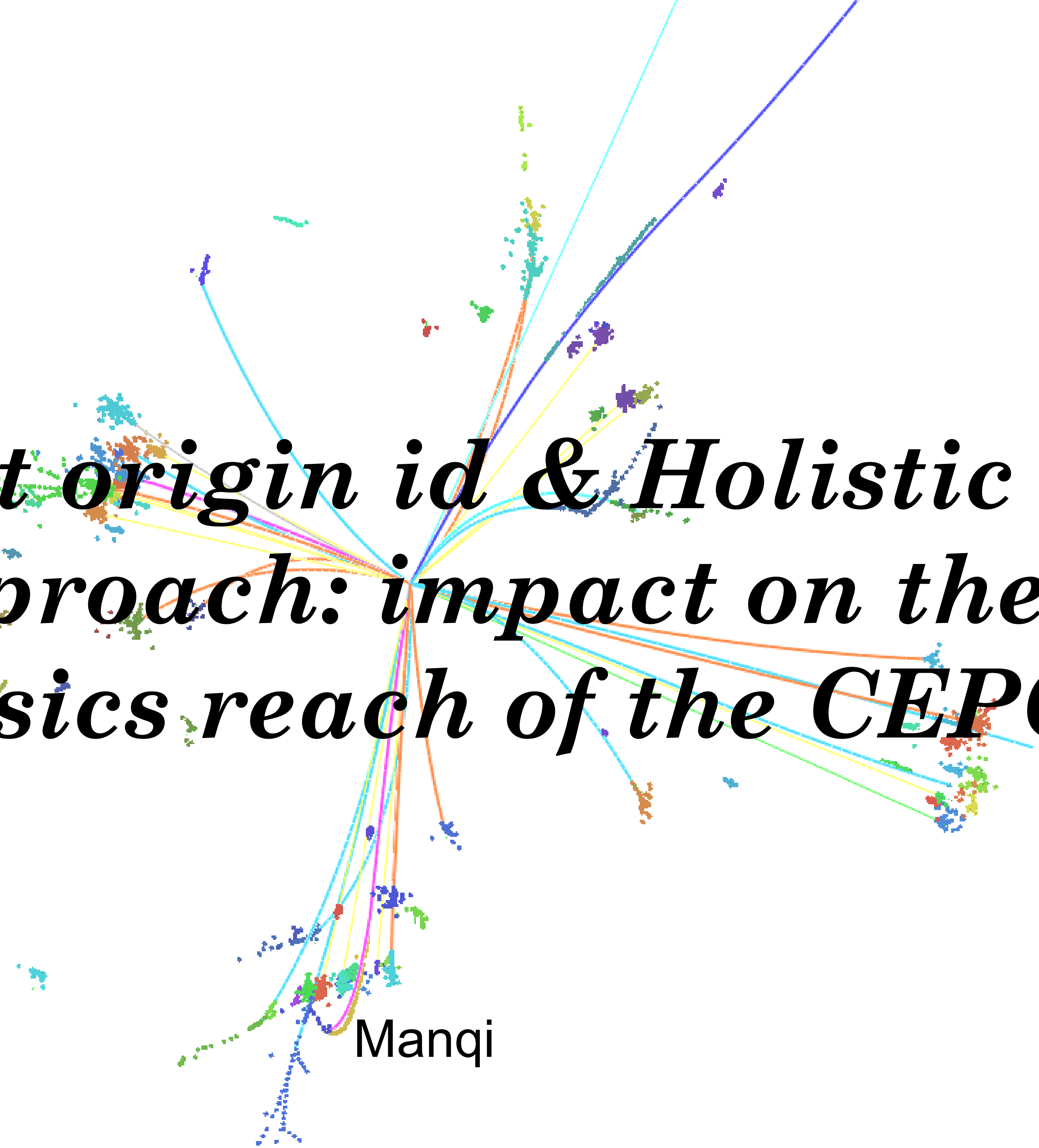
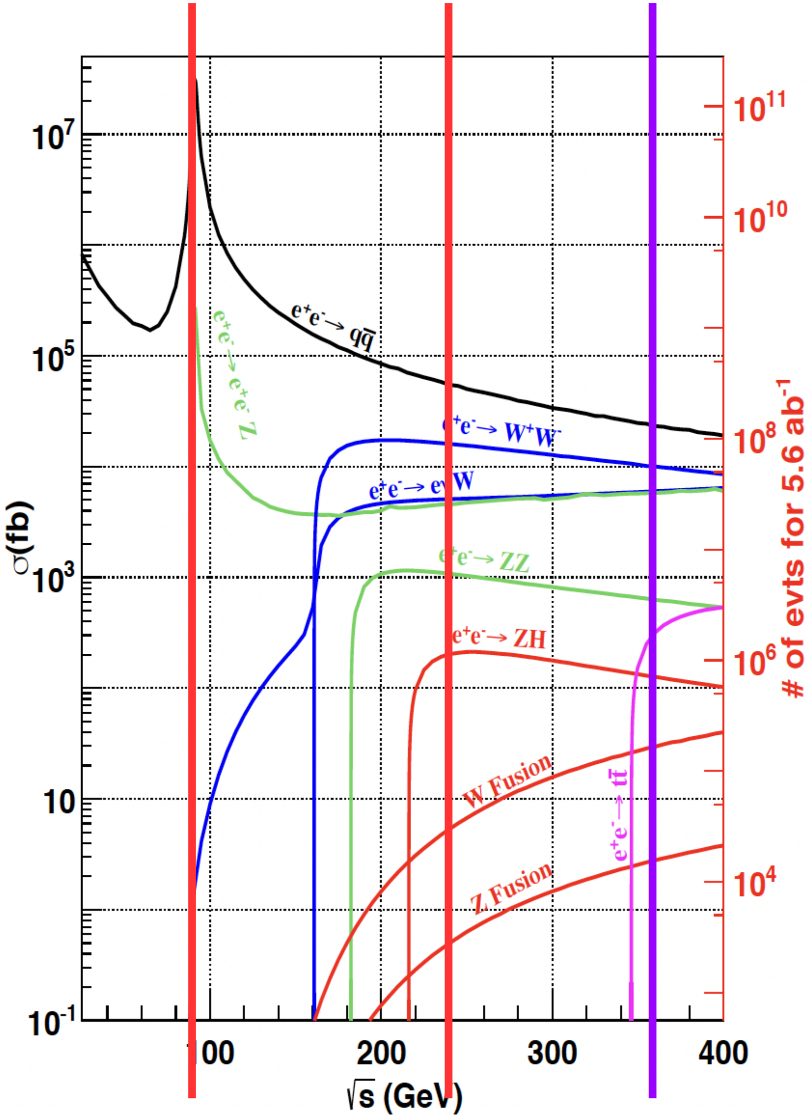
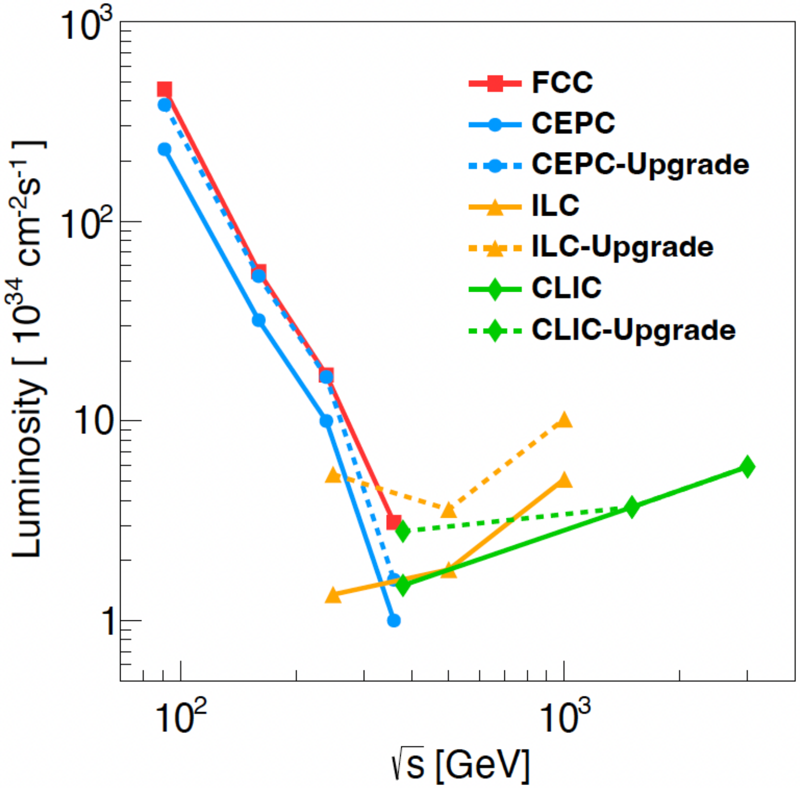


Jet origin id & Holistic approach: impact on the Physics reach of the CEPC

Manqi

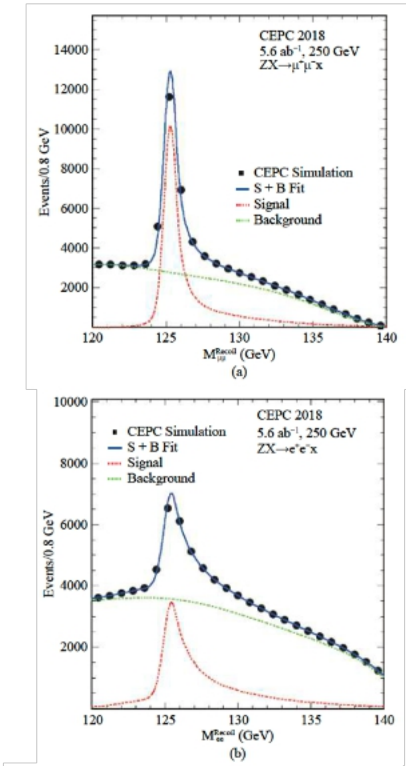
Yields ~ Xsec * Lumi * Time

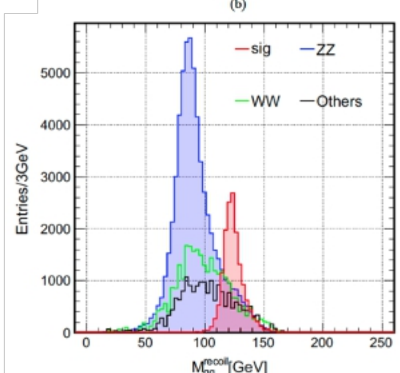


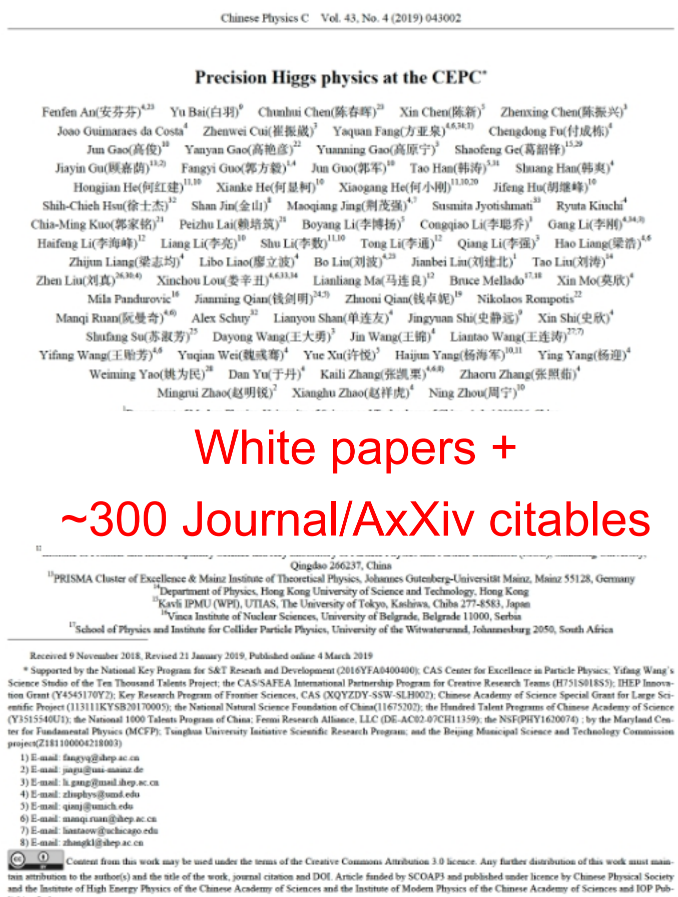


- CEPC: 100 km main ring circumference
- 4 Million Higgs (10 years)
- ~ 1 Giga W (1 year) + 4 Tera Z (2 years)
- Upgradable: Top factory (500 k ttbar)

CEPC Physics study






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

Precision Higgs physics at the CEPC*

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

White papers +

~300 Journal/AxXiv citables

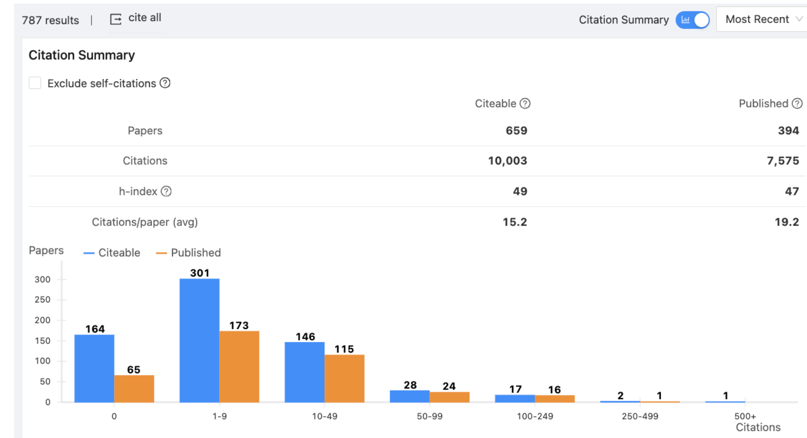


Table 2.1: Precision of the main parameters of interests and observables at the CEPC, from Ref. [1] and the references therein, where the results of Higgs are estimated with a data sample of 20 ab^{-1}. The HL-LHC projections of 3000 fb^{-1} data are used for comparison. [2]

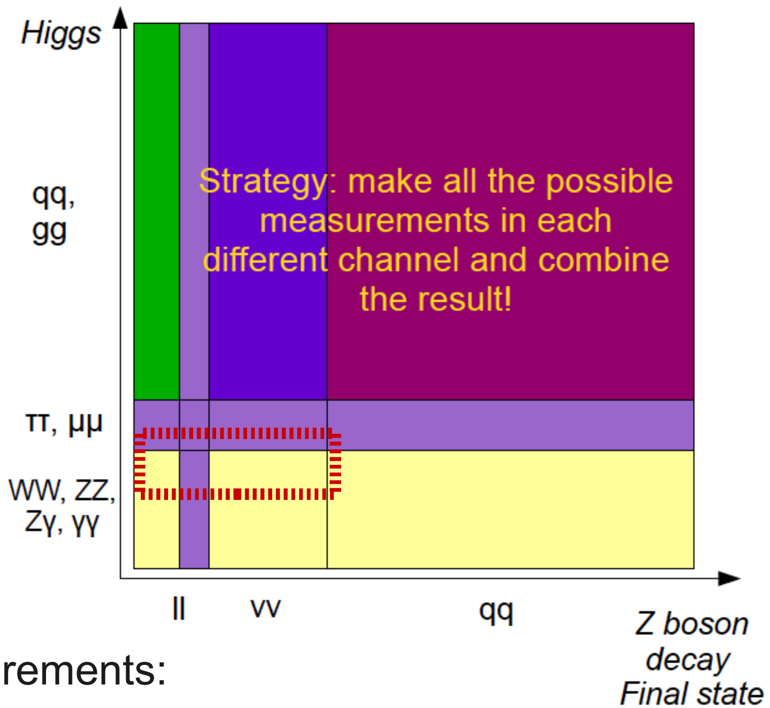
Higgs			W, Z and top		
Observable	HL-LHC projections	CEPC precision	Observable	Current precision	CEPC precision
M_H	20 MeV	3 MeV	M_W	9 MeV	0.5 MeV
Γ_H	20%	1.7%	Γ_W	49 MeV	2 MeV
$\sigma(ZH)$	4.2%	0.26%	M_{top}	760 MeV	$\mathcal{O}(10)$ MeV
$B(H \to bb)$	4.4%	0.14%	M_Z	2.1 MeV	0.1 MeV
$B(H \to cc)$	-	2.0%	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.
- …

Performance requirements

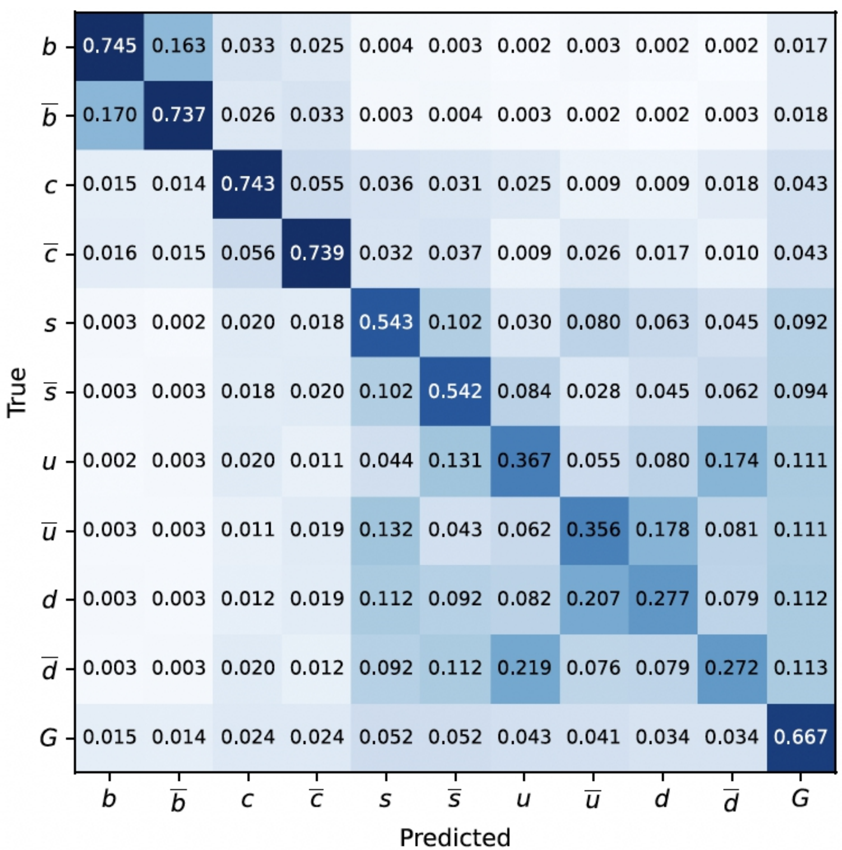
- To reconstruct all kinds of Physics Object
  - Identification & Measurements
  - Objects:
    - Lepton, Photons, Kaon,
    - pi-0, Tau, Lambda, Kshort,
    - Heavy flavor hadrons,
    - **Jets**
    - Missing energy/momentum
    - Exotics...
- Massive Four in Standard Model:
  - Z & W: ~ 70% goes to a pair of jets
  - Higgs: ~90% final state with jets (ZH events)
  - Top: t → W + b

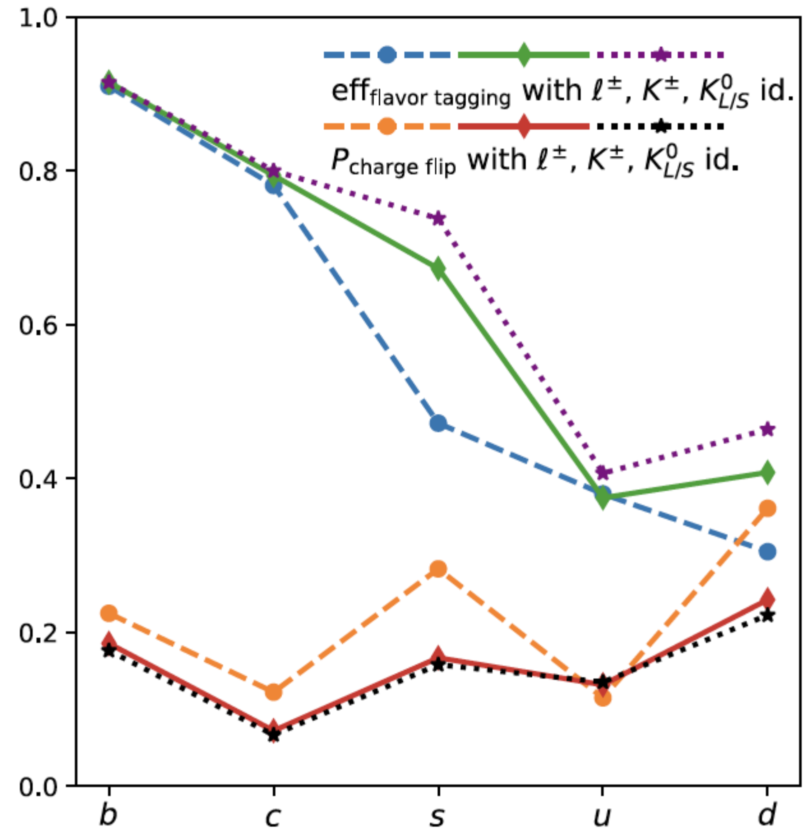


- Requirements:
  - **1-1 correspondence**

    Excellent pattern. Reco. & Object id
  - Larger acceptance, Excellent intrinsic resolutions, Extremely stable...
- Be addressed by detector design, technology, and reconstruction algorithm

Holistic Reco: Jet origin id

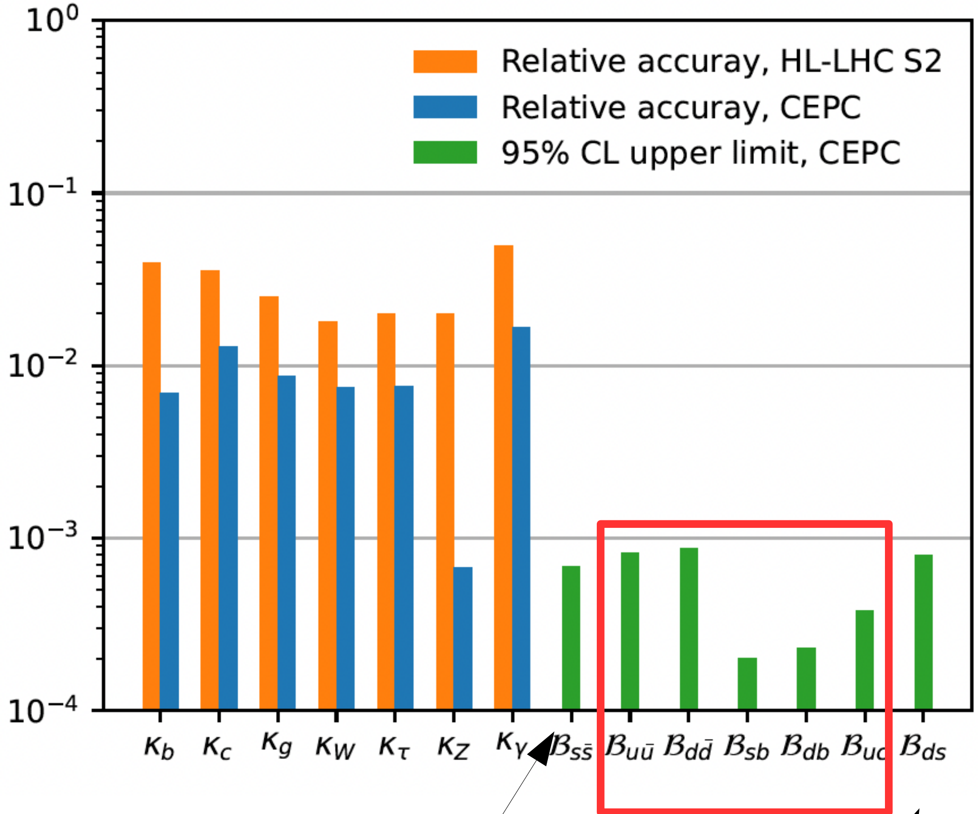


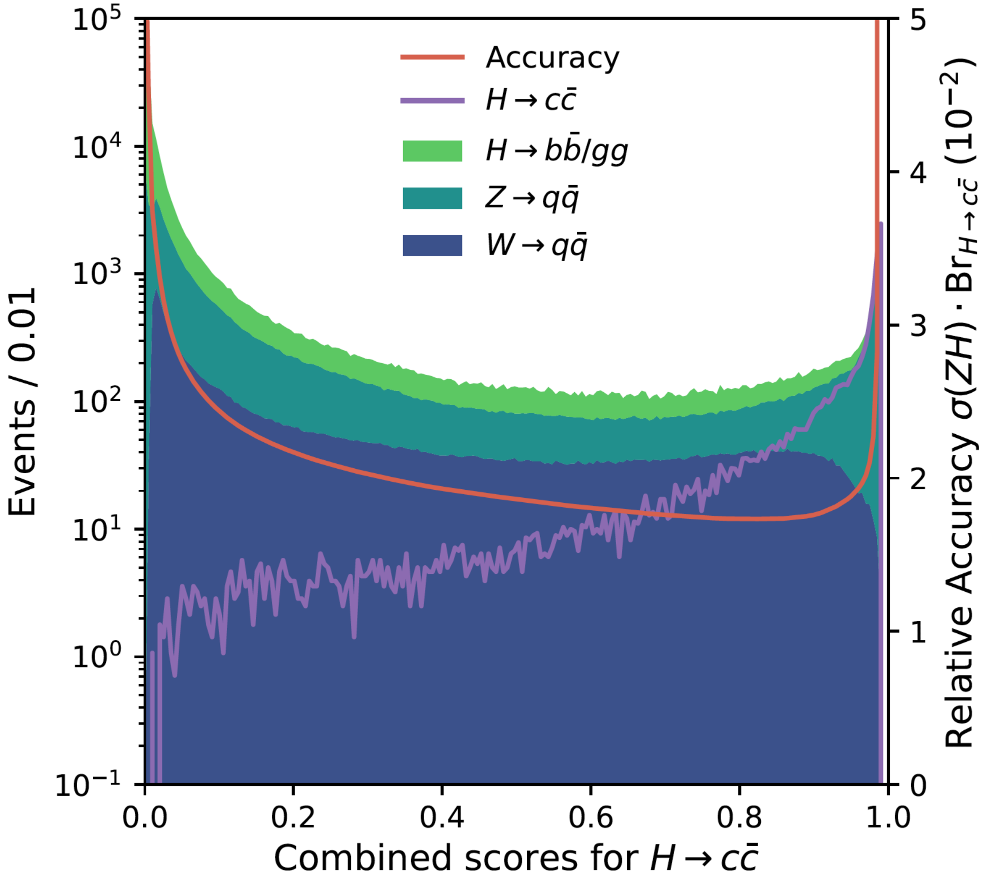


- 11 categories (5 quarks + 5 anti quarks + gluon) identification, realized at Full Simulated di-jet events at CEPC CDR baseline with Arbor + ParticleNet
- Published in PRL *132, 221802 (2024)*. Comment from the referee: *"demonstrate the world-leading performance of tagger", "a "game changer" and opens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

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

Impact on Physics: Higgs & W





Improved by ~3 times

Improved by 1-2 orders of magnitudes

Presumably... firstly quantified

- Compared to Conventional :
  - vvH, H→cc: 3% → 1.7%
  - Vcb: 0.75% → 0.5%
  - Applicable to Vcs, Vts, etc.

Updated result on $\sin^2\theta^l_{eff}$ measurement

Table 2. Sensitivity S of different final state particles.

$\sqrt{s}$/GeV	S of $A^{e/\mu}_{FB}$	S of A^{d}_{FB}	S of A^{u}_{FB}	S of A^{s}_{FB}	S of A^{c}_{FB}	S of A^{b}_{FB}
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

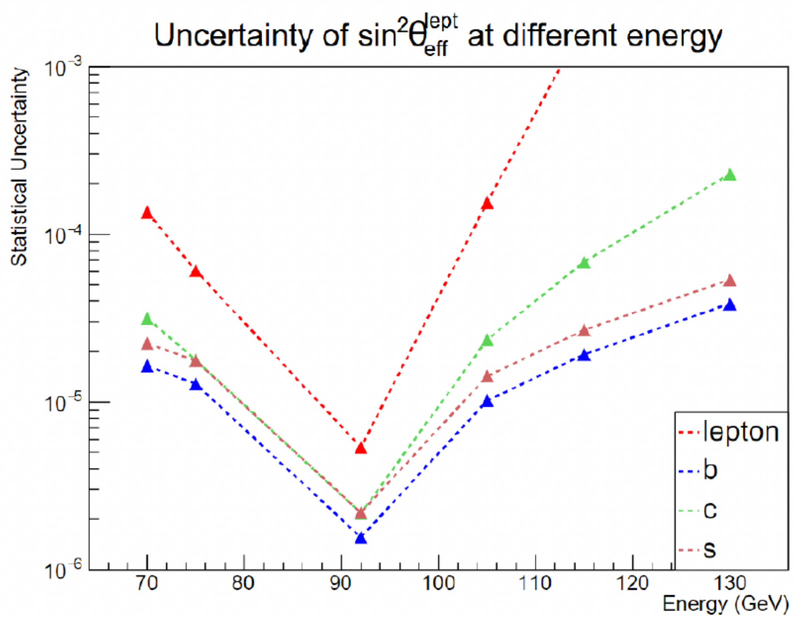
Table 3. Cross section of process $e^+e^- \to f\bar{f}$ calculated using the ZFITTER package. Values of the fundamental parameters are set as $m_Z = 91.1875$ GeV, $m_t = 173.2$ GeV, $m_H = 125$ GeV, $\alpha_s = 0.118$ and $m_W = 80.38$ GeV.

$\sqrt{s}$/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 $\sin^2\theta^l_{eff}$ measurement.

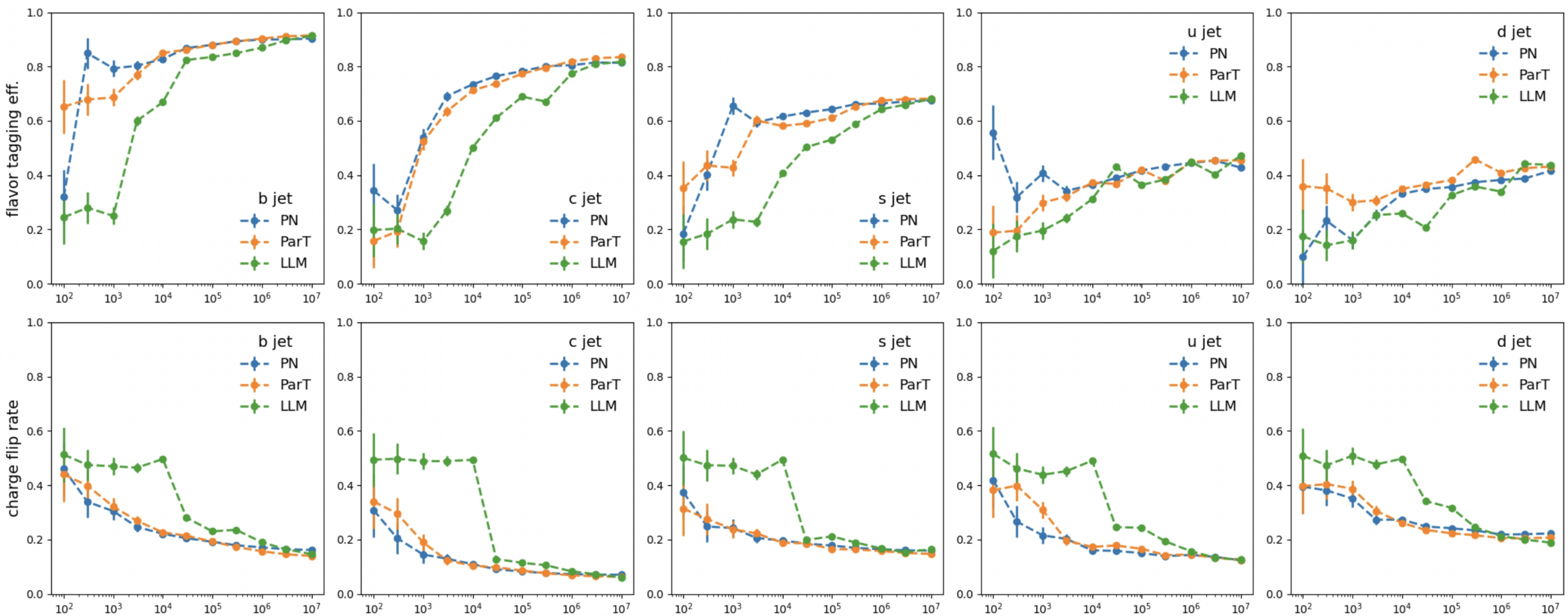
(Using one-month data collection, ~ *4e12/24 Z events* at Z pole)



$\sqrt{s}$	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}

...+ Significant impact on Flavor Physics measurements, i.e., those with Bs oscillation...

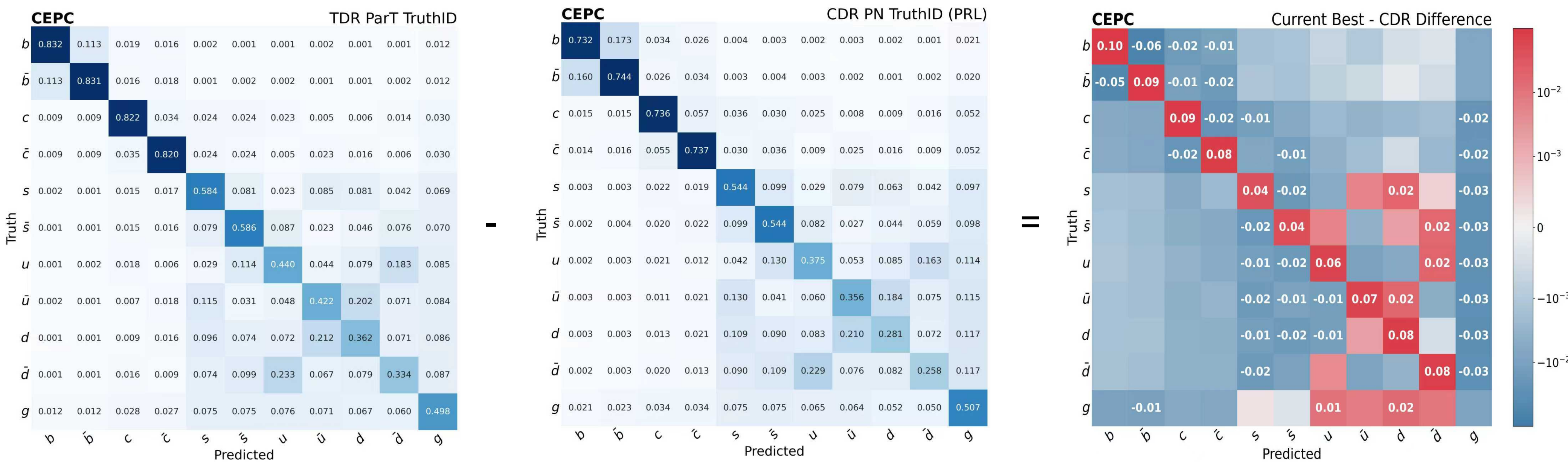
From specialized Models to LLM



- Comparable result with different scaling behavior
- Para. Numbers: PN 360k, ParT 2.4M, BINBBT(Large Language Base Model) 150 M
- More details at: *https://arxiv.org/pdf/2412.00129*

超对称
Super Symmetry Technologies

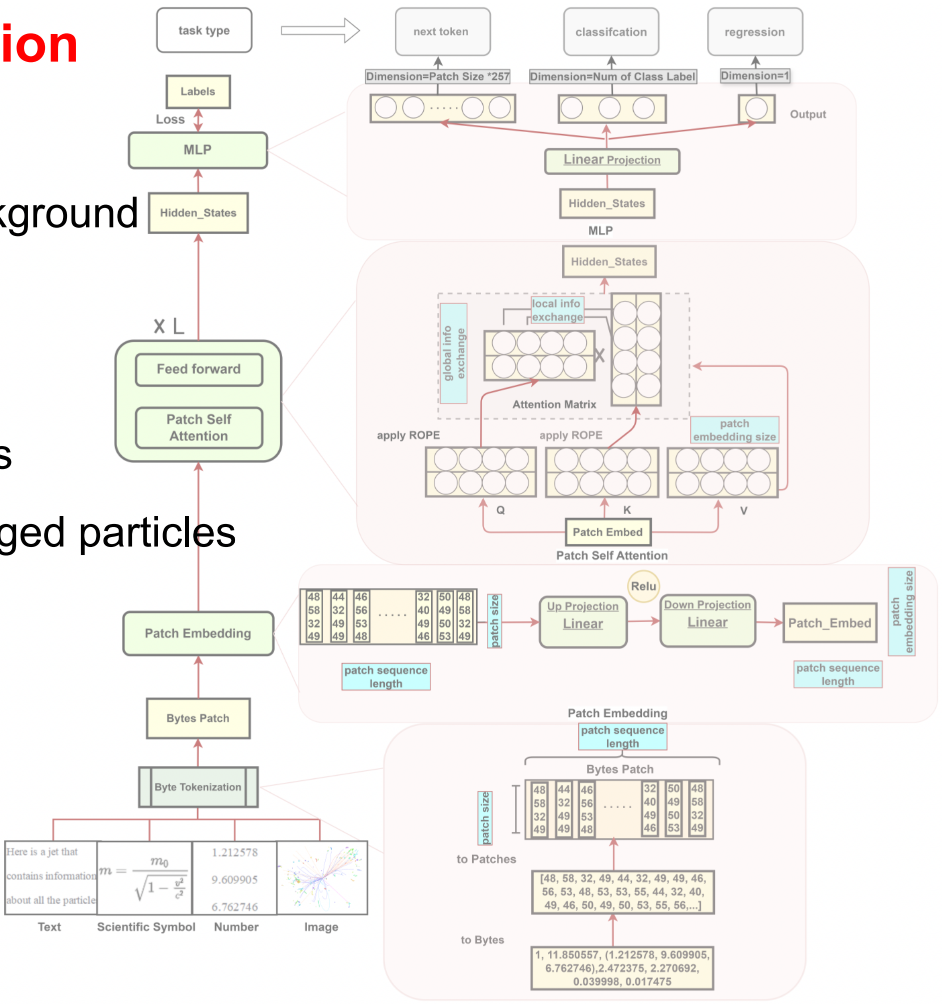
Recent updates... preliminary



- Current Best: ~ 10% improvements in M11
  - Change AI architecture, with extend input variables
  - Vertex optimization
  - ...
- To do:
  - *Scan on generator/hadronization models,*
  - *Better reconstruction of intermediate particles (pi0, phi, Lambda, Kshort, etc)...*

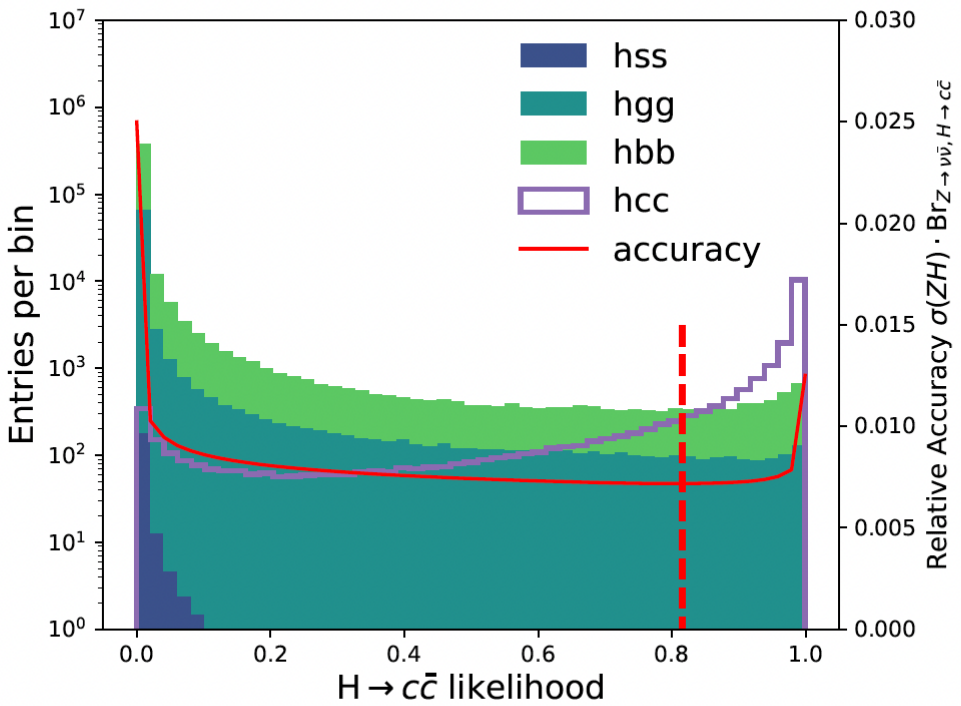
Holistic approach

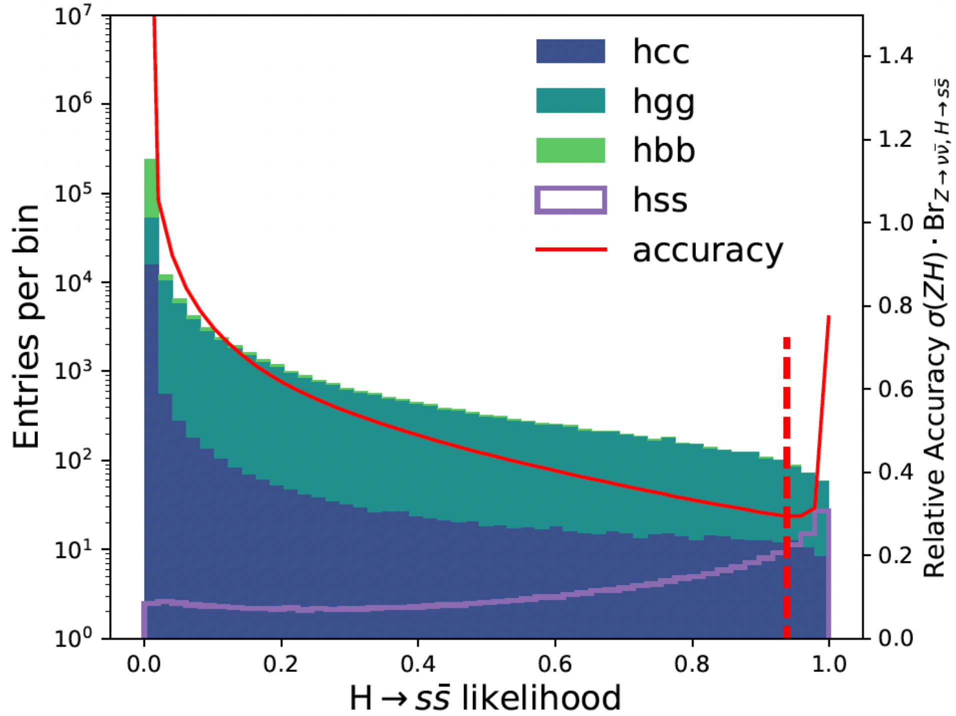
- Provide all reconstructable for **classification**
  - Reco: Jet origin identification
  - Analysis: to distinguish the signal from the background
- In the context of 1-1 correspondence/PFA, inputs =
  - 4 momentum + Pid of all reconstructed particles
  - Track impact parameters of reconstructed charged particles
  - Potentially: parenting info
    - Photon to Pi-0, pions to kaon...
    - Color Singlet (from Z or H)
    - ...
  - Uncertainties (as suggested by Vincent)



- Challenge: high quality simulation, knowledge of Detector response & Theory/interpretation models...

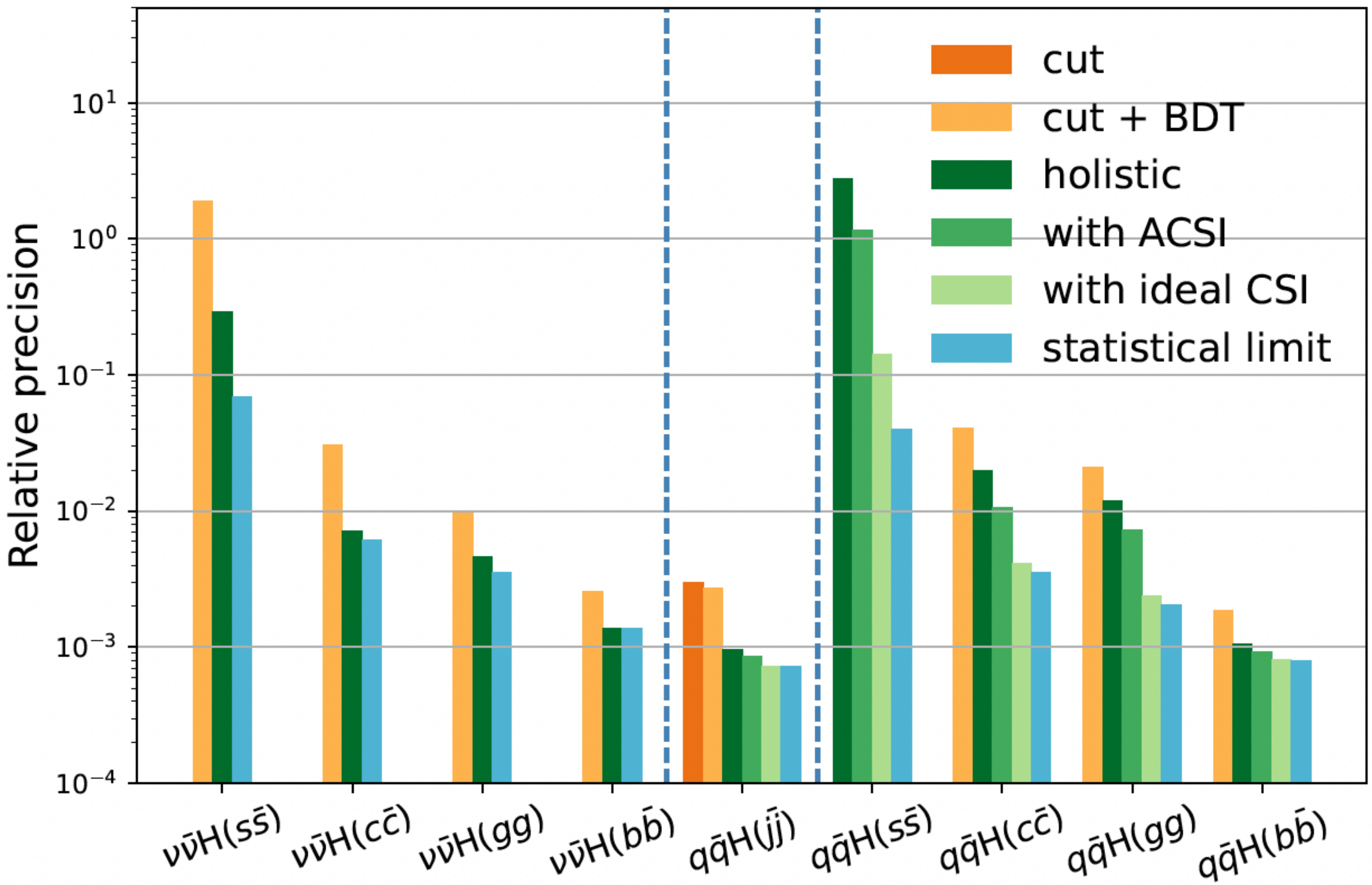
Holistic Analysis: vvH, H→2 jet





- vvH, H→bb/cc/gg/ss measurements: 4 kinds classification
- Simplified analysis with irreducible background...
- Accuracies: 2-6 times better than previous studies (include other bkgrd, BDT based, etc)
- H→ss: close to confirmation!

Holistic approach + ACSI



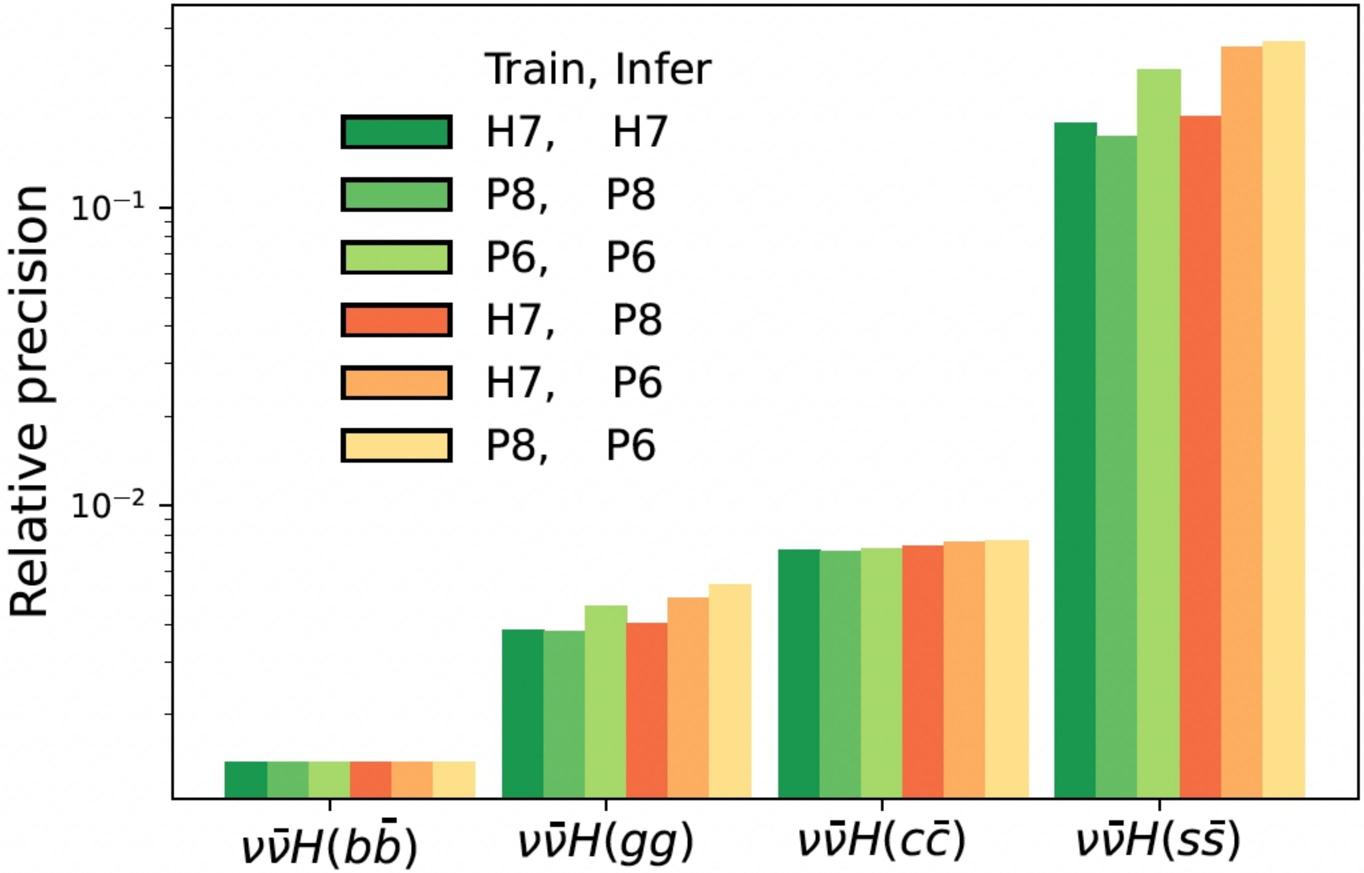
Holistic + ACSI: improves the accuracy by 2 – 6 times

ACSI makes a leap even from Holistic, but still has significant room to improve...

H→ss within the reach...

https://arxiv.org/pdf/2506.11783

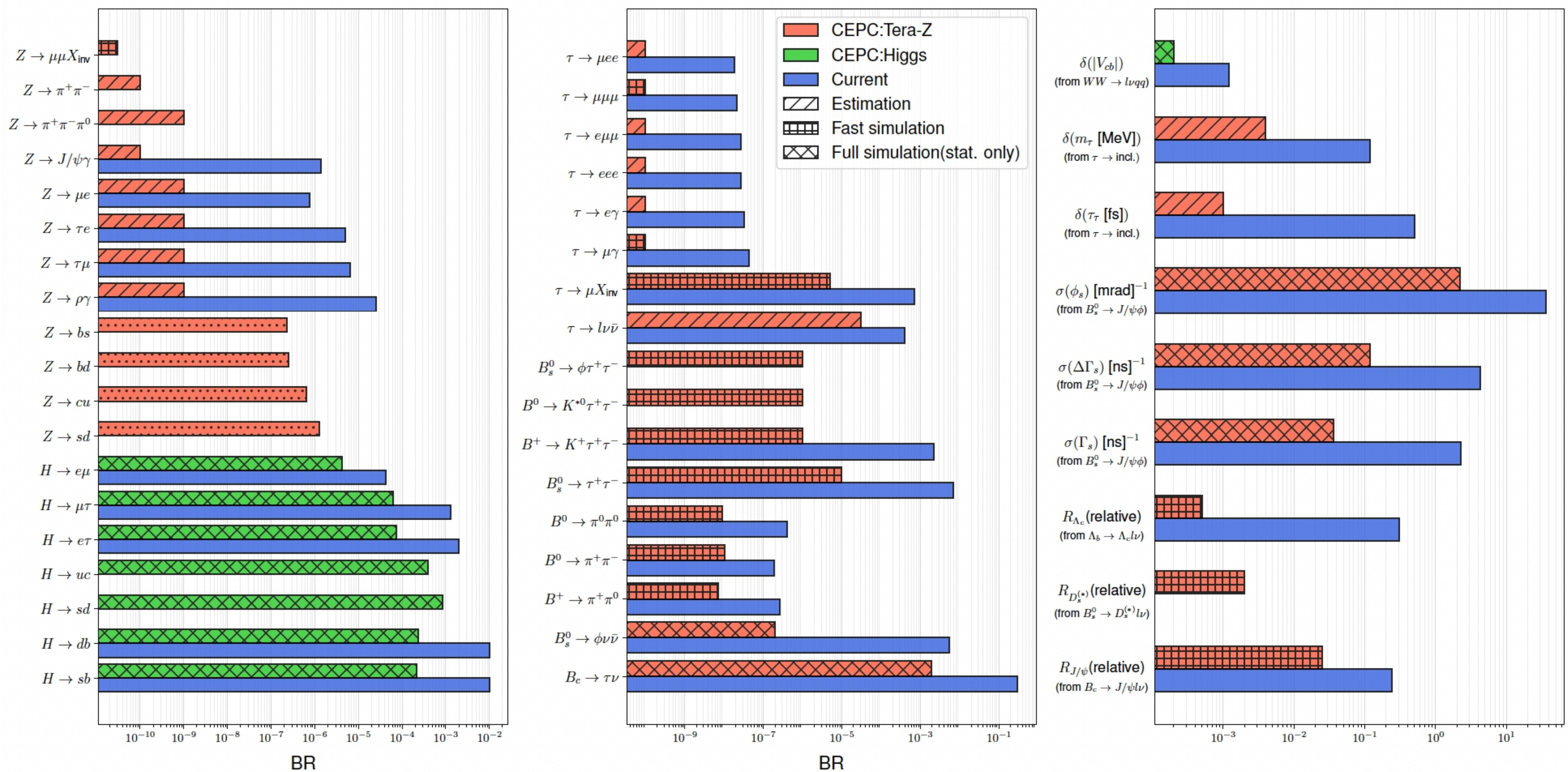
Supervised learning: need High Quality MC



The Holistic approach is in principle free from human intervene...

Human define the goal (the signal), AI serves as the mean...

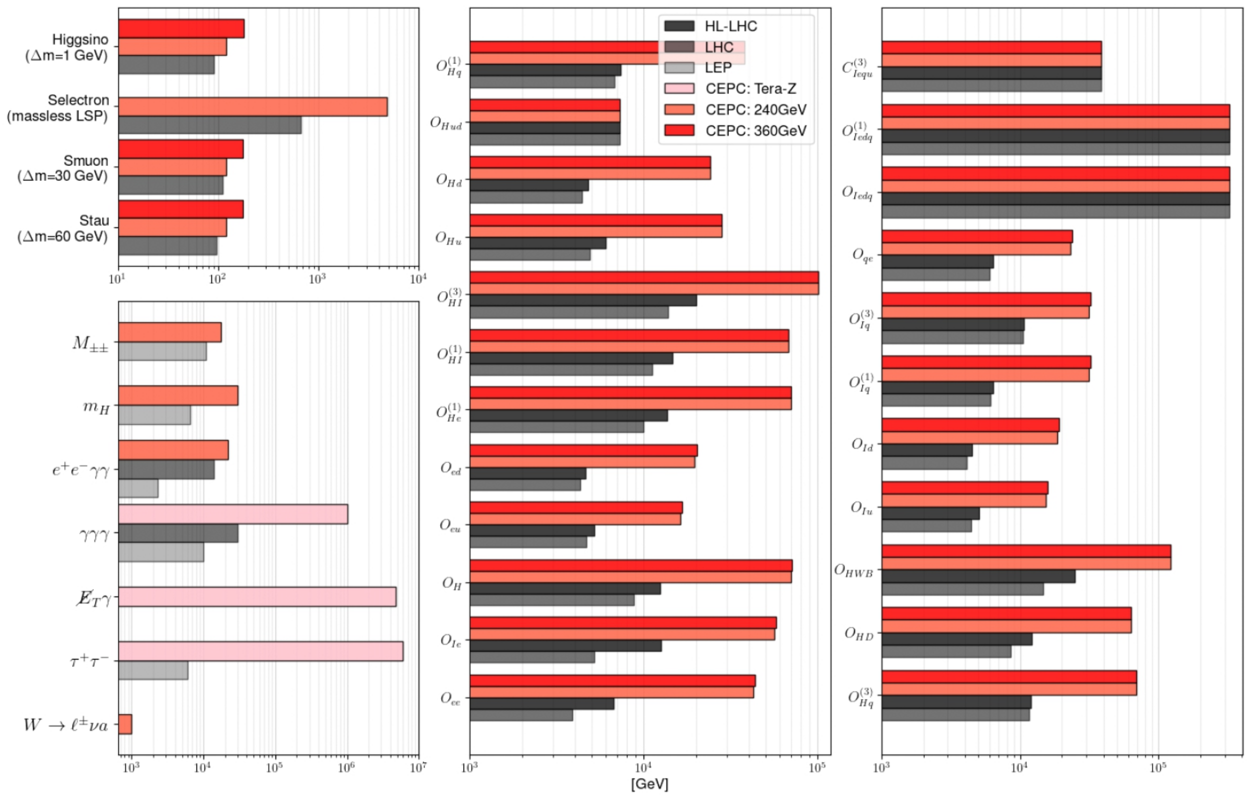
Flavor physics

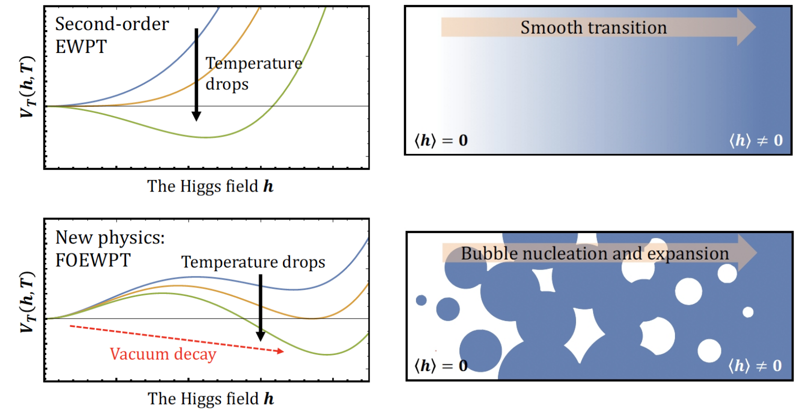


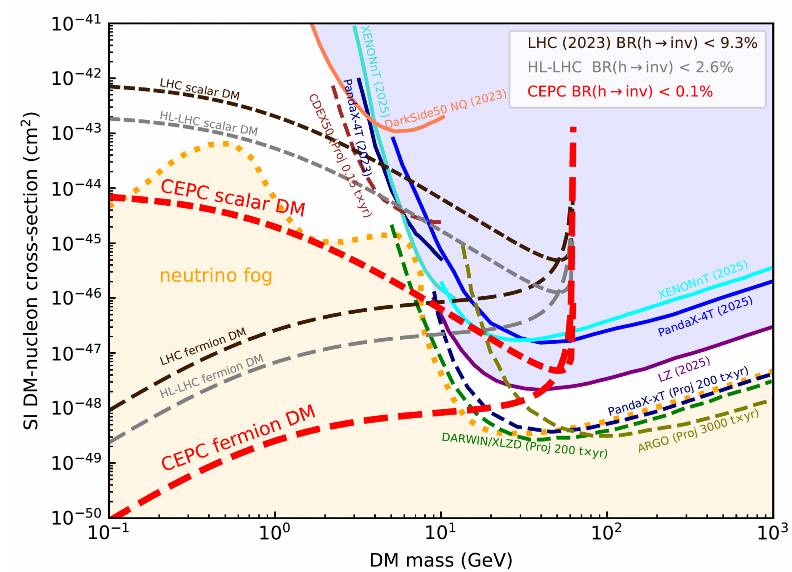
See the non-seen: i.e, Bc→tauv, Bs→Phivv
Orders of magnitudes improvements (1 – 2.5 orders...).
Access New Physics with energy scale of 10 TeV, or even above

https://arxiv.org/pdf/2412.19743

Direct New physics search







Matter Origin, Dark matter...
Access to NP ~ 100 TeV...

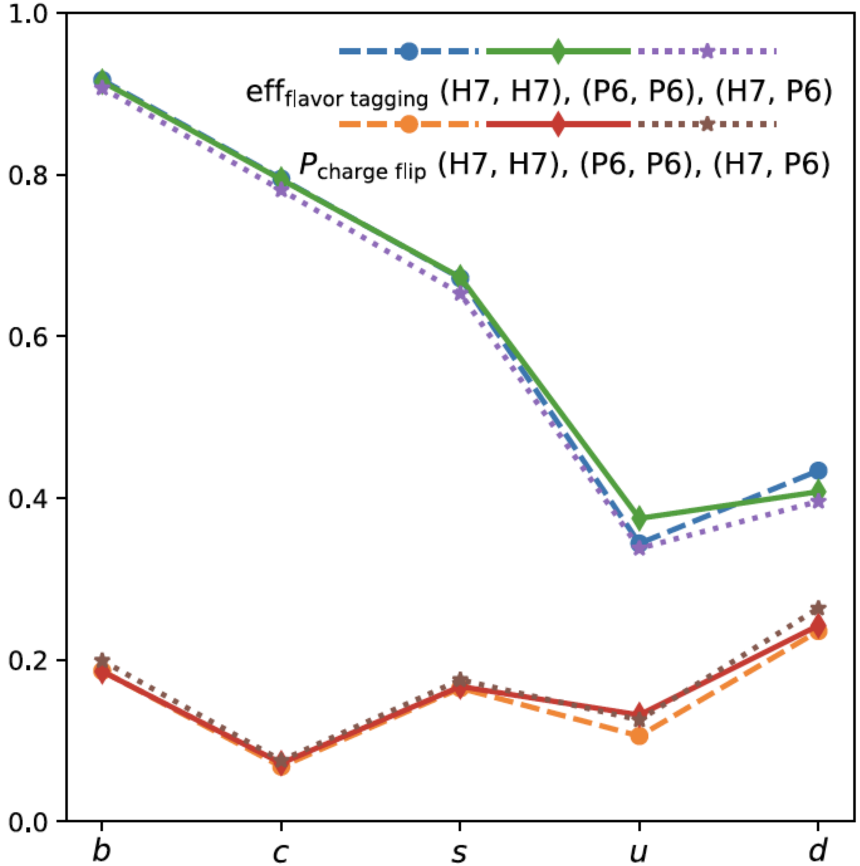
https://arxiv.org/pdf/2505.24810

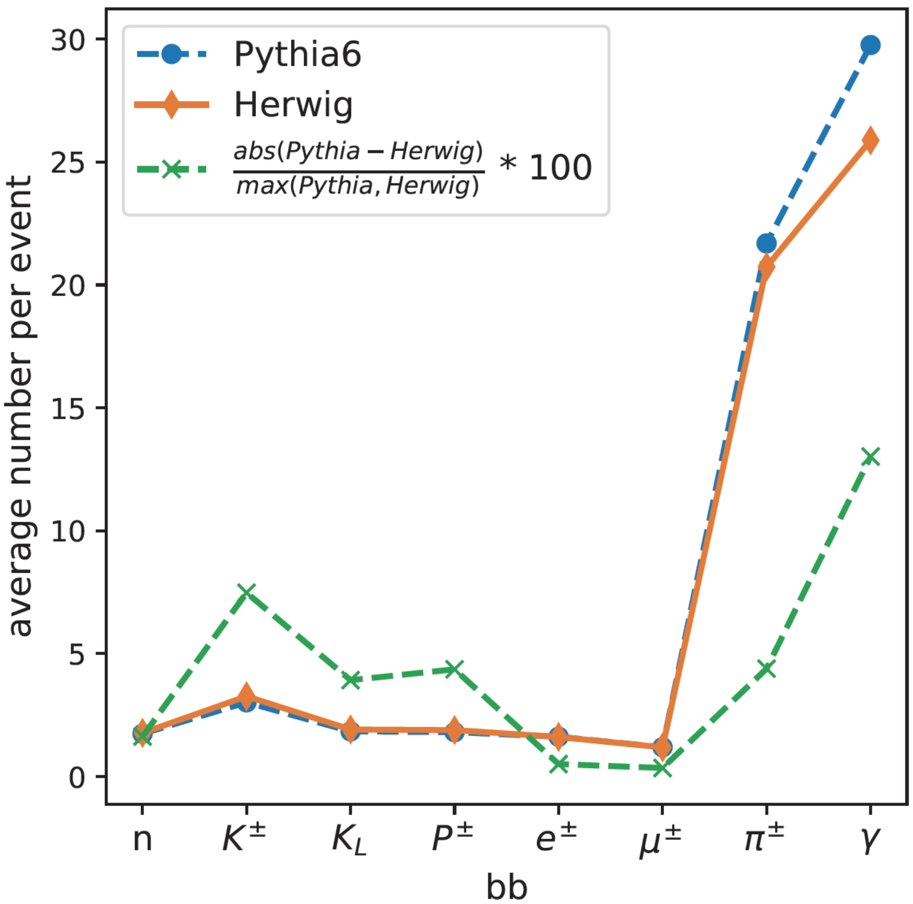
Summary

- ***... Higgs factory has strong discovery power to NP, its detector & reconstruction should and could have excellent performance...***
- ***AI could strongly enhance the discovery power of Higgs factory: 3 times & more...***
  - *Holistic approach*
    - *Reco: Jet origin id, 'see' the quark & gluons...*
    - *Analysis: Processing in principle free from Human intervene.*
      - *+ ACSI for full hadronic events*
- ***Multiple challenges need to be addressed... with intriguing prospects...***
  - *Precise Simulation is critical to utilize supervised learning, which request profound understanding of relevant factors – be developed iteratively*
  - *To explore other methodologies: non/weakly-supervised, enhanced, LLM...*
  - *Lots more to explore, with unsupervised, LLM, ... rich interplay & synergies.*
  - ***...***

Back up

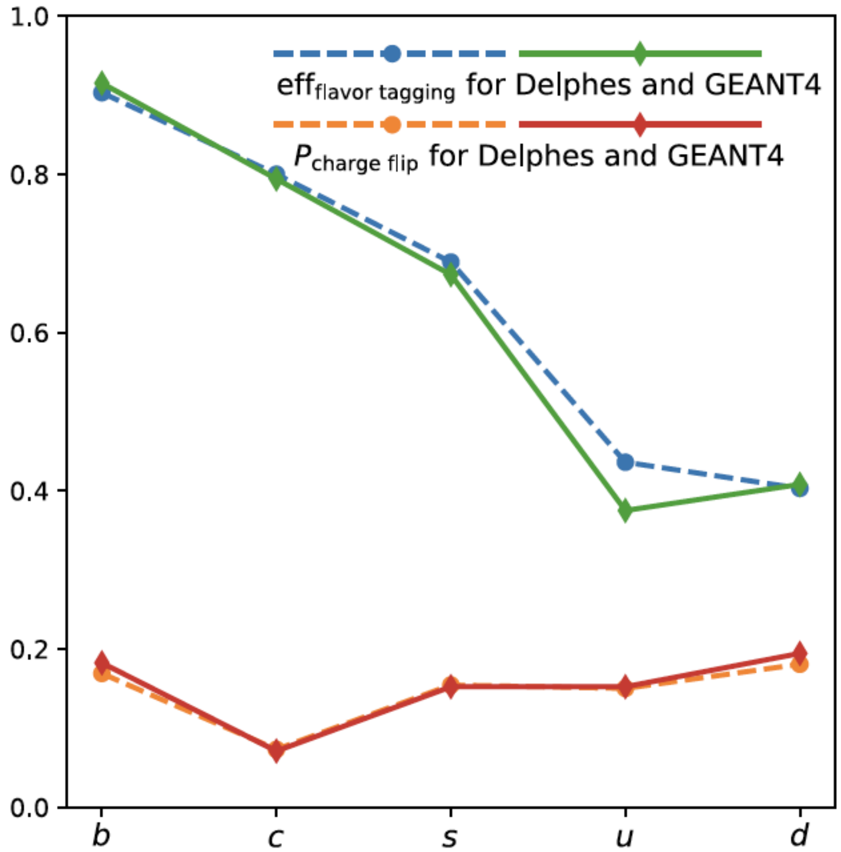
V.S. Hadronization models



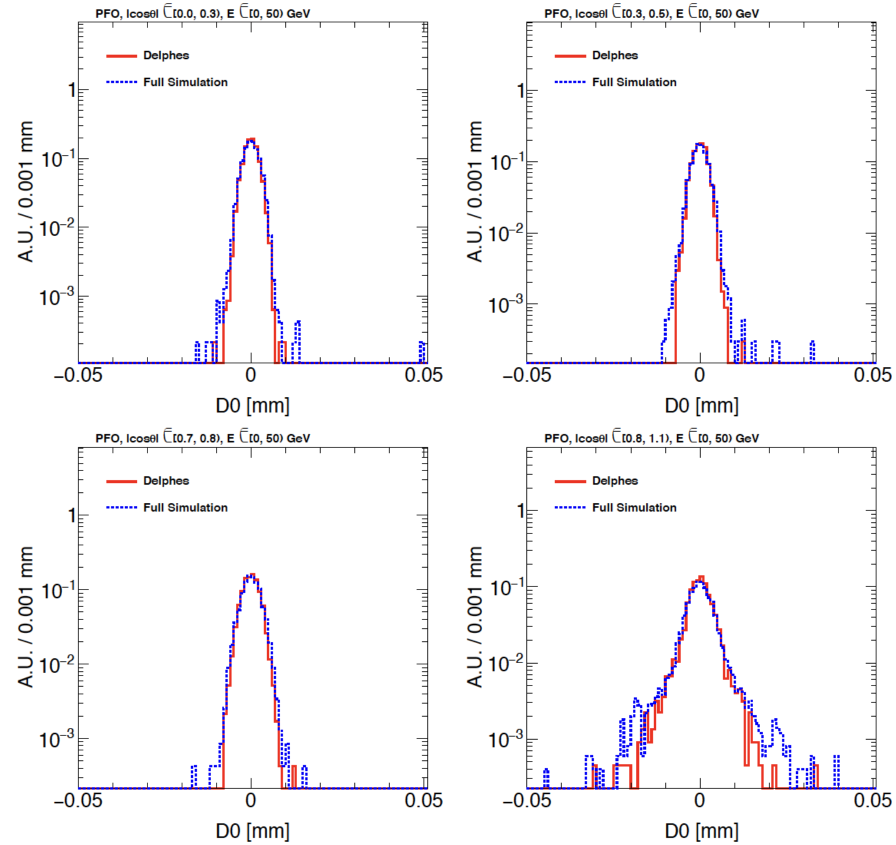


- Much severer descriptions.. in exclusive measurements (i.e., specific hadron generation, decay, etc)

Fast/Full Simulation



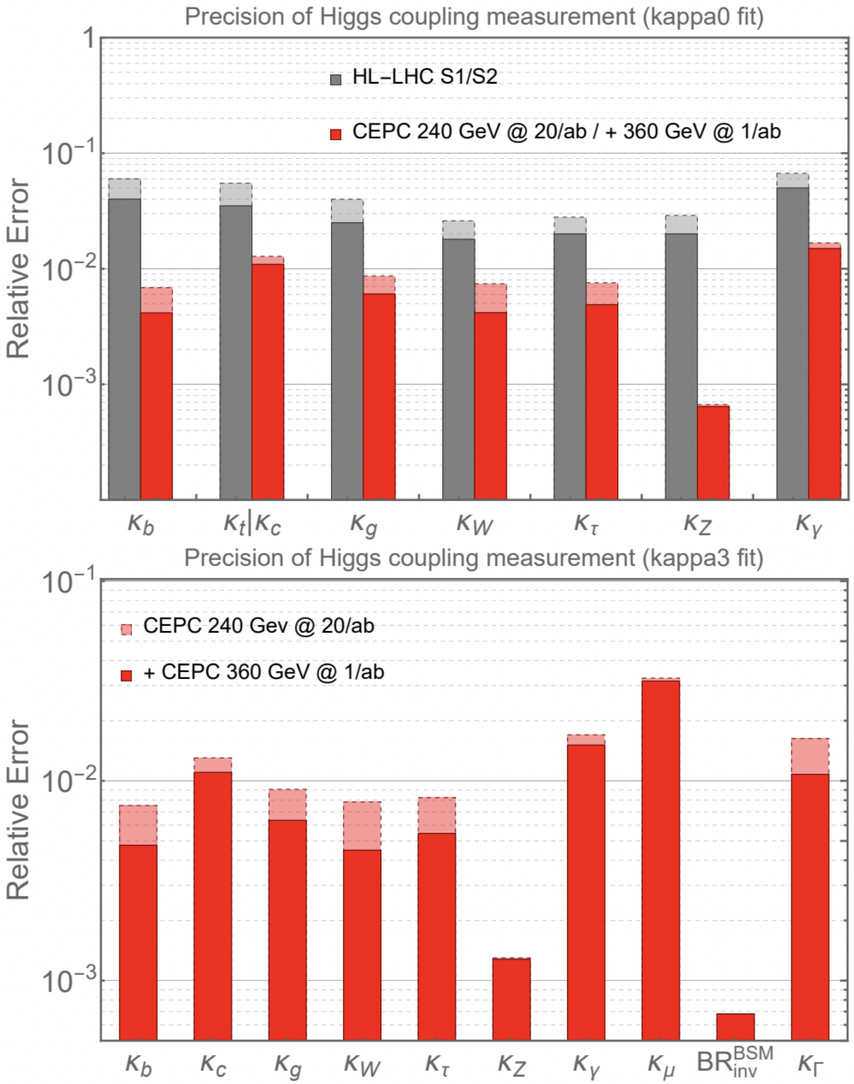
Z->μμ (91.2 GeV)



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

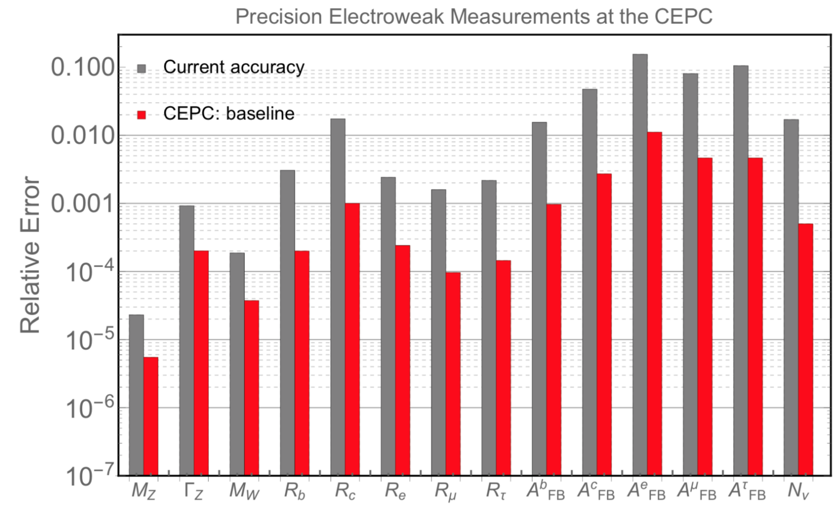
Higgs

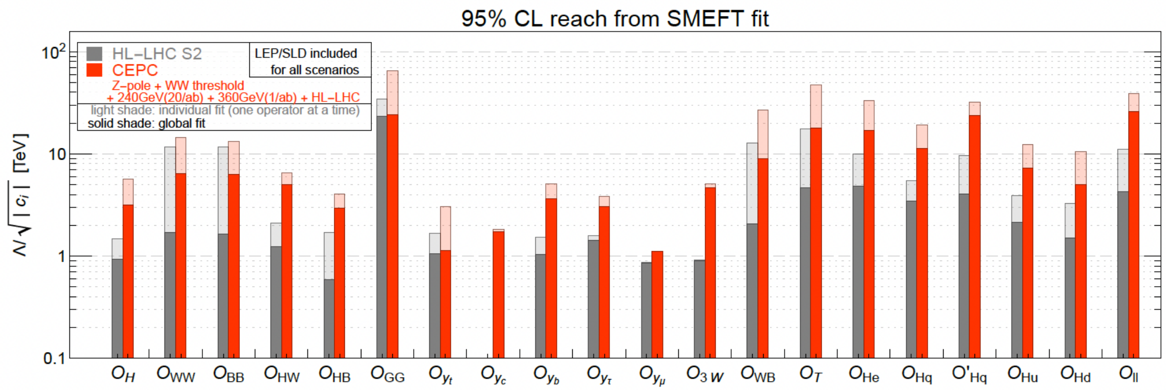
	240 GeV, 20 ab^{-1}		360 GeV, 1 ab^{-1}		
	ZH	**vvH**	ZH	**vvH**	**eeH**
inclusive	**0.26%**		**1.40%**	\	\
H→bb	**0.14%**	**1.59%**	**0.90%**	**1.10%**	**4.30%**
H→cc	**2.02%**		**8.80%**	**16%**	**20%**
H→gg	**0.81%**		**3.40%**	**4.50%**	**12%**
H→WW	**0.53%**		**2.80%**	**4.40%**	**6.50%**
H→ZZ	**4.17%**		**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%**		
$\mathrm{Br}_{upper}(H \to inv.)$	**0.07%**				
Γ_H	**1.65%**		**1.10%**		

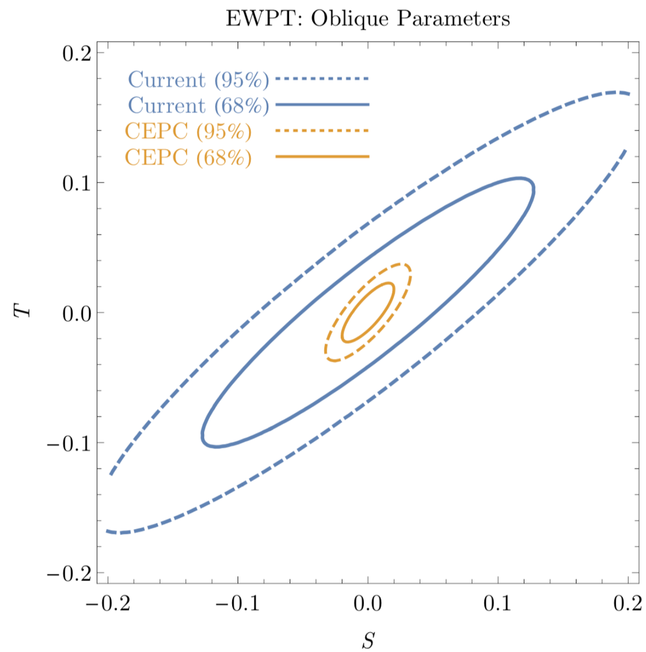


Plus EW & SMEFT

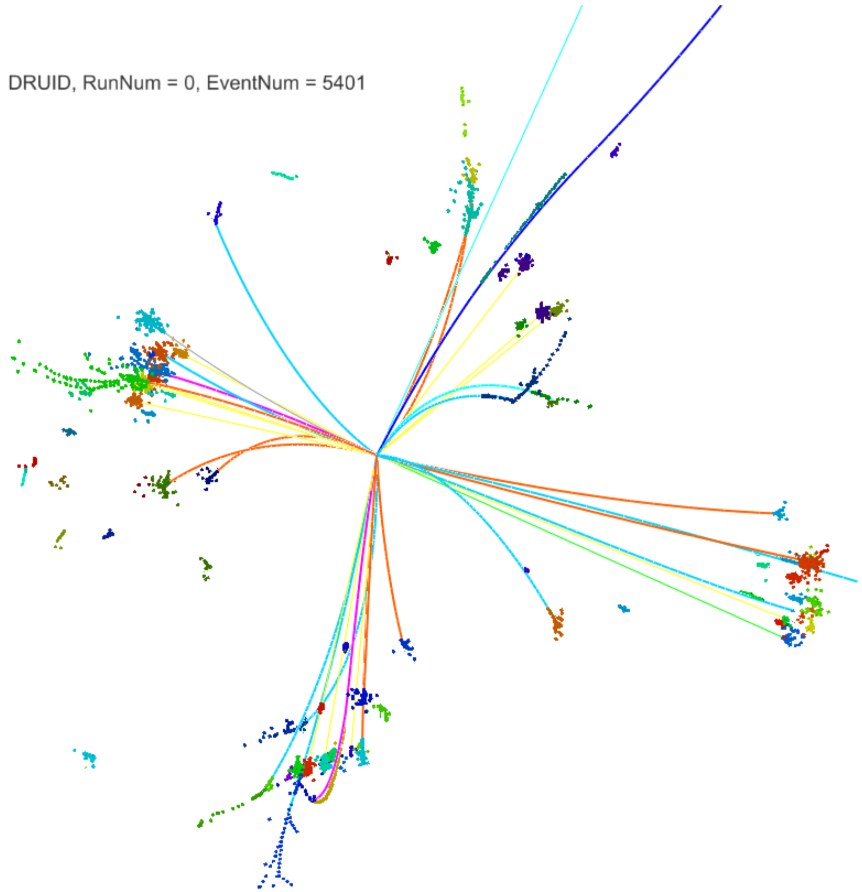
Observable	current precision	CEPC precision (Stat. Unc.)	CEPC runs	main systematic
Δm_Z	2.1 MeV [37–41]	0.1 MeV (0.005 MeV)	Z threshold	E_{beam}
$\Delta\Gamma_Z$	2.3 MeV [37–41]	0.025 MeV (0.005 MeV)	Z threshold	E_{beam}
Δm_W	9 MeV [42–46]	0.5 MeV (0.35 MeV)	WW threshold	E_{beam}
$\Delta\Gamma_W$	49 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

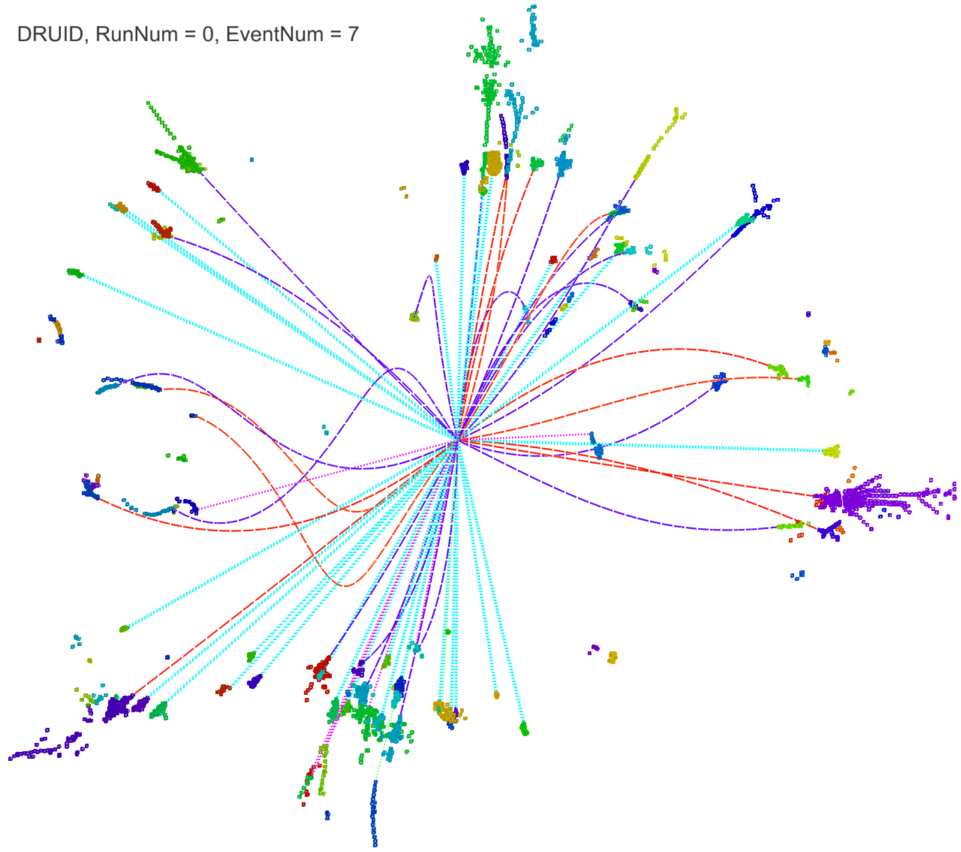






Color Singlet Identification





at full hadronic ZH event

CSI: bottleneck for measurement at full hadronic events

JHEP
PUBLISHED FOR SISSA BY SPRINGER

RECEIVED: *March 11, 2022*
REVISED: *September 9, 2022*
ACCEPTED: *November 11, 2022*
PUBLISHED: *November 16, 2022*

JHEP11(2022)100

The Higgs→ $b\bar{b}, c\bar{c}, gg$ measurement at CEPC

Yongfeng Zhu, Hanhua Cui and Manqi Ruan

Institute of High Energy Physics, Chinese Academy of Sciences,
19B Yuquan Road, Beijing 100049, China
University of Chinese Academy of Sciences,
19A Yuquan Road, Beijing 100049, China

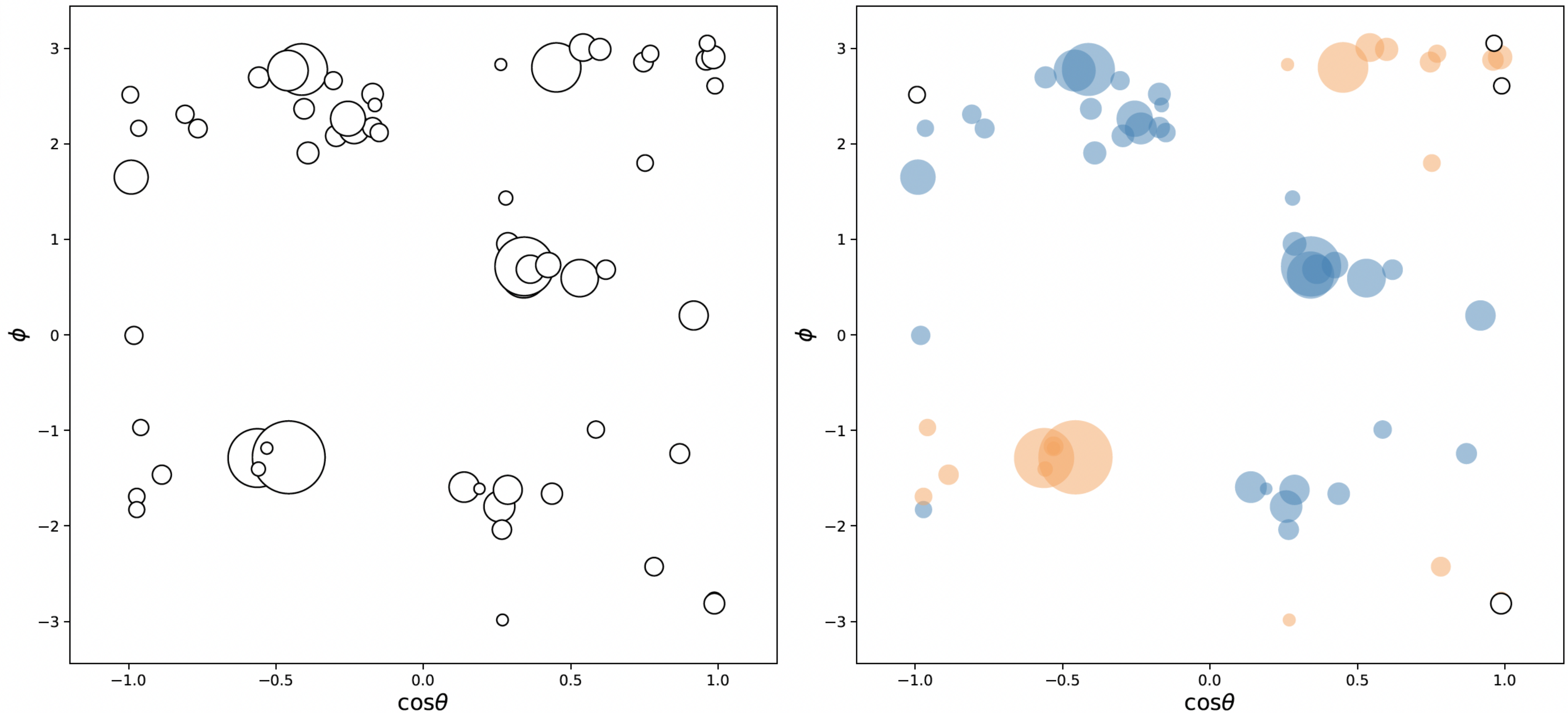
E-mail: ruanmq@ihep.ac.cn

Z decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
$Z \to e^+e^-$	1.57%	14.43%	10.31%
$Z \to \mu^+\mu^-$	1.06%	10.16%	5.23%
$Z \to q\bar{q}$	0.35%	7.74%	3.96%
$Z \to \nu\bar{\nu}$	0.49%	5.75%	1.82%
combination	0.27%	4.03%	1.56%

Table 3. The signal strength accuracies for different channels.

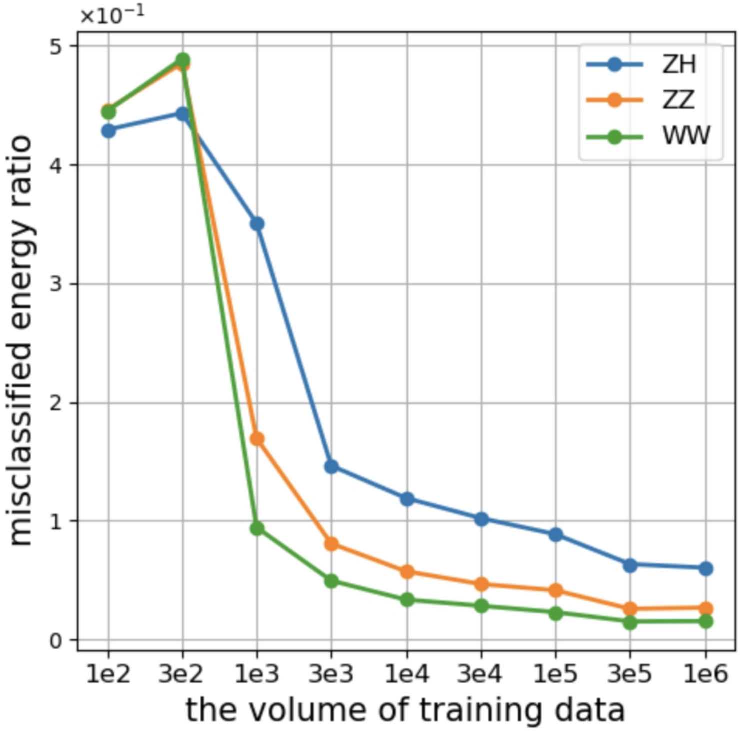
- H→cc & gg measurements at qqH channel is much worse vvH channels, despite the former has 3.5 times more signal statistic
- Reason: Failure of Color Singlet Identification – to distinguish the decay products of each Color Singlet
  - Z & H for 240/250 GeV Higgs factory
  - Which Higgs boson for Higgs self-coupling (i.e., at vvHH events at 500 GeV, etc)

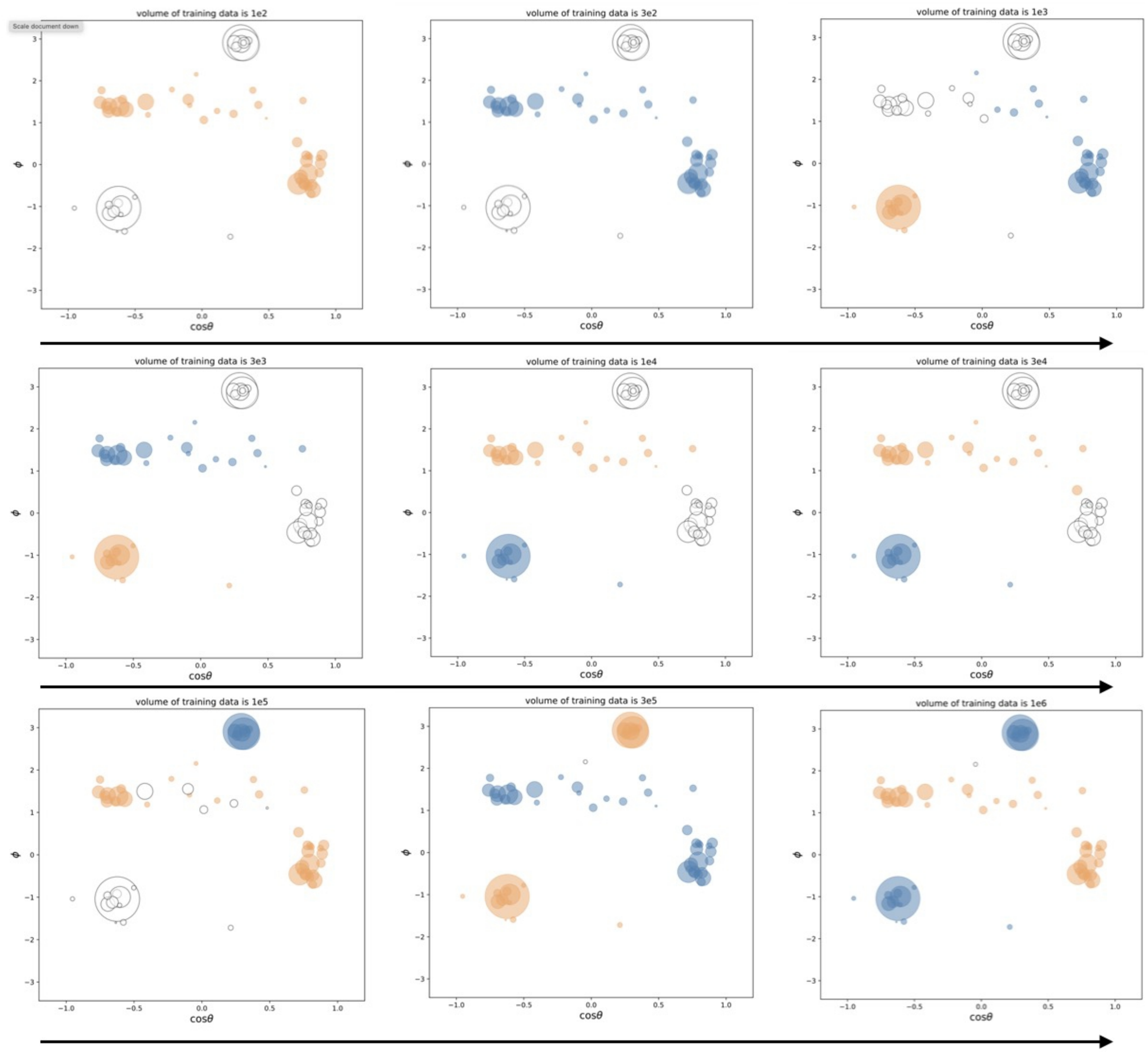
Advanced CSI using AI



Yongfeng, Hao, Yuexin, etc

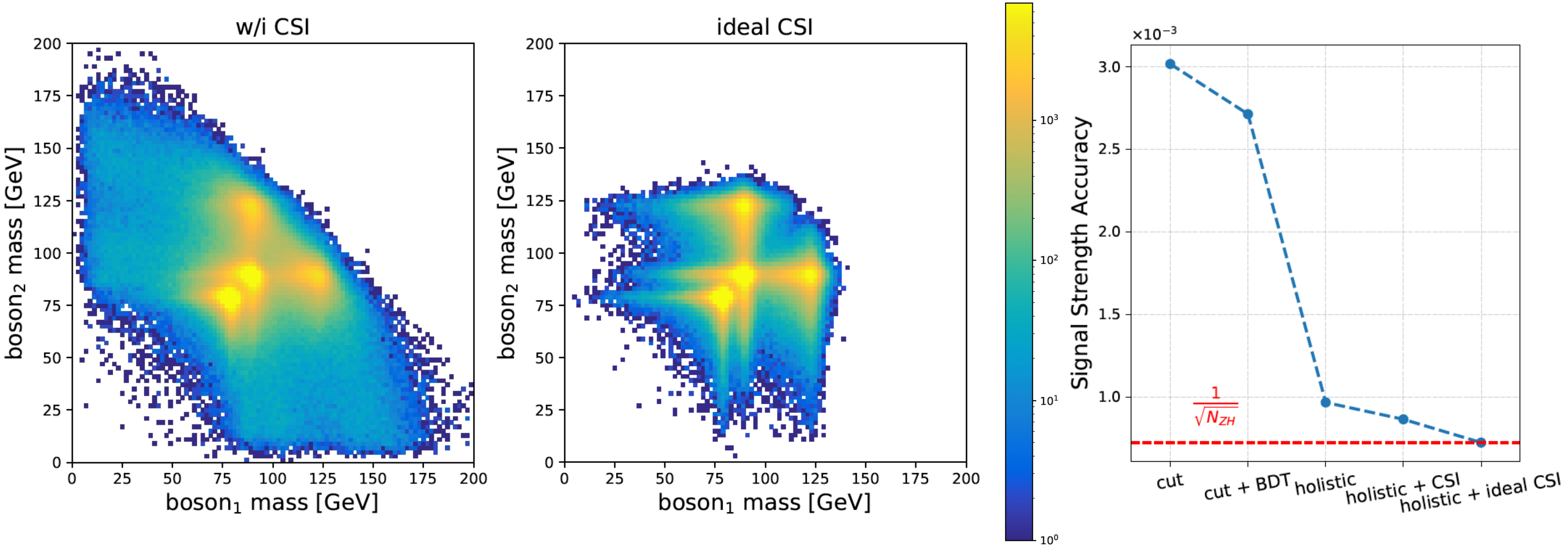
Scaling...


×10⁻¹
misclassified energy ratio
the volume of training data
1e2
3e2
1e3
3e3
1e4
3e4
1e5
3e5
1e6
ZH
ZZ
WW


volume of training data is 1e2
volume of training data is 3e2
volume of training data is 1e3
volume of training data is 3e3
volume of training data is 1e4
volume of training data is 3e4
volume of training data is 1e5
volume of training data is 3e5
volume of training data is 1e6
cosθ
ϕ

A toy analysis: identify full hadronic ZH signal from ZZ + WW background

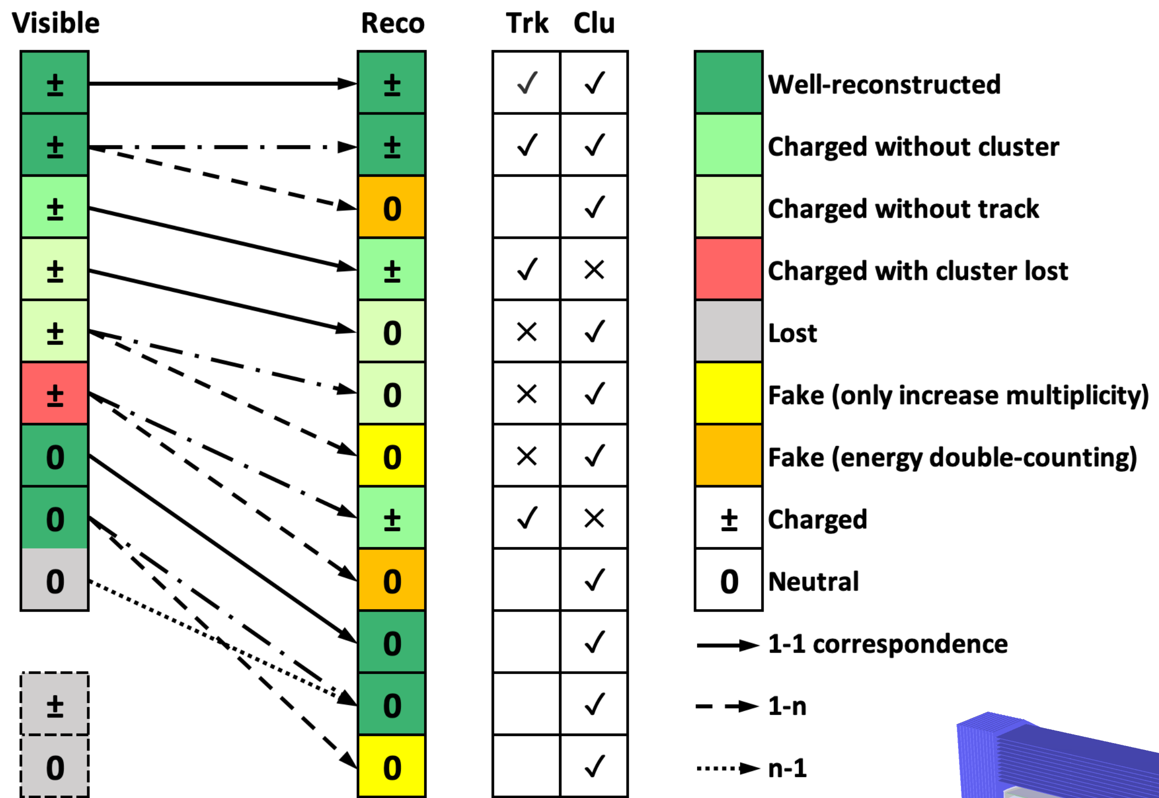
540k ZH + 3.1M ZZ + 47 M WW full hadronic events (~ 5.6 iab), result scale to 20 iab



Holistic: use all the reconstructable info to category signal & different background

1-1 correspondence reconstruction

Final state particles



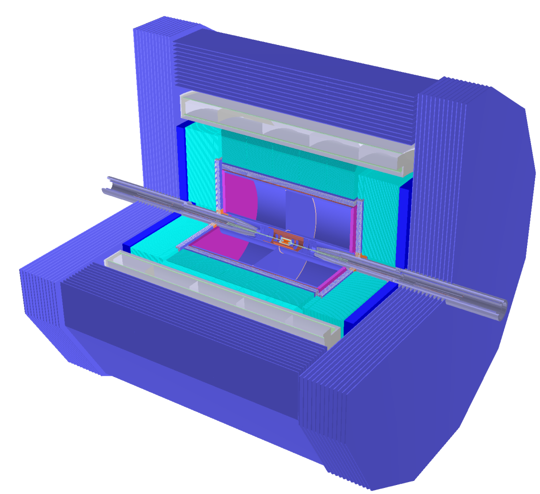
https://arxiv.org/abs/2411.06939

Computer Physics Communications 314 (2025) 109661
Contents lists available at ScienceDirect
Computer Physics Communications
journal homepage: www.elsevier.com/locate/cpc

Computational Physics

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

Yuexin Wang [a,b], Hao Liang [a,c,d], Yongfeng Zhu [e], Yuzhi Che [a,f], Xin Xia [a,c], Huilin Qu [g], Chen Zhou [e], Xuai Zhuang [a,c], Manqi Ruan [a,c,*]



Necessary studies...

- Beam induced backgrounds: comparative studies...
- Event building with realistic detector time response, including electronic pulse shape & time sequence...
- TPC & Tracker:
  - Dependence of dE/dx or dN/dx performance on the shifting distance & readout threshold/Noise
  - Ion distortion VS shielding & possible correction
  - B-Field mapping
  - Mechanic stability
  - Low Pt track reconstruction
- Calorimeter
  - SiPM: response uniformity & Dynamic range, especially towards large Tile/Bar configuration in ECAL
  - Requirement on the Attenuation length for scintillating materials...
  - Homogenates in space & stability in time
  - Development of Energy & Time Estimator...
- Dead zone/dead channel tolerance
- Performance degrading with different Noise: rates, intrinsic, and radiation relevant ones
- Calibration Procedure & Monitoring methodologies...