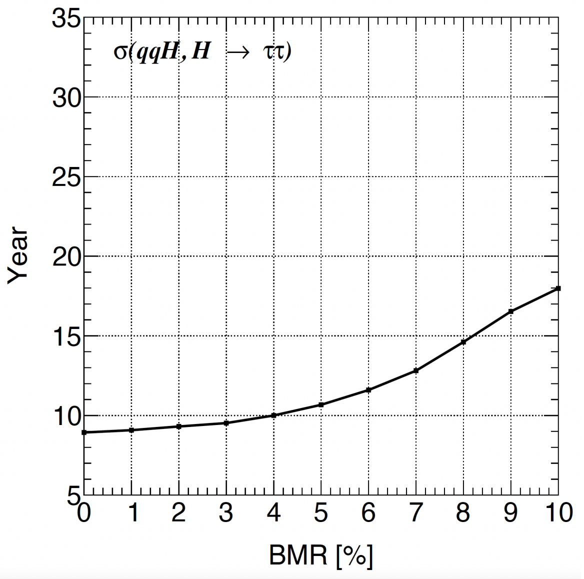








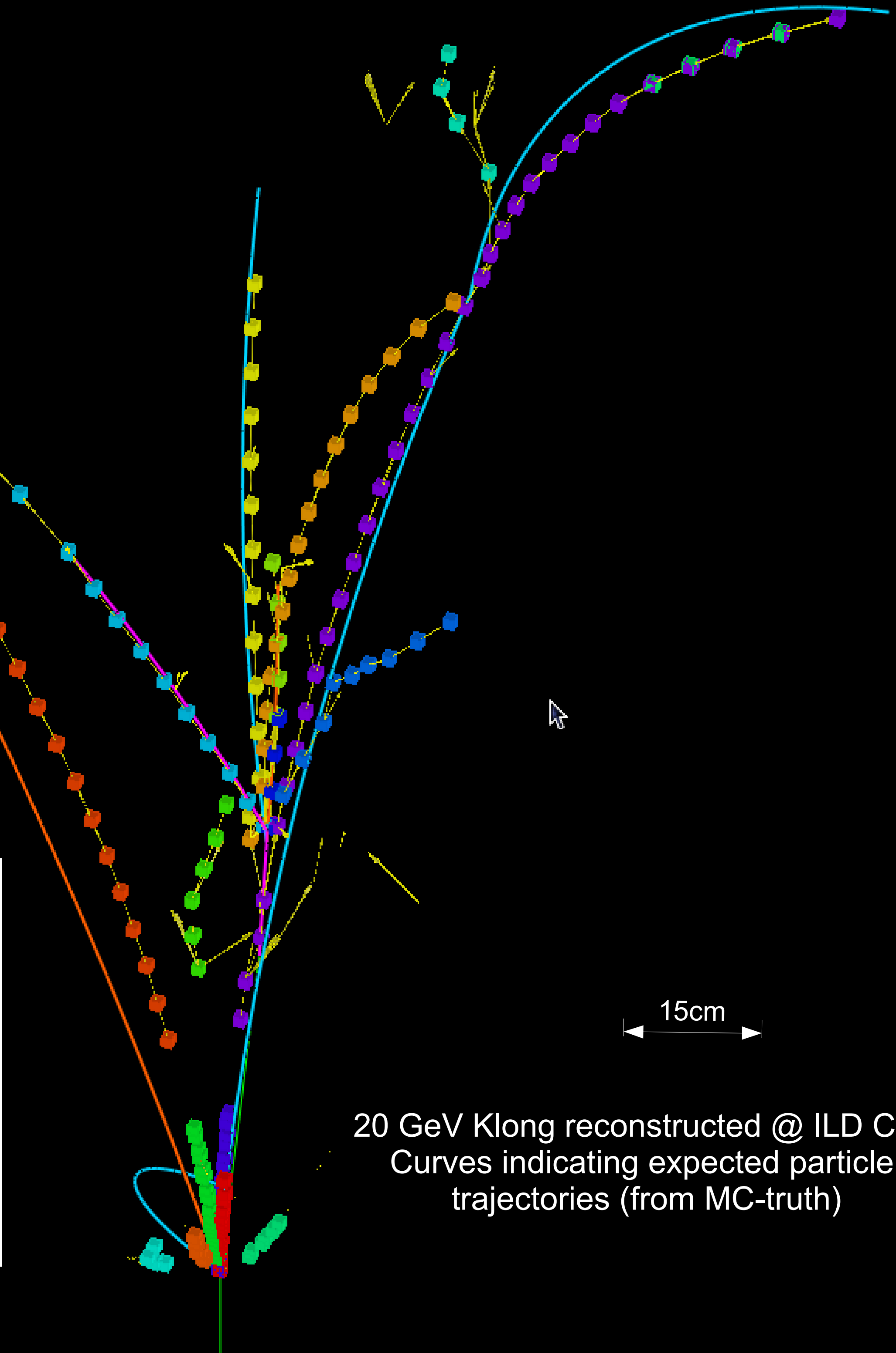




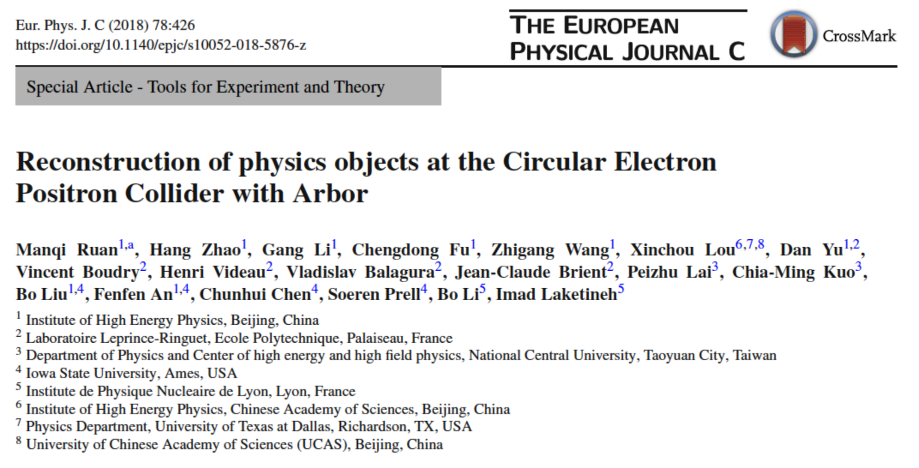
Arbor

Tree topology of particle shower

Ori. Idea from Henri Videau @ ALEPH



20 GeV Klong reconstructed @ ILD Calo
Curves indicating expected particle
trajectories (from MC-truth)



Eur. Phys. J. C (2018) 78:426
https://doi.org/10.1140/epjc/s10052-018-5876-z

THE EUROPEAN
PHYSICAL JOURNAL C

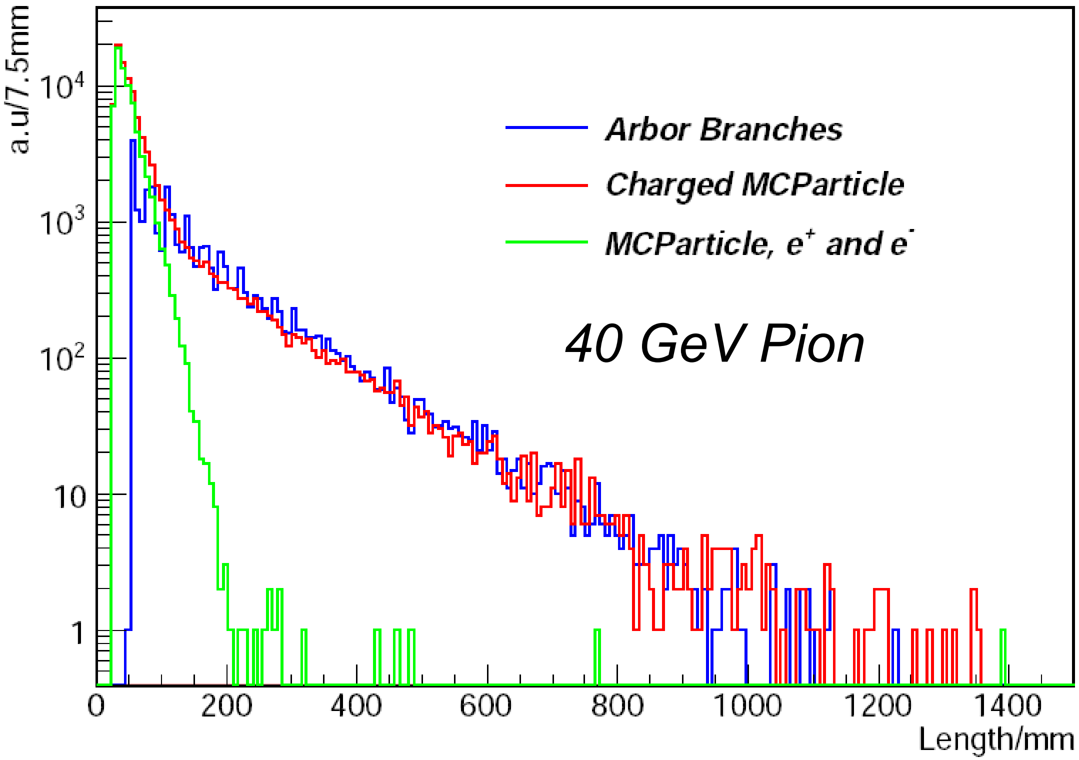
Special Article - Tools for Experiment and Theory

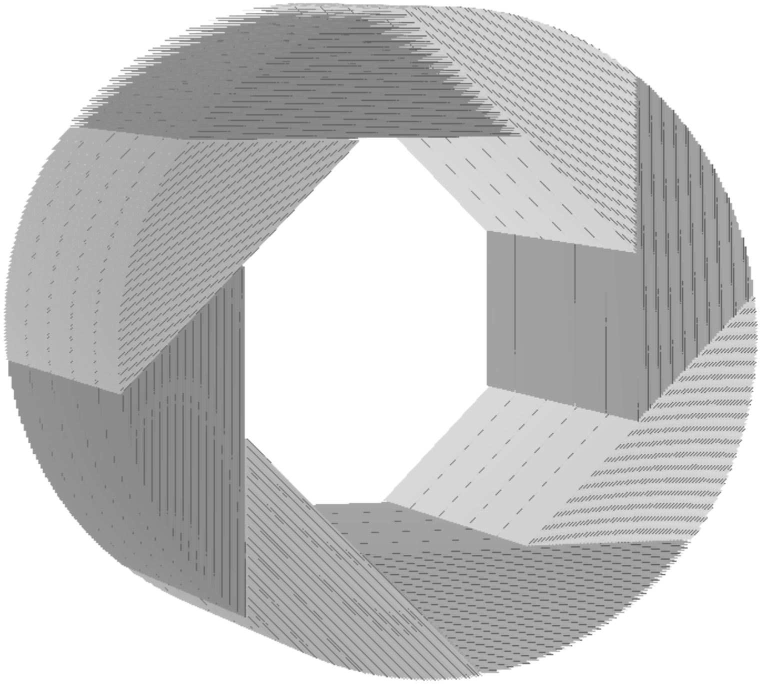
Reconstruction of physics objects at the Circular Electron Positron Collider with Arbor

Manqi Ruan[1,a], **Hang Zhao**[1], **Gang Li**[1], **Chengdong Fu**[1], **Zhigang Wang**[1], **Xinchou Lou**[6,7,8], **Dan Yu**[1,2], **Vincent Boudry**[2], **Henri Videau**[2], **Vladislav Balagura**[2], **Jean-Claude Brient**[2], **Peizhu Lai**[3], **Chia-Ming Kuo**[3], **Bo Liu**[1,4], **Fenfen An**[1,4], **Chunhui Chen**[4], **Soeren Prell**[4], **Bo Li**[5], **Imad Laketineh**[5]

[1] Institute of High Energy Physics, Beijing, China
[2] Laboratoire Leprince-Ringuet, Ecole Polytechnique, Palaiseau, France
[3] Department of Physics and Center of high energy and high field physics, National Central University, Taoyuan City, Taiwan
[4] Iowa State University, Ames, USA
[5] Institute de Physique Nucleaire de Lyon, Lyon, France
[6] Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China
[7] Physics Department, University of Texas at Dallas, Richardson, TX, USA
[8] University of Chinese Academy of Sciences (UCAS), Beijing, China

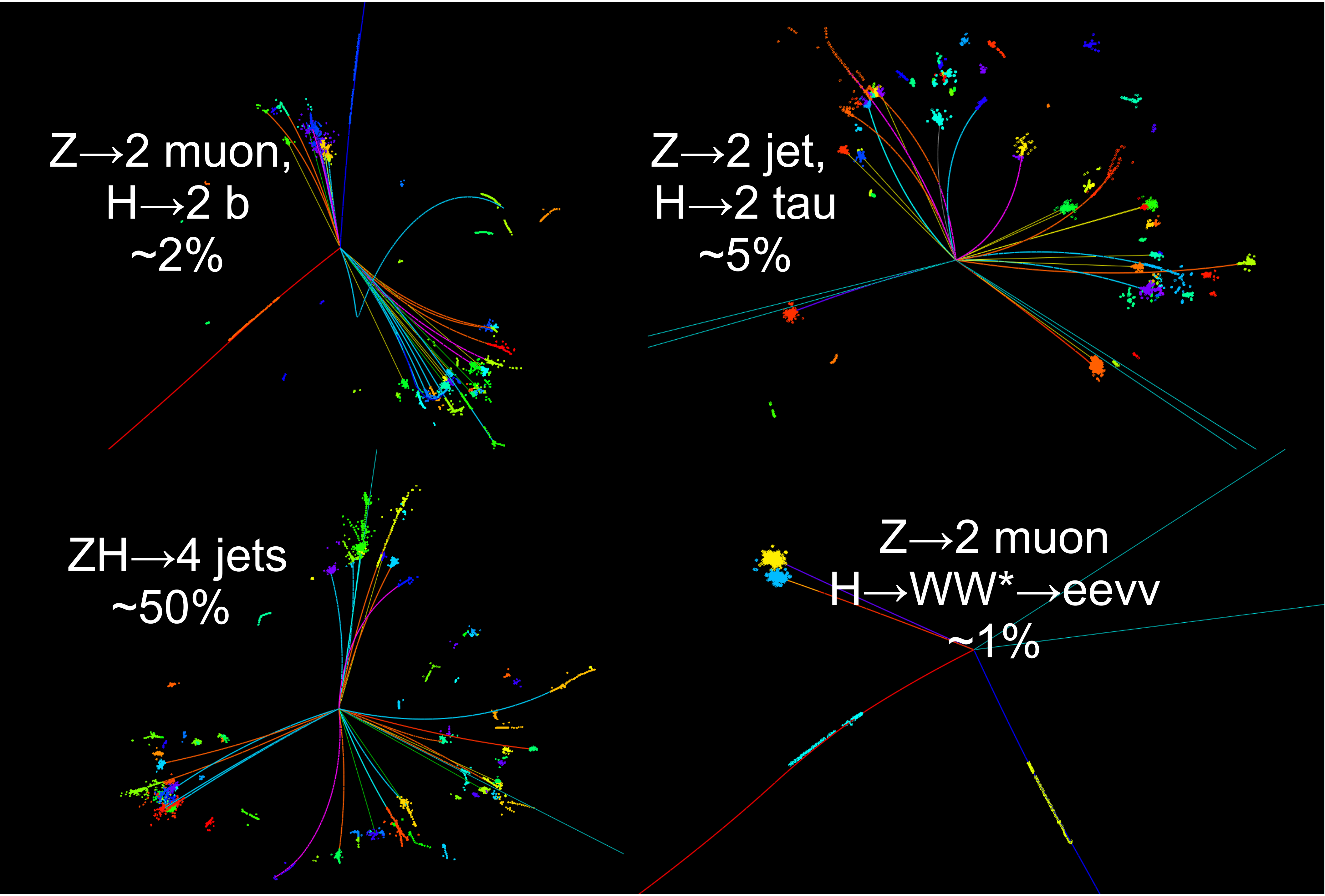
Validation: Arbor Branch Length Vs MC Truth





Arbor: successfully tag sub-shower structure

Samples: Particle gun event at ILD HCAL (readout granularity 1cm^2 & layer thickness 2.65cm)
Length:
Charged MCParticle: spatial distance between generation/end points
Arbor branch: sum of distance between neighboring cells

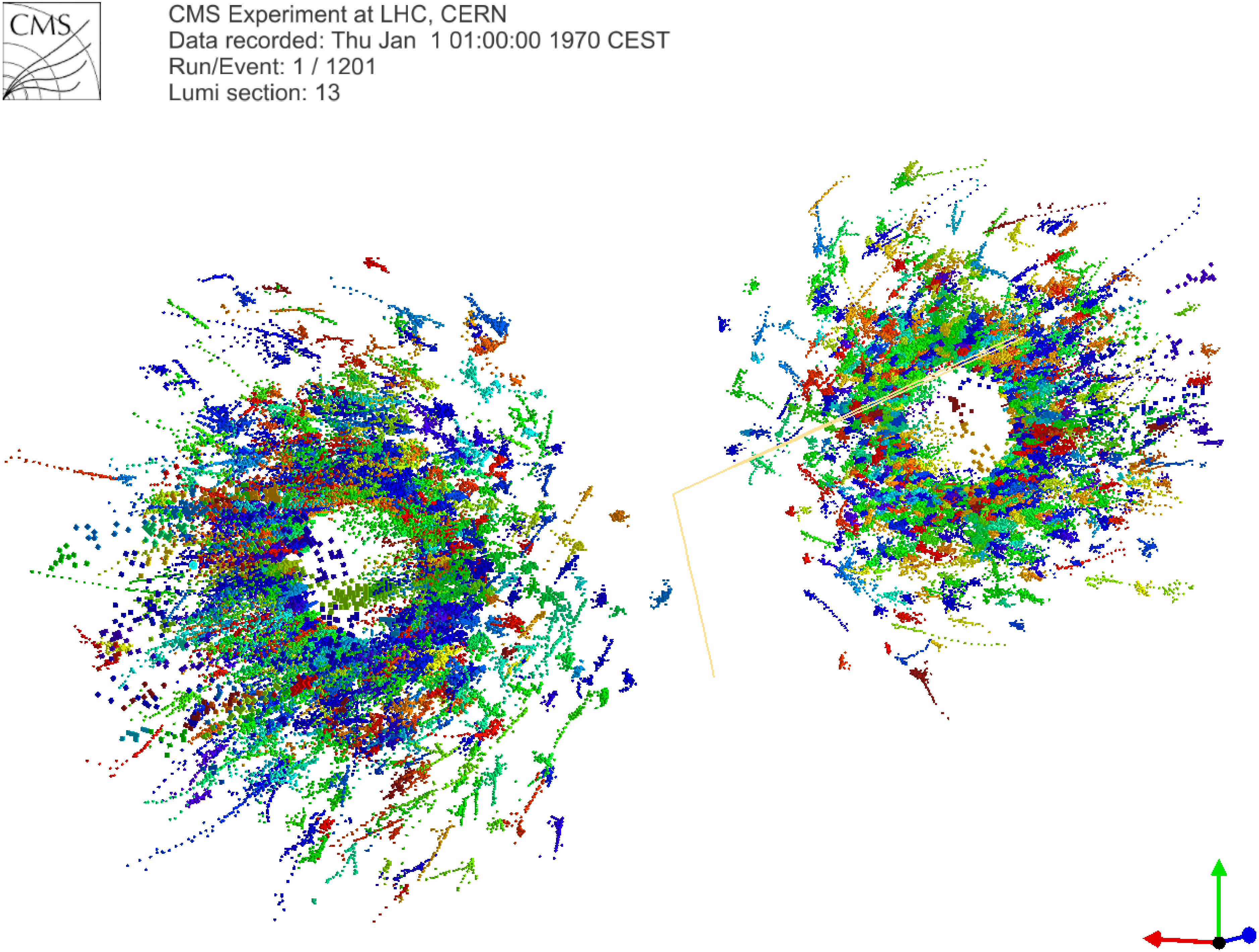


Z→2 muon,
H→2 b
~2%

Z→2 jet,
H→2 tau
~5%

ZH→4 jets
~50%

Z→2 muon
H→WW*→eevv
~1%


CMS
CMS Experiment at LHC, CERN
Data recorded: Thu Jan 1 01:00:00 1970 CEST
Run/Event: 1 / 1201
Lumi section: 13

s-jets: dependency on Leading hadron