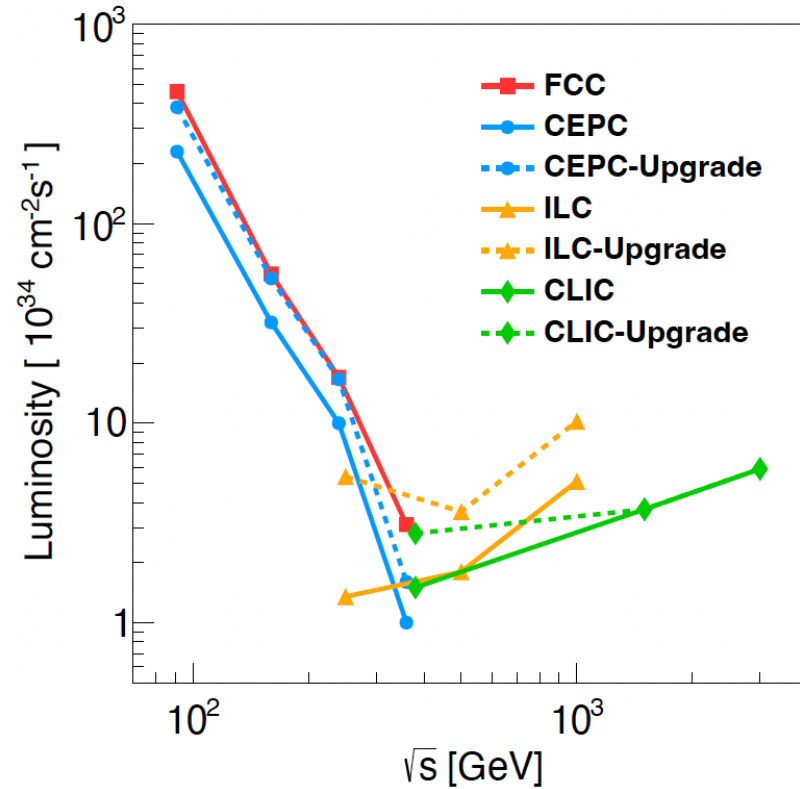
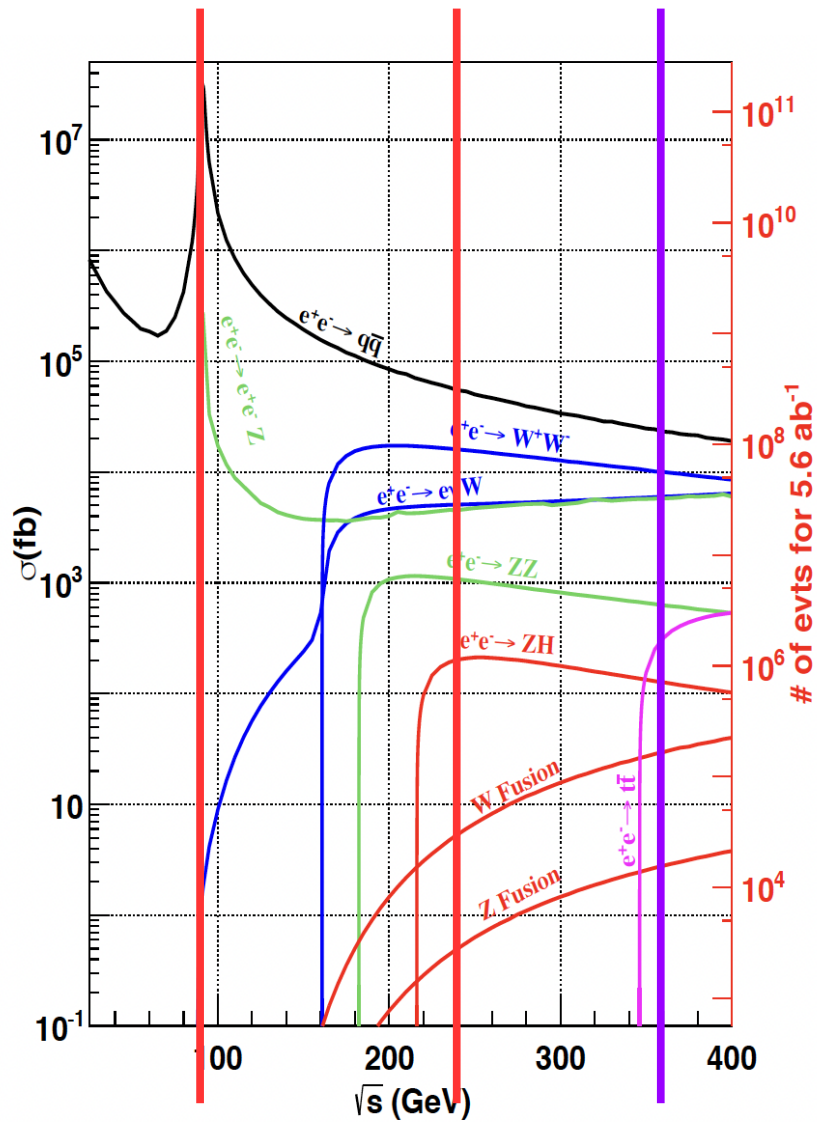




*Learning from all particles,
and optimization of PFA
oriented concepts*

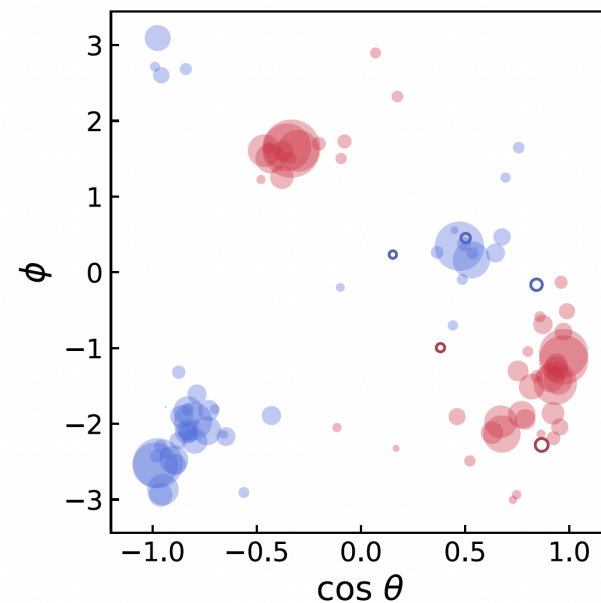
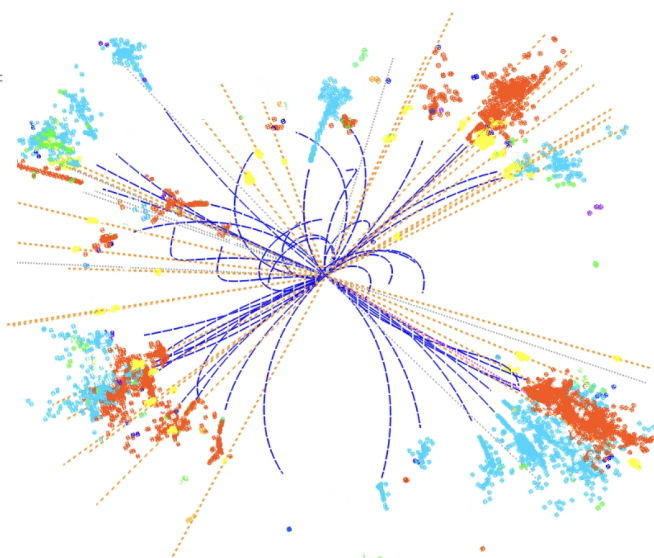
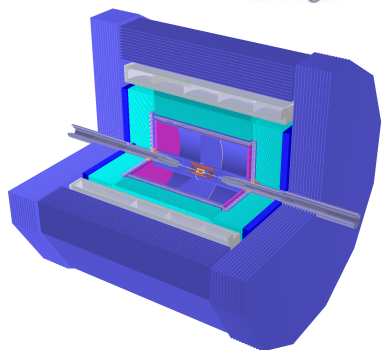
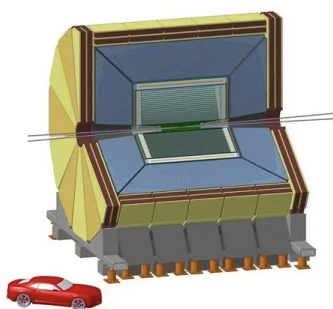
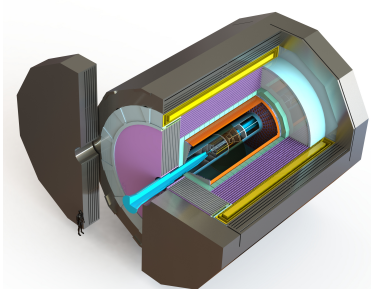
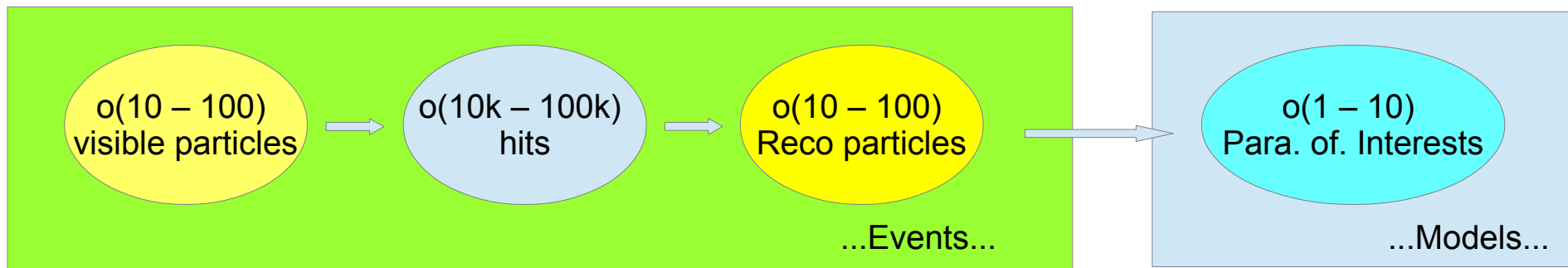
Manqi

Yields \sim Xsec * Lumi * Time

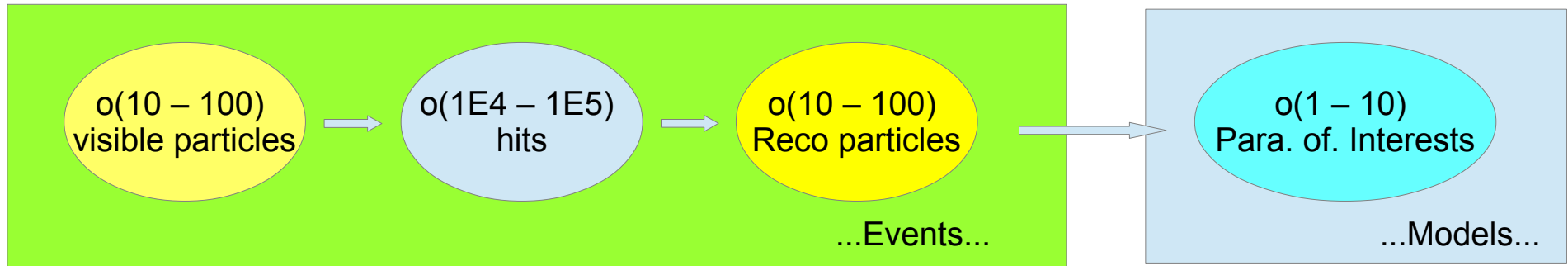


- CEPC: 100 km main ring circumference
- 4 Million Higgs (10 years)
- ~ 1 Giga W (1 year) + 4 Tera Z (2 years)
- Upgradable: Top factory (500 k ttbar)

Information flow

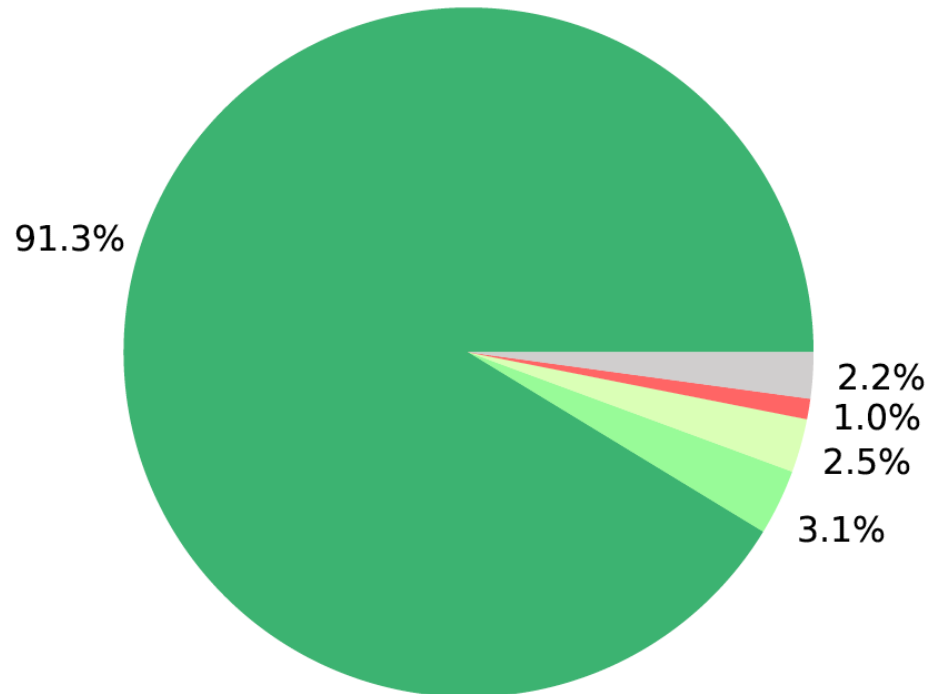
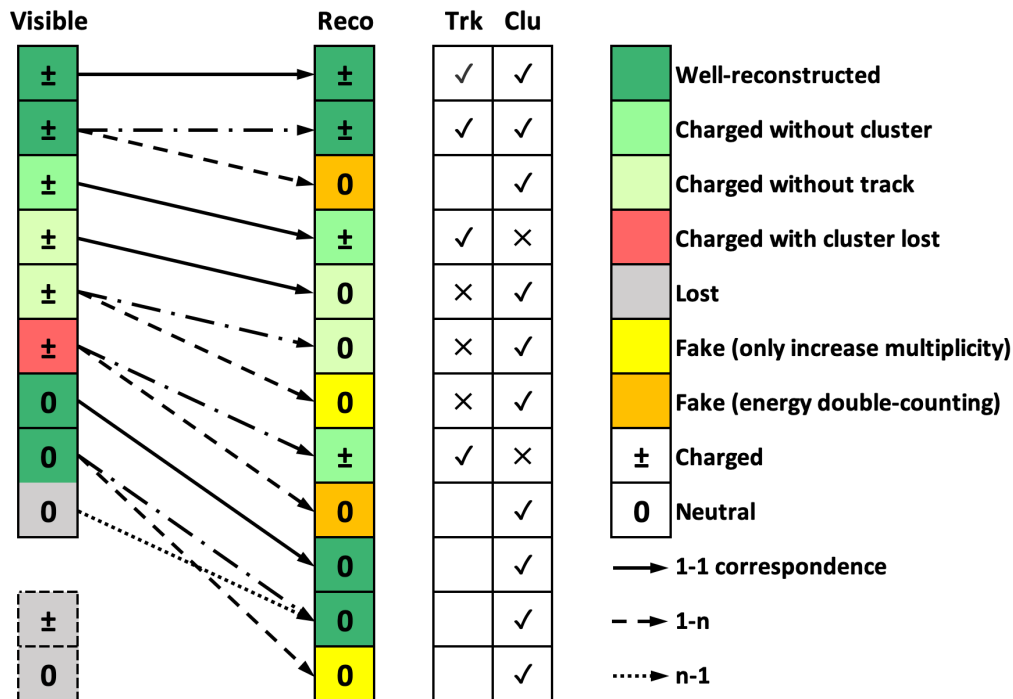


Learning from all particles



- Reconstructed particles quality...: **1-1 correspondence reco + detector design**
- **Holistic approach** using all reco particles
 - Physics object reconstruction: Jet origin id
 - Identifying the nature of physics events: Signal & Backgrounds
 - Advanced color singlet identification
 - ...

1-1 correspondence reconstruction



<https://arxiv.org/abs/2411.06939>

Computer Physics Communications 314 (2025) 109661

Contents lists available at ScienceDirect

Computer Physics Communications

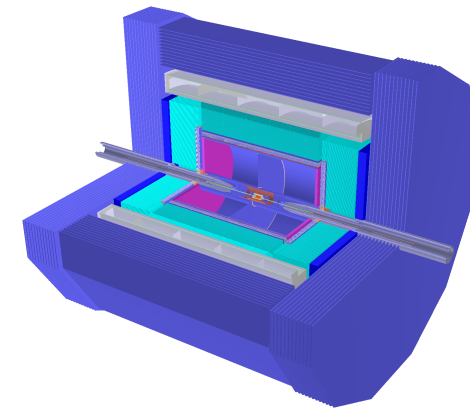
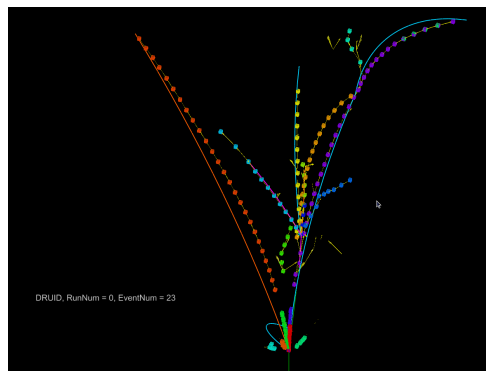
journal homepage: www.elsevier.com/locate/cpc



Computational Physics

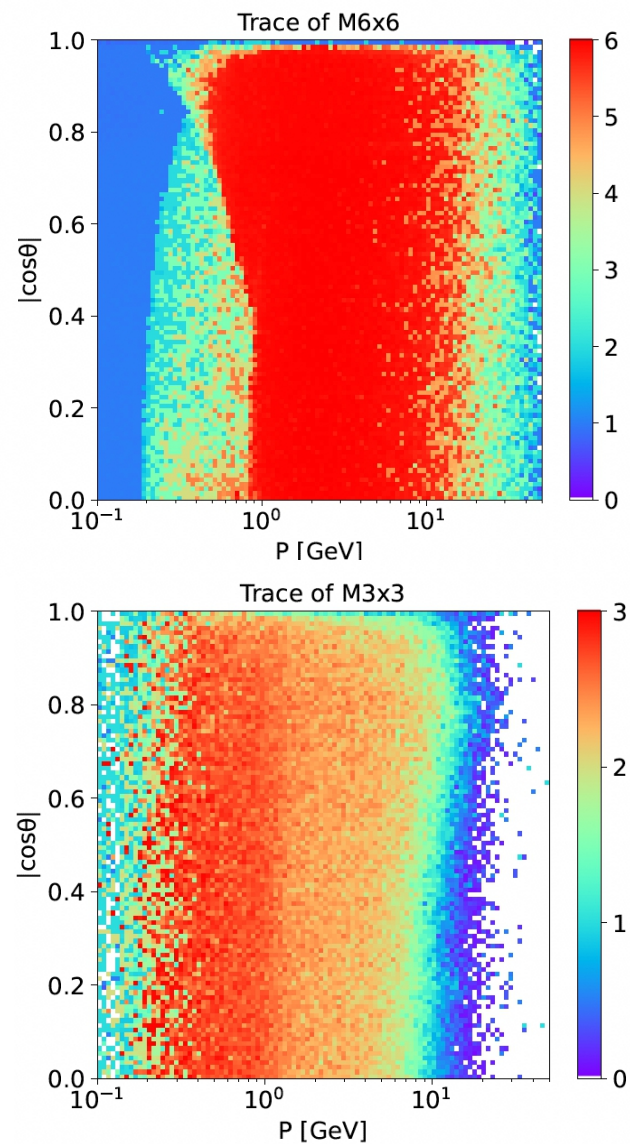
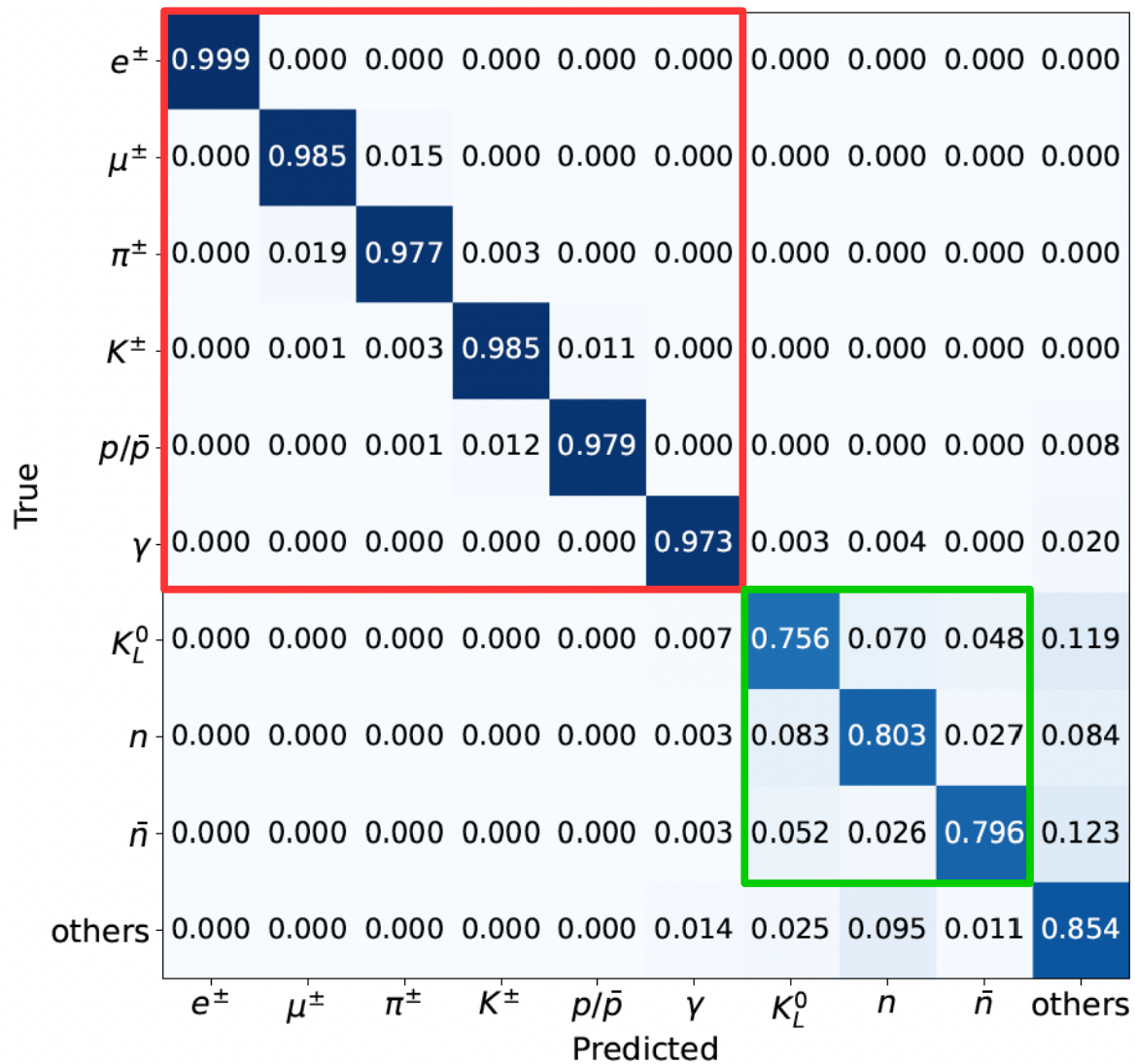
One-to-one correspondence reconstruction at the electron-positron Higgs factory

Yuexin Wang^{a,h}, 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,*}

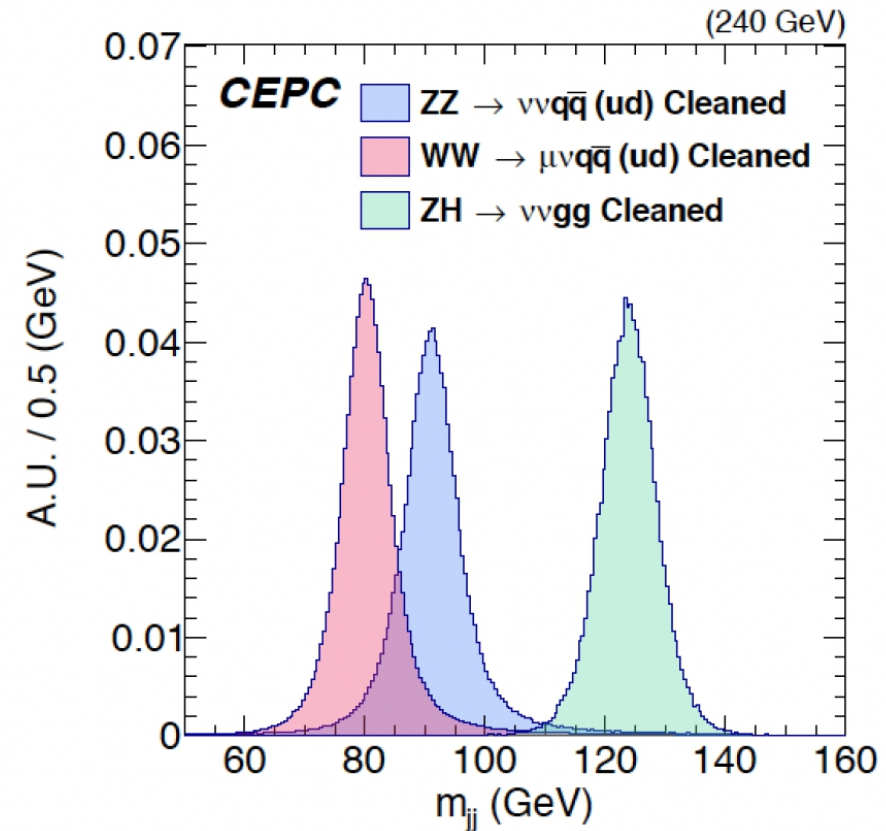
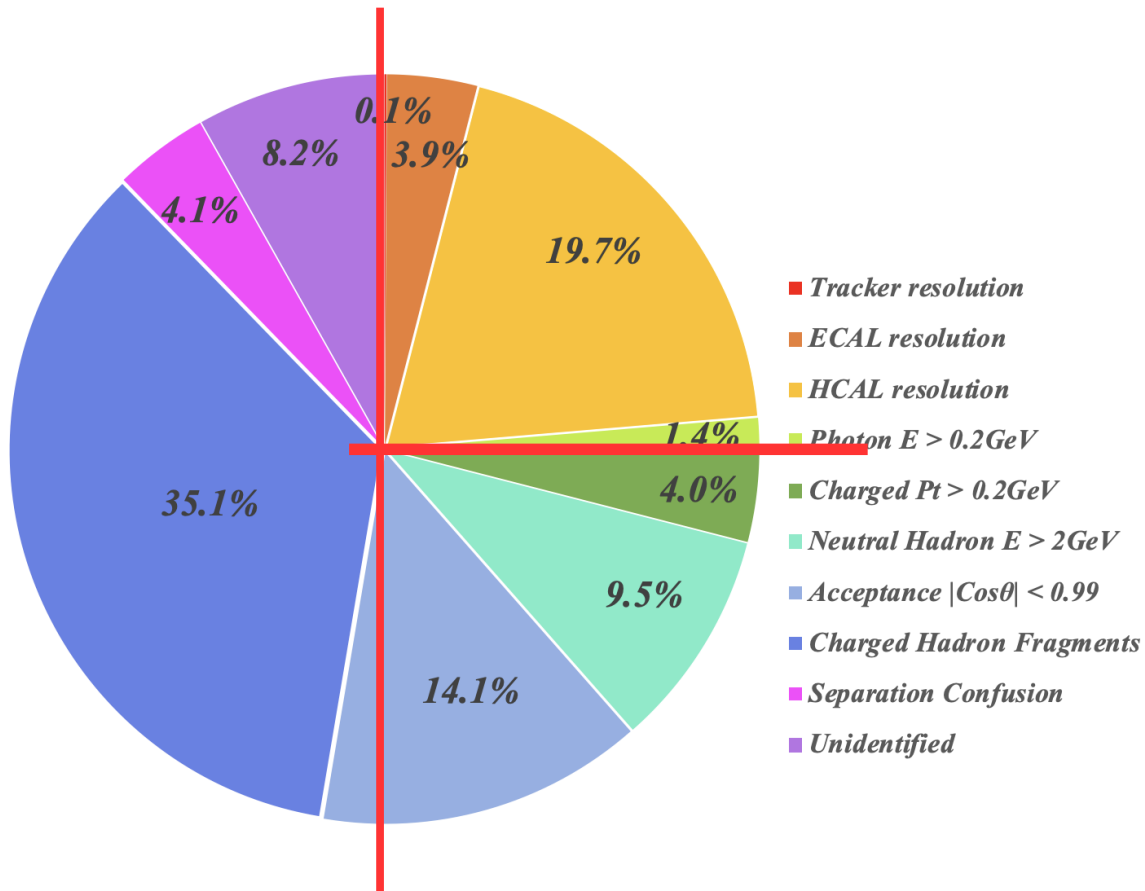


AURORA det. Concept with 5d calo.

Pid: differential performance



BMR & Decomposition



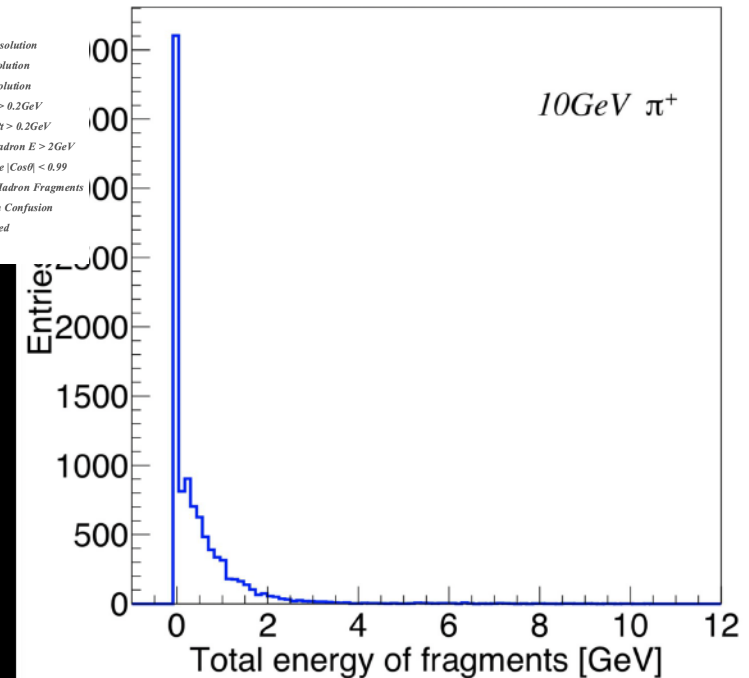
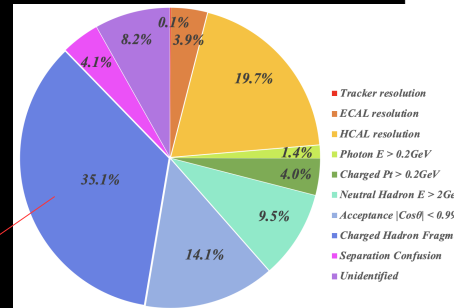
CEPC CDR baseline: TPC + SiW ECAL + RPC HCAL, BMR of 3.8%

Confusion & Acceptance & Intrinsic Resolution: 50%, 25%, 25%

Key contributions: Fragments (leading confusion) & Intrinsic HCAL resolution

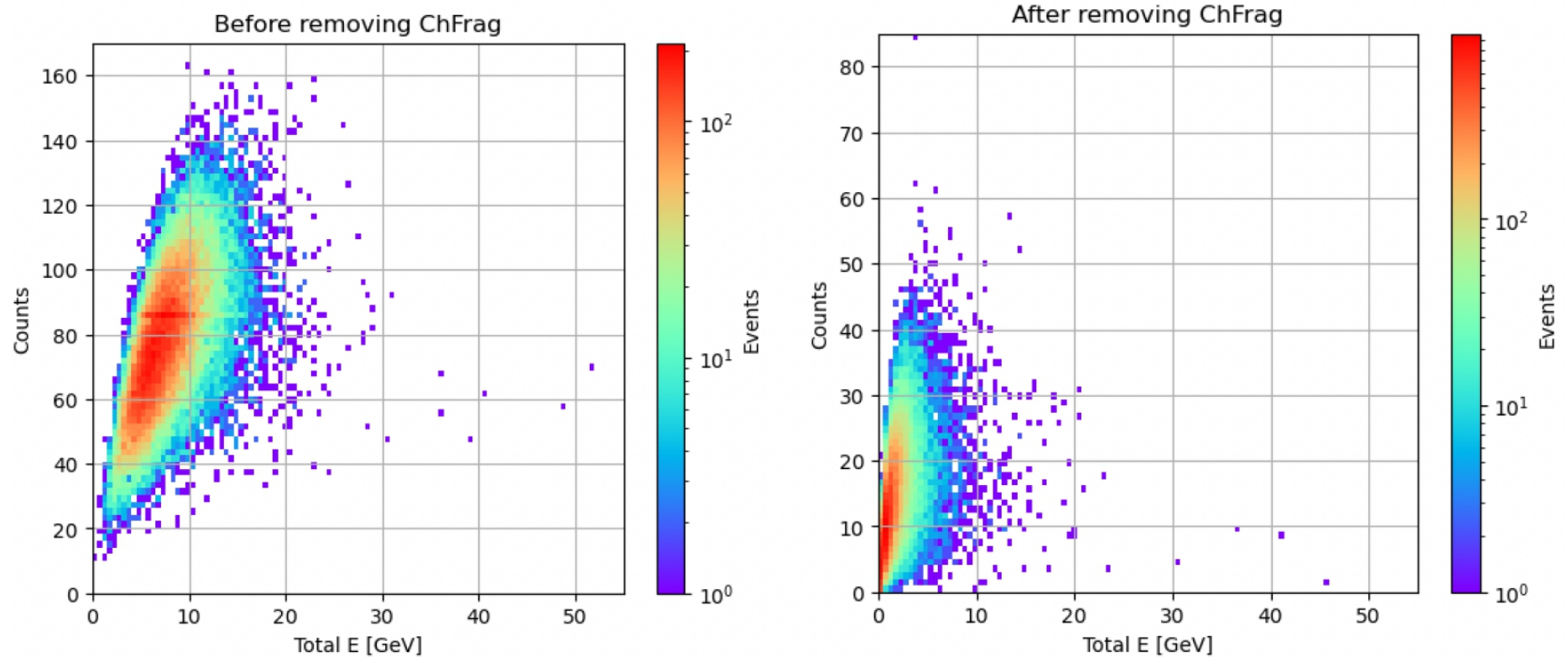
Cluster splitting: the most severe confusions

DRUID, RunNum = 0, EventNum = 0



Time/pattern recognition may help a lot, in identify the charged cluster fragmentations without arise the threshold for the neutral hadron significantly...

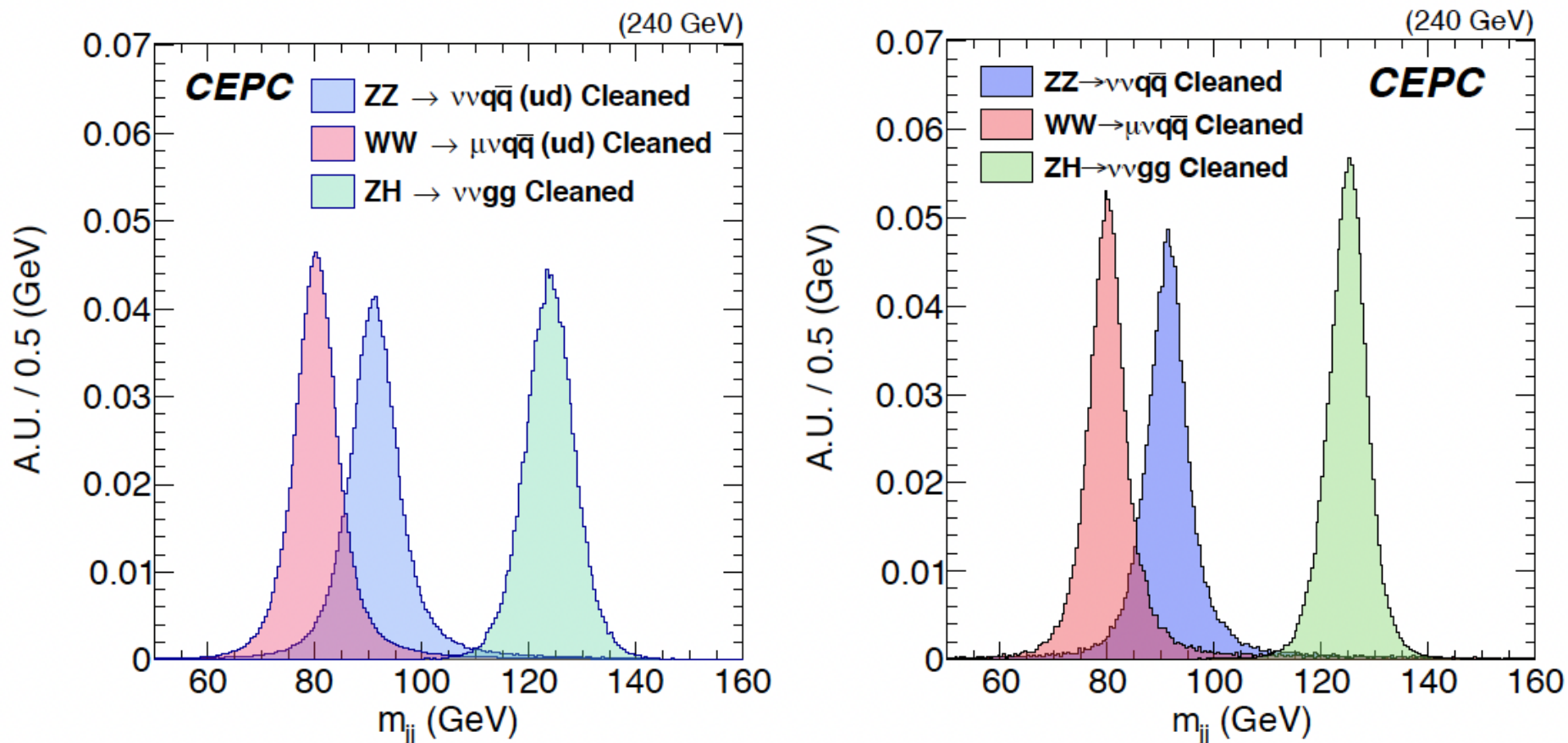
Confusion: frag. Identification & veto



Fake particle originated Confusion reduced by 1 order of magnitude, at nominal vvH , $H \rightarrow gg$ event, at the cost of create mis-vetoed energy of < 1 GeV.

Frag Total Energy (MPV/Mean): 6.3/7.6 GeV \rightarrow 0.7/1.4 GeV

BMR of 2.7% reached

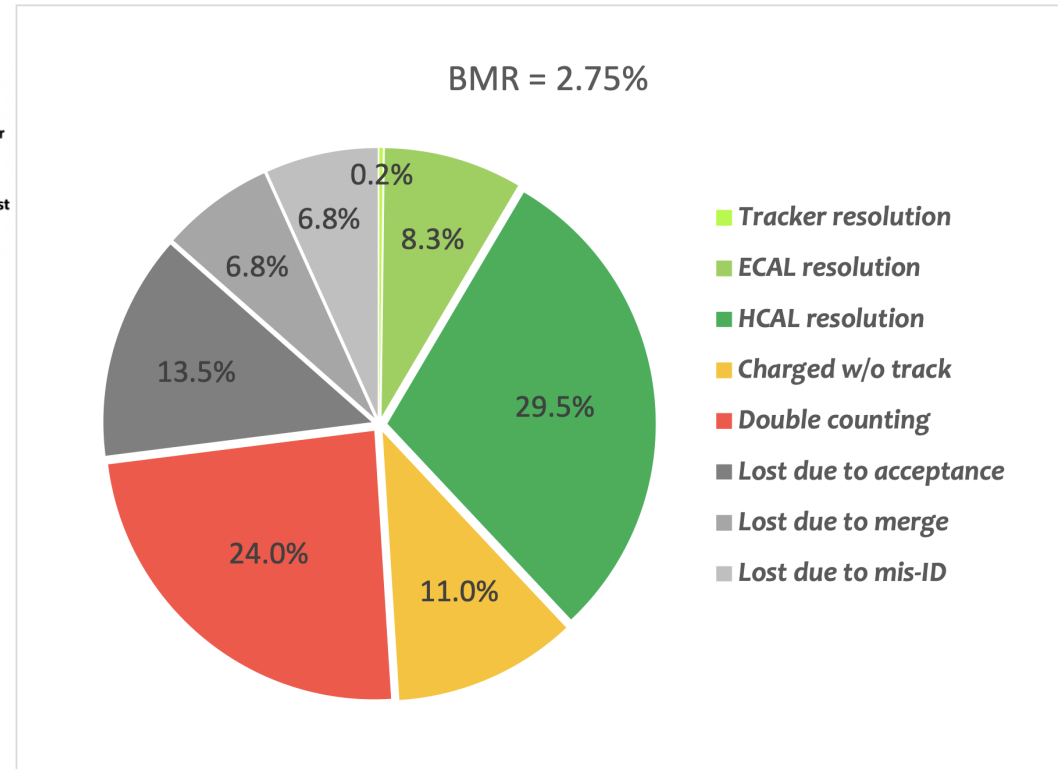
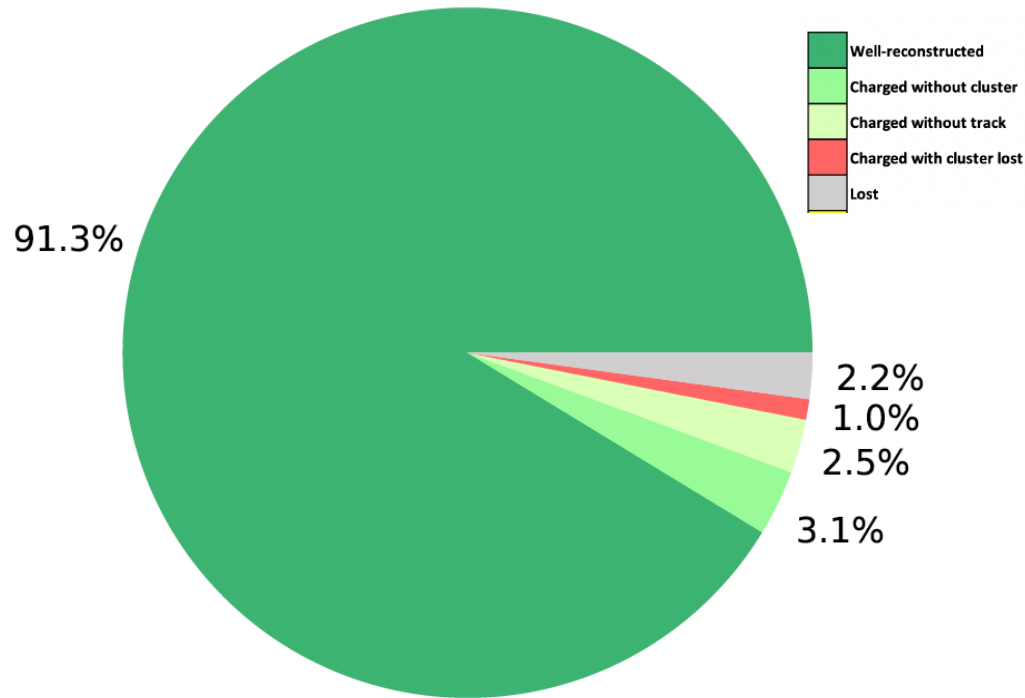


Detector change (usage of high density scintillating glass HCAL): BMR 3.7% \rightarrow 3.4%;

AI enhanced reconstruction: 3.4% \rightarrow 2.8%.

Recent update: further optimization + Pid, etc, current value **\sim 2.6%**

BMR decomposition @ AURORA



1-1 corresponding type: contributing to the BMR via resolution: $\sim o(0.1 - 0.001)$ of its mean value

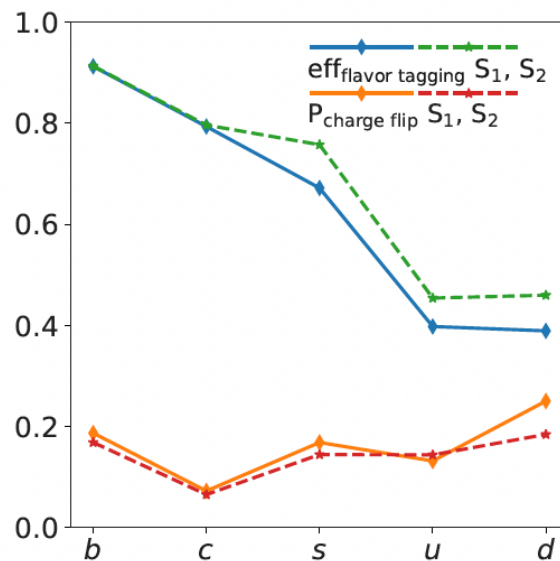
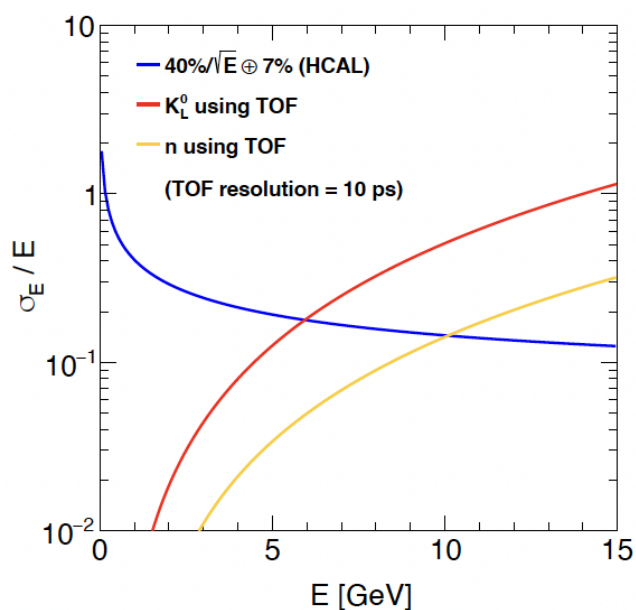
Double Counting & Lost type: contributing to the BMR $\sim o(1)$ to its mean value

Perspectives with 1-1 correspondence

Jet (hadronic events) with Calo

Jet with PFA
Charged in Tracker
Neutral in Calo

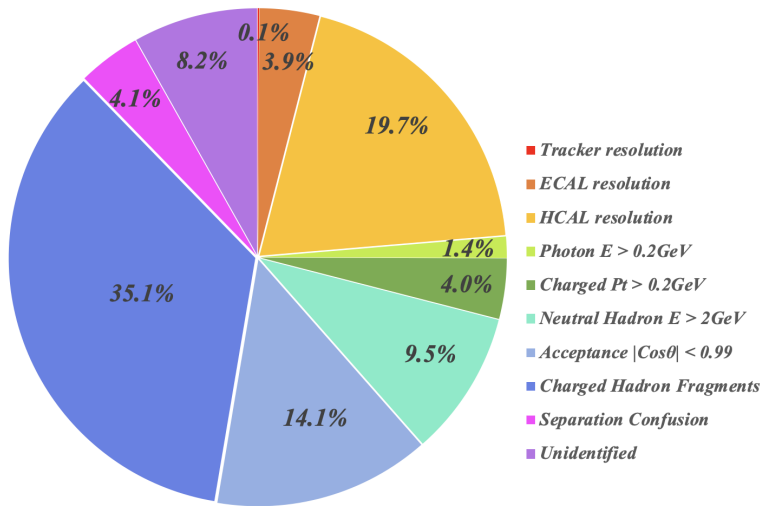
Jet with 1-1
Charged in Tracker + **ToF**
Neutral in Calo + **ToF**



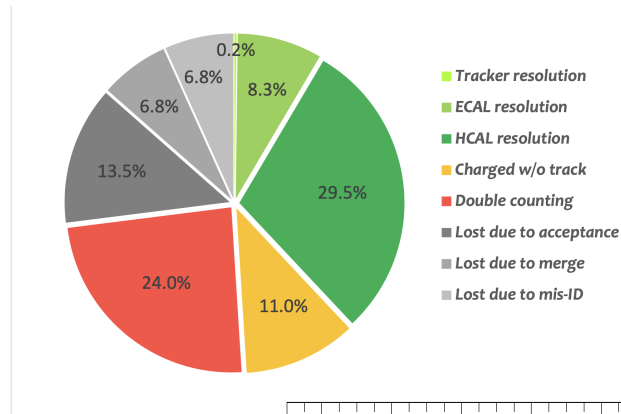
- 5d calo is critical: ToF for all visible particle, thus Pid...
 - Assume Low energy neutrons & secondary particles can be tamed... still challenge...

BMR decomposition & evolution...

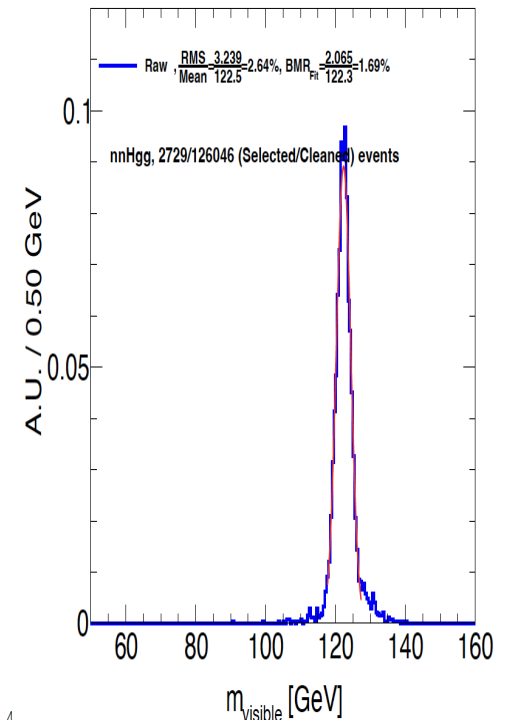
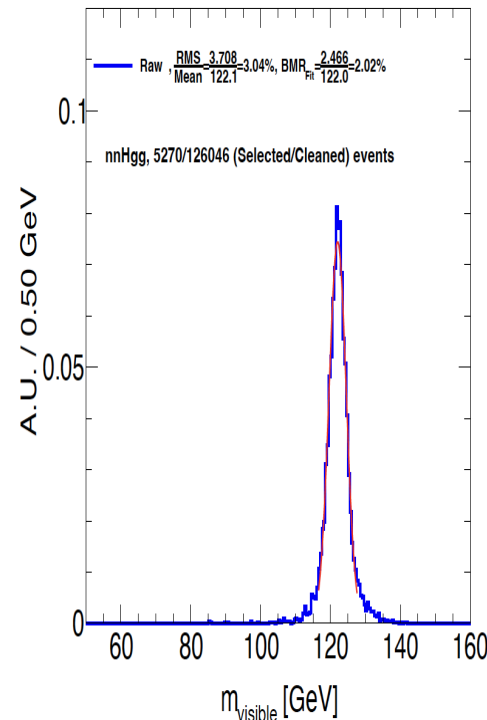
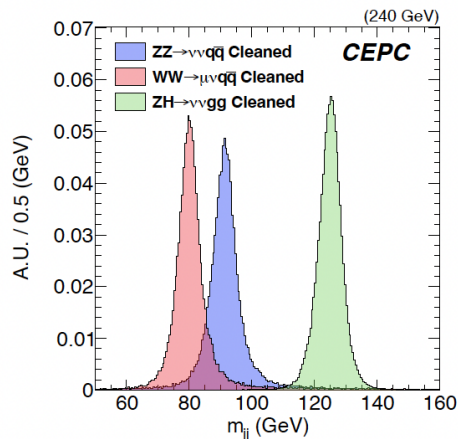
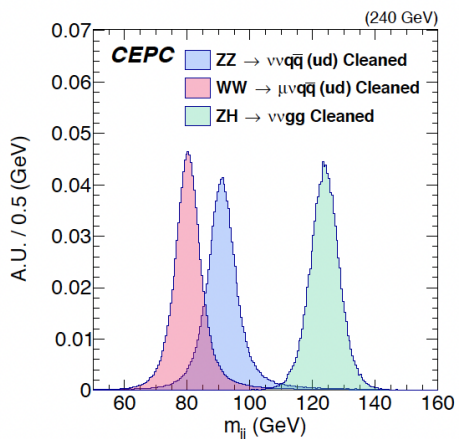
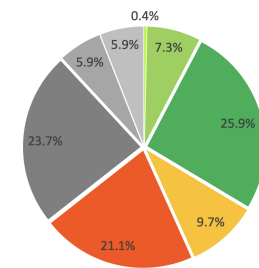
2016 - CDR: BMR ~ 4%



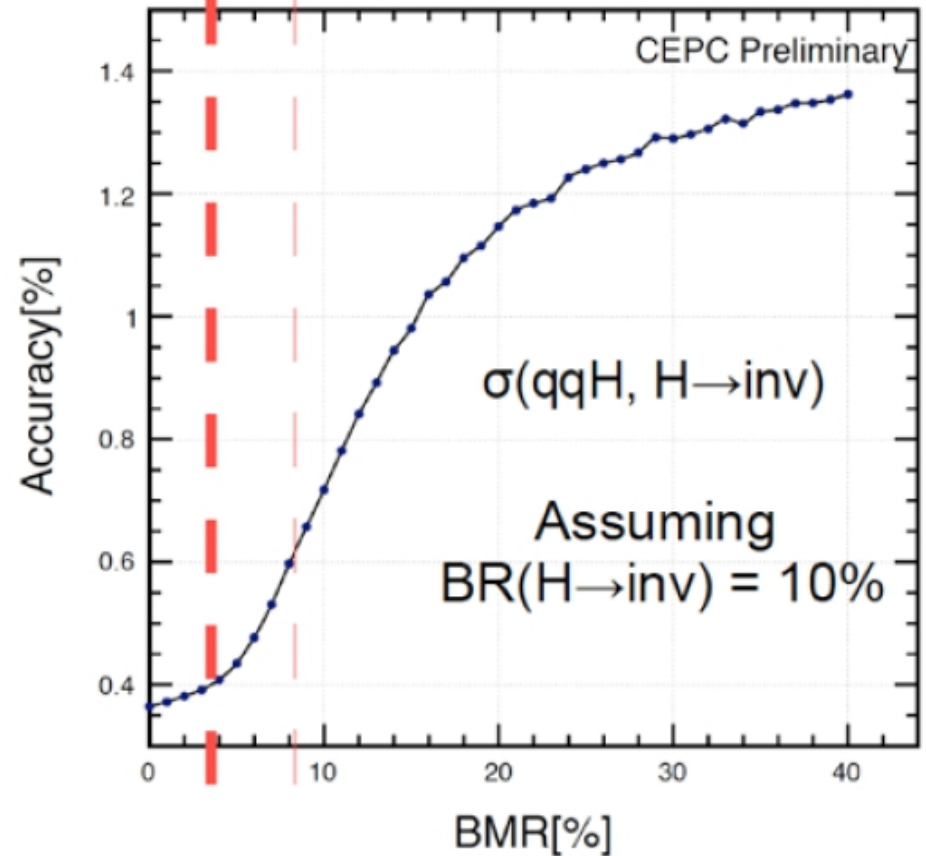
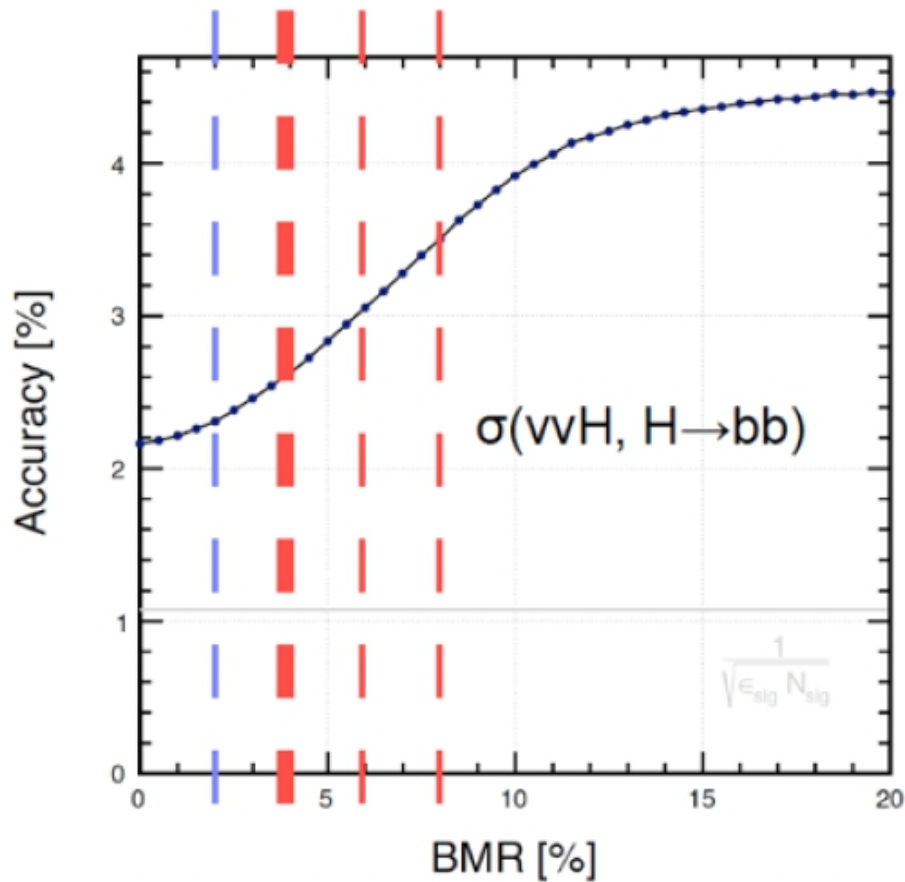
2024 - AURORA: BMR ~ 2.7%



Future: BMR ~ 2.0%



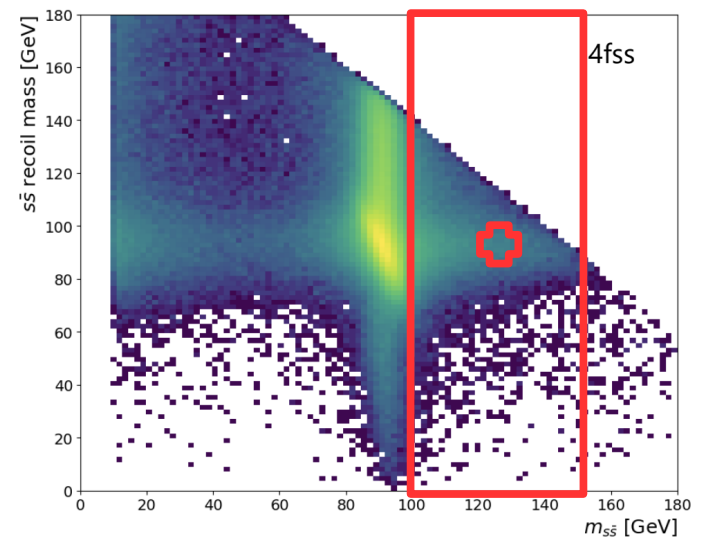
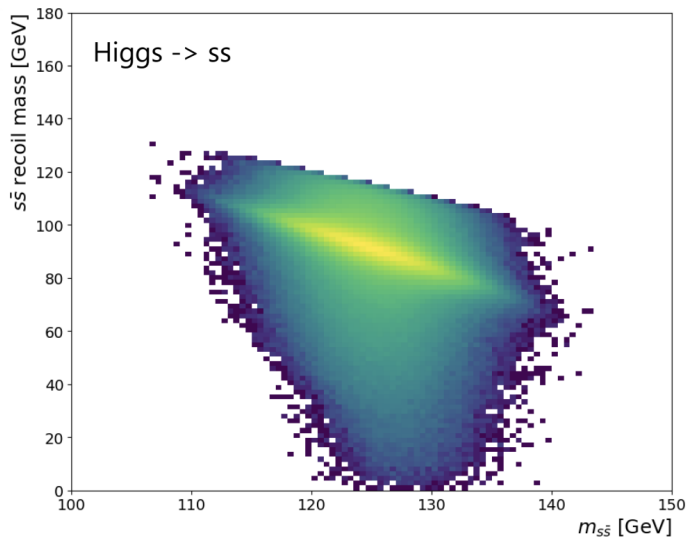
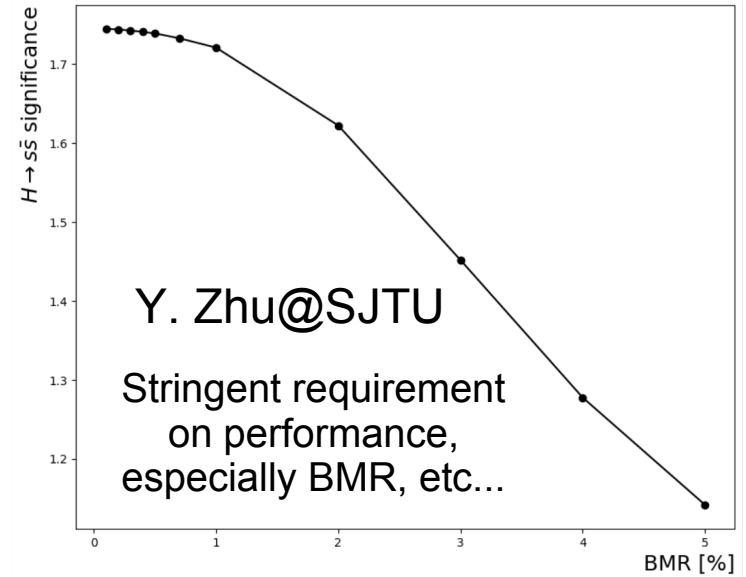
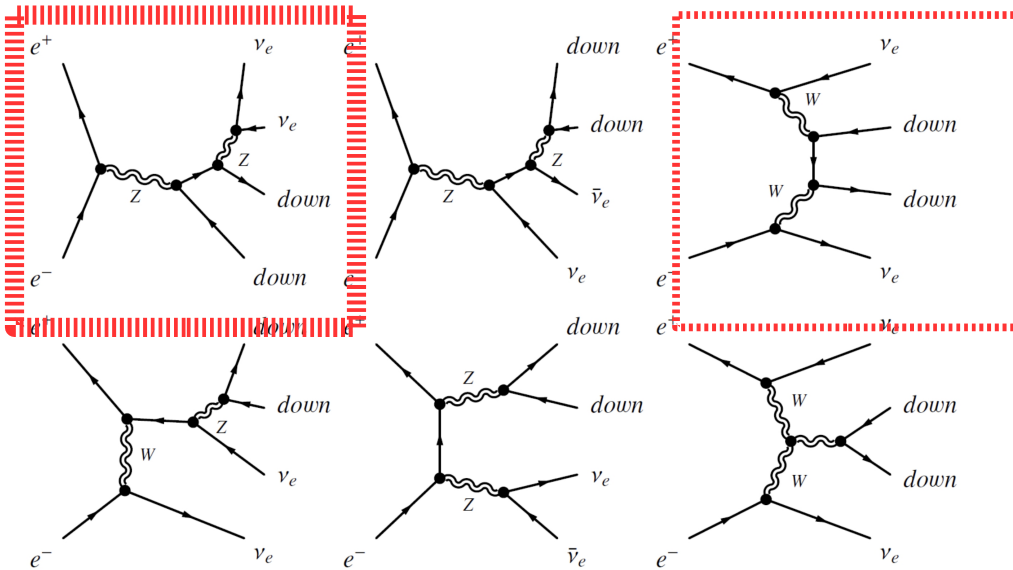
Impact on physics benchmarks...



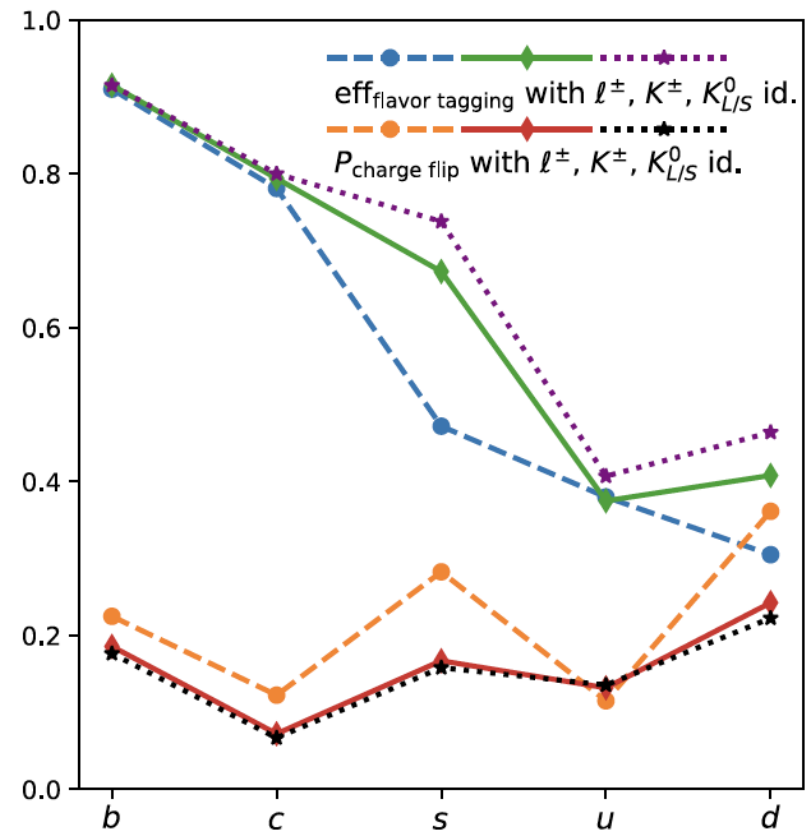
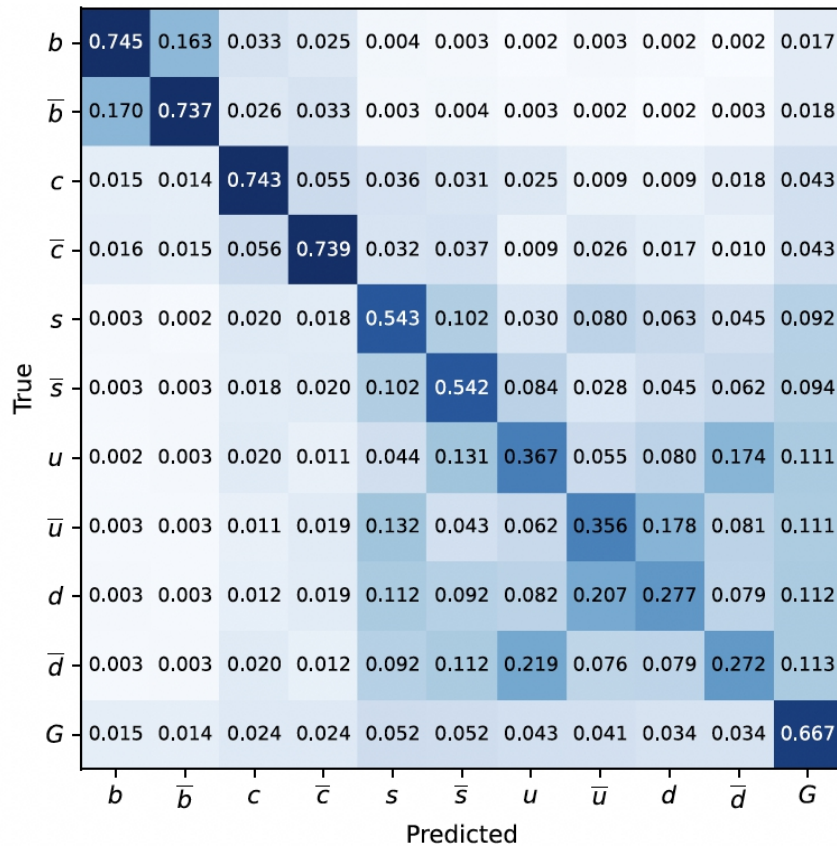
Accuracies of Higgs measurements improved by $\sim 10\%$ with conventional analysis...
Critical for $g(HZZ)$ & new physics detection...

Personal Anticipation: larger impact with sophisticated Analysis, i.e., holistic analysis.

$H \rightarrow ss$: anticipated 3-8 sigmas



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>

A old diagnosis of TRD performance

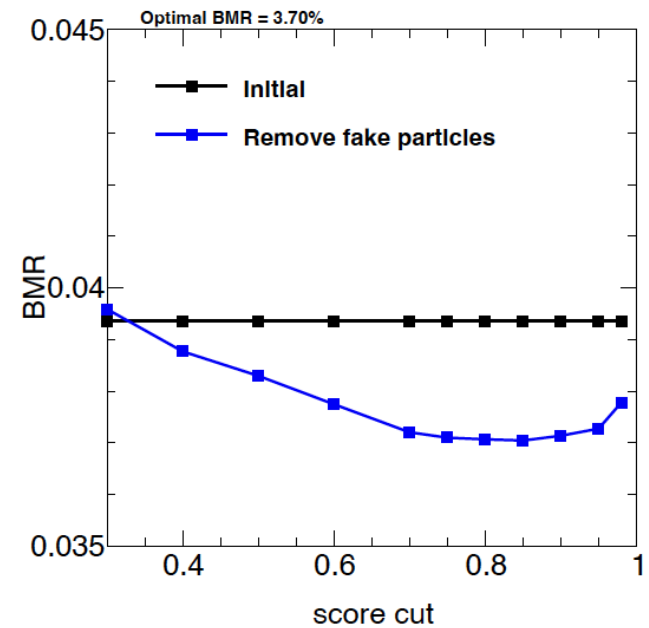
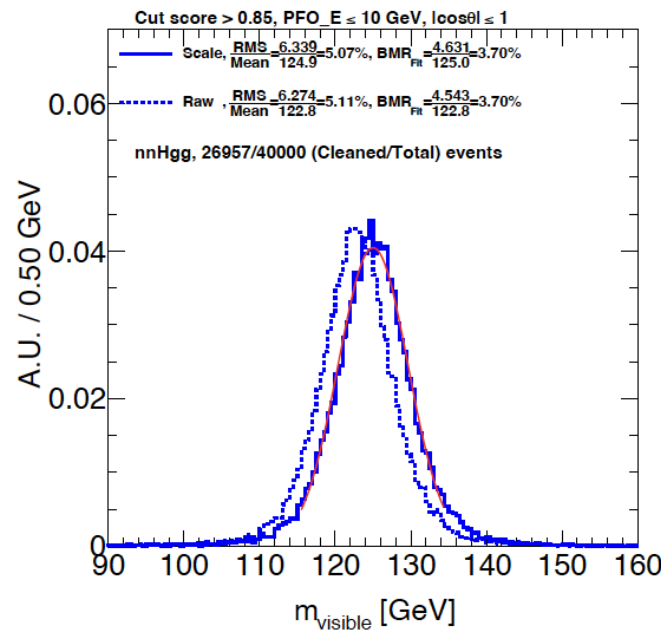
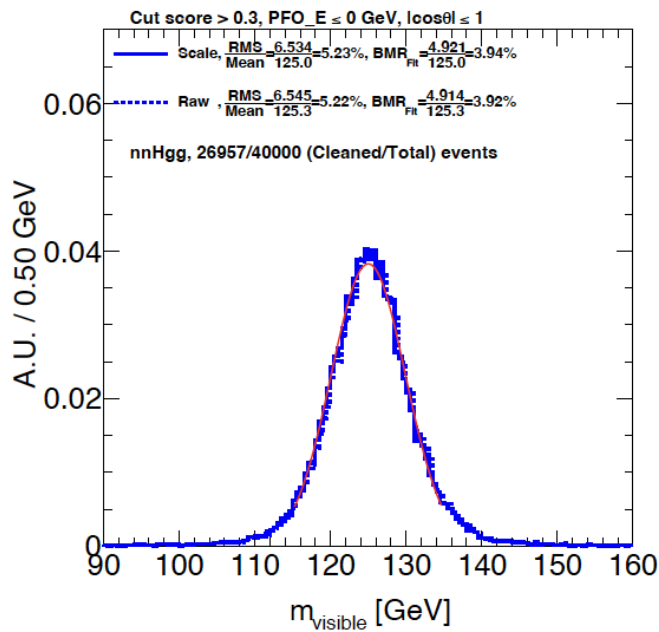
- In the standard of 1-1 correspondence
 - Visible energy decomposition
 - BMR decomposition <https://indico.ihep.ac.cn/event/25992/>
 - Pid
- *This diagnosis needs dedicated MCTruth info, Many Thanks to Fangyi for preparing the sample & update the software*
- *The simulation still has subtitles... especially in the characterize the 2ndary generated in simulation → to be updated.*
- *Based on Ref-TDR setup at Spring, 2025, 2 times more Xstal bars compared to final version*

Ref-TDR ChFrag veto

BMR 3.94% -> 3.7% (rel. 6%)

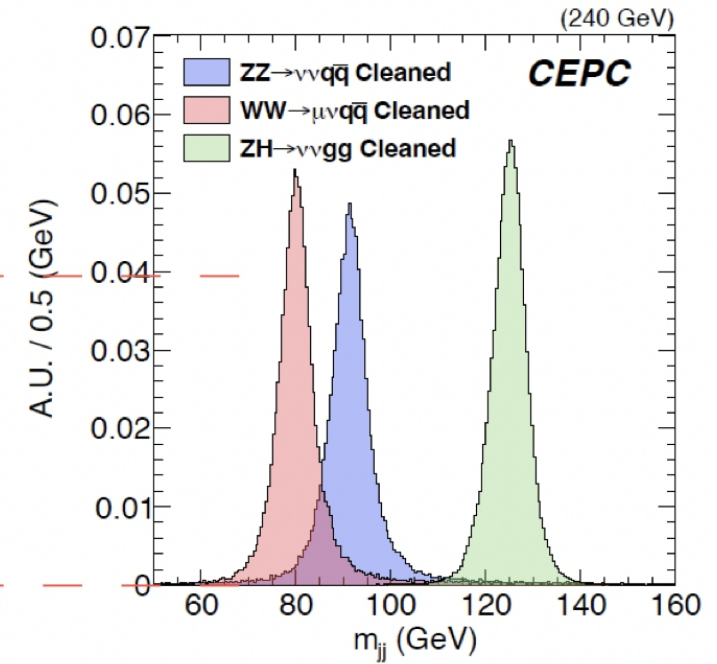
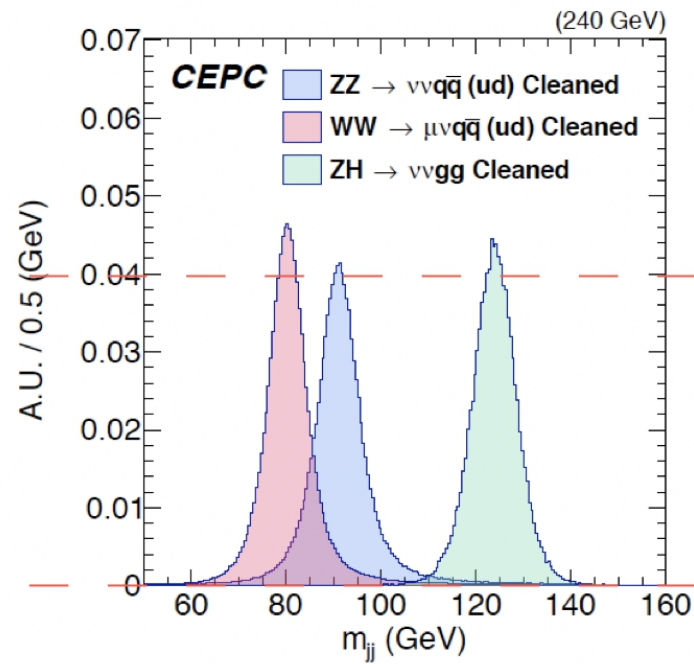
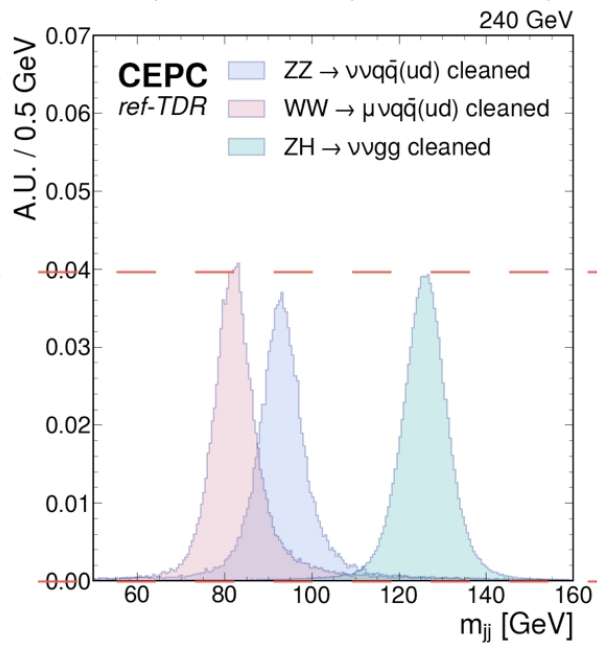
Fangyi's version:

https://code.ihep.ac.cn/guofangyi/cepcsw-release/-/tree/CyberPFA-6.0.8-dev?ref_type=heads
branch CyberPFA-6.0.8-dev

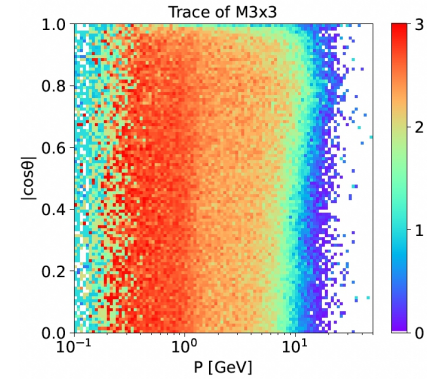
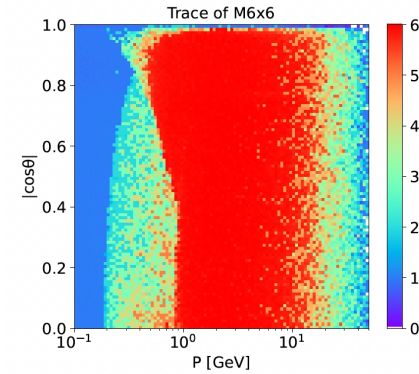
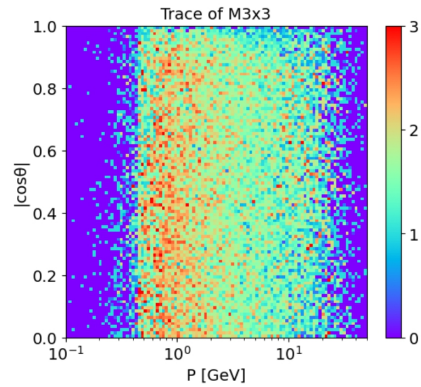
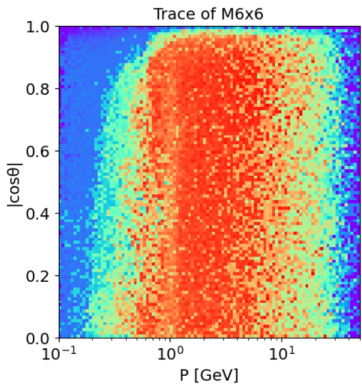
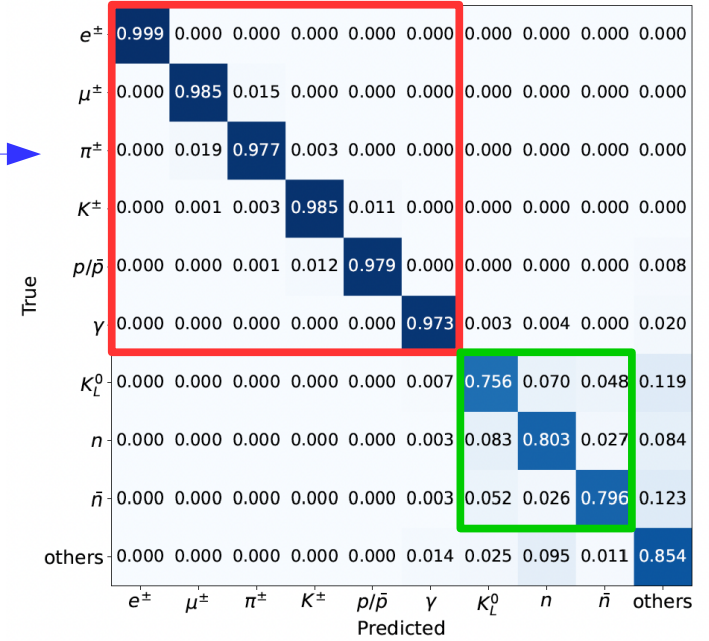
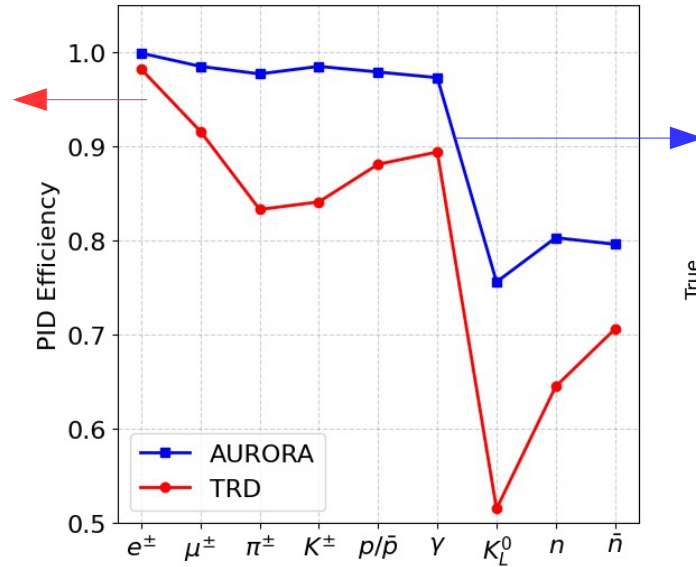
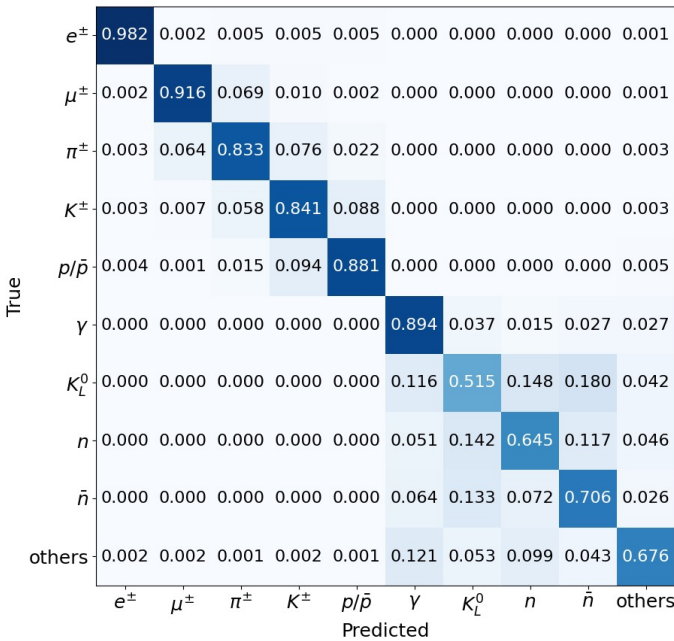


ChFrag Veto: compared to AURORA (3.7/3.4→2.7), much less efficient in TRD as the leading bottleneck is not the fragments

BMR comparison



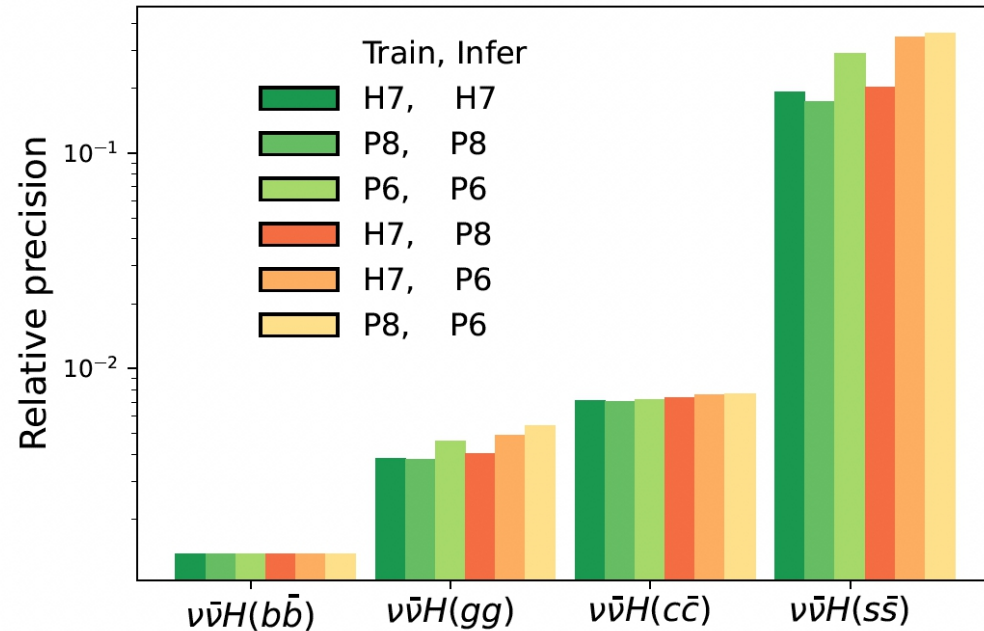
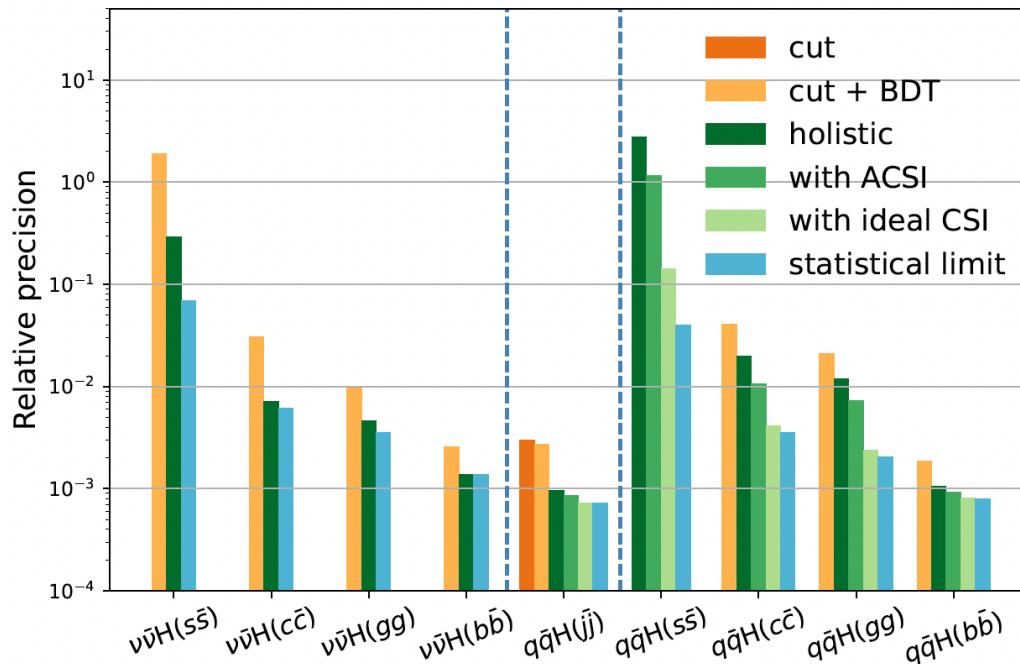
Pid



Kaon id: TDR has larger inner TPC radius. To be verified & confirmed quantitatively.

Lepton & neutral Kaon id: relatively limited info. From ECAL in TRD.

Holistic approach + ACSI

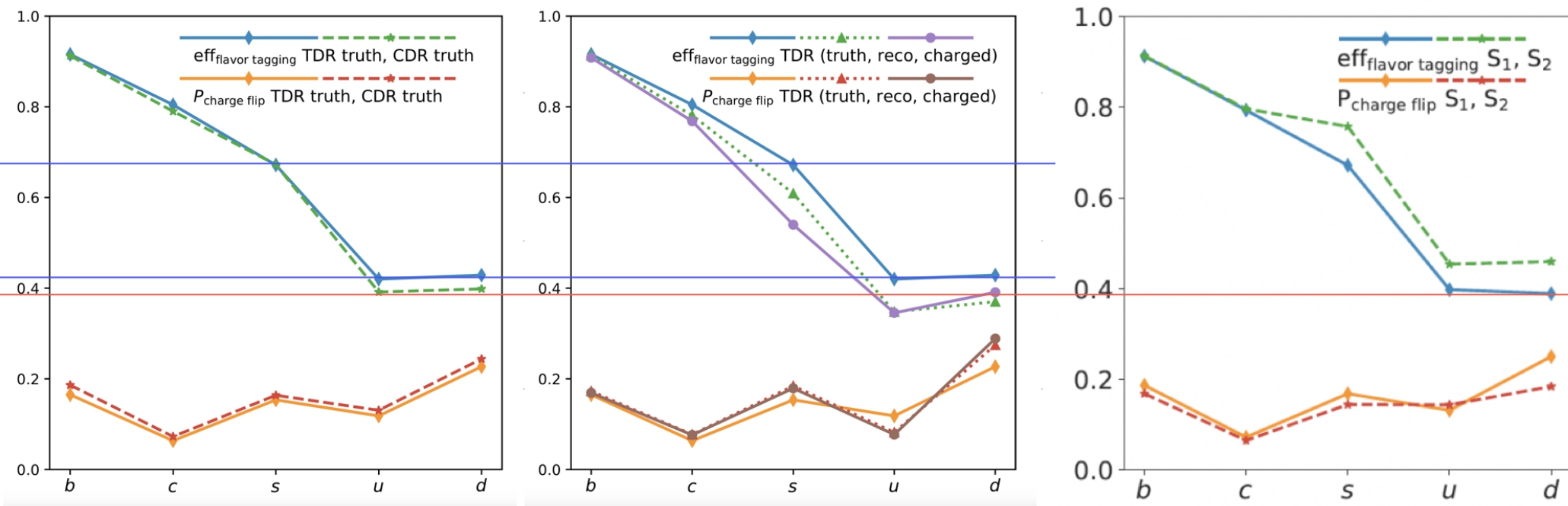


Holistic + ACSI: improves the accuracy by 2 – 6 times,
in principle free from human intervene (if the simulation is good enough...)
Strongly depends on Hadronization modes...

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

$H \rightarrow s\bar{s}$ within the reach at toy analysis: clarify the conditions to confirm this decay mode.

Jol at TRD, CDR & AURORA (ideal)

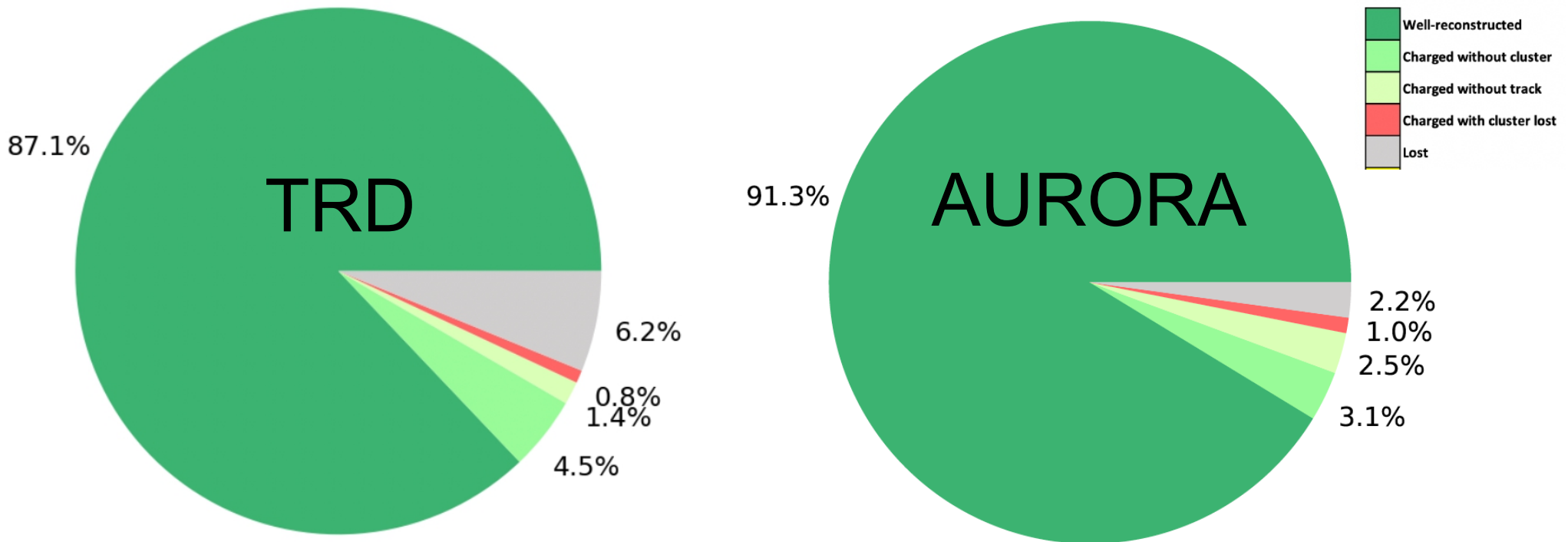


Using truth Pid, TRD has better Jol than CDR detector, as it uses longer Barrel + stitching VTX

Pid at TRD is limited, will degrade the H→ss measurement... (software version 0401, not 1-1)

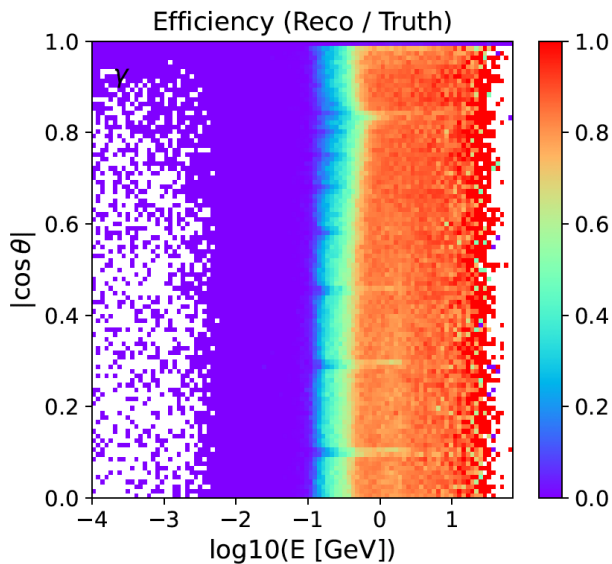
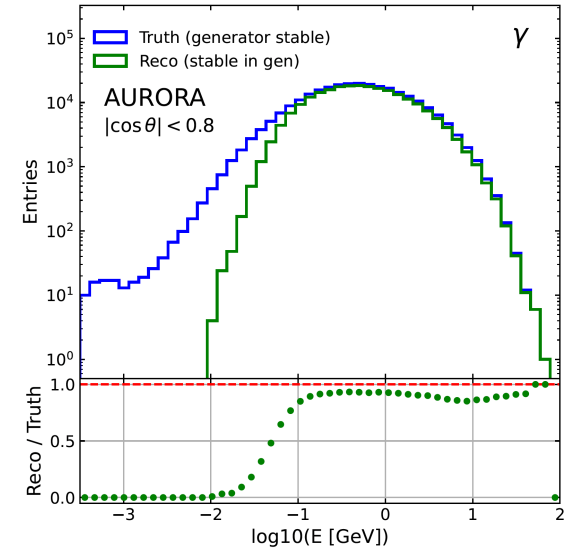
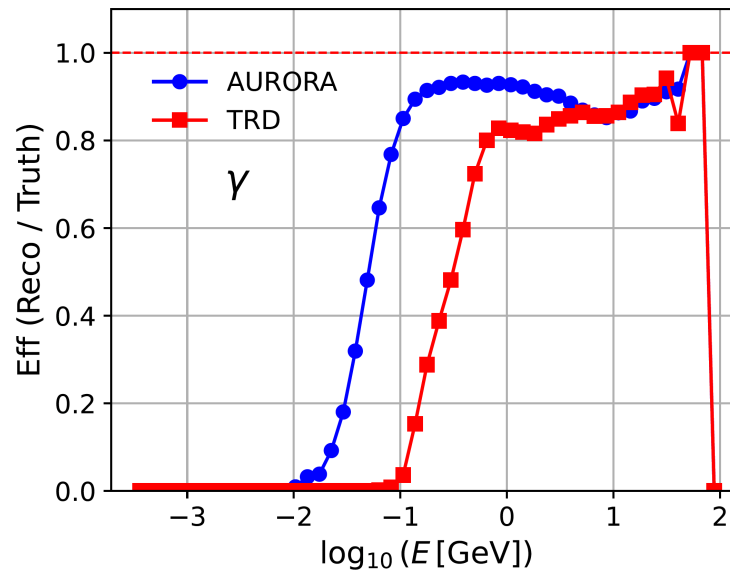
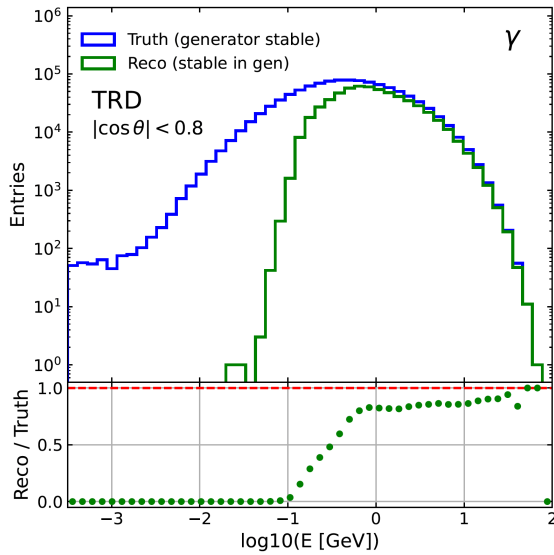
Neutral Hadron ID has strong impact on Light Quark ID: highly appreciated in H→ss

Visible Energy

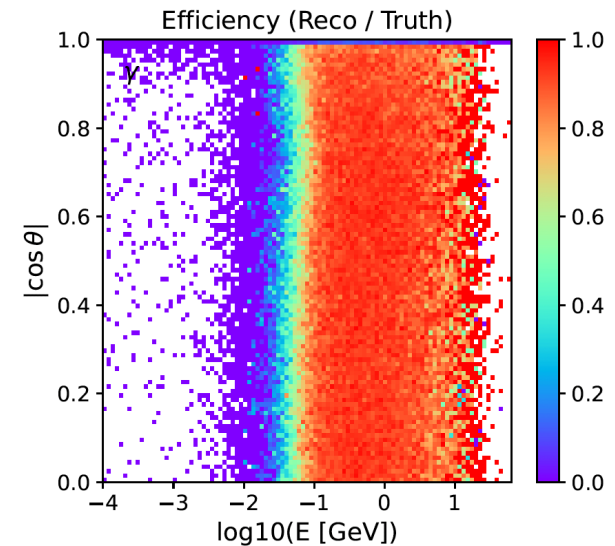


- TDR Ref-det:
 - Ch. wo. Cluster: increased by 50%
 - Ch. wo. Track: reduced to half (Tracking in TRD is actually better)
 - Ch. wi. Cluster Lost: (Double counting) reduced by 20%
 - Lost contribution increased **3 times** (**5 time** if subtract 1% of irreducible Lost due to Acceptance)

Photon reco. efficiency

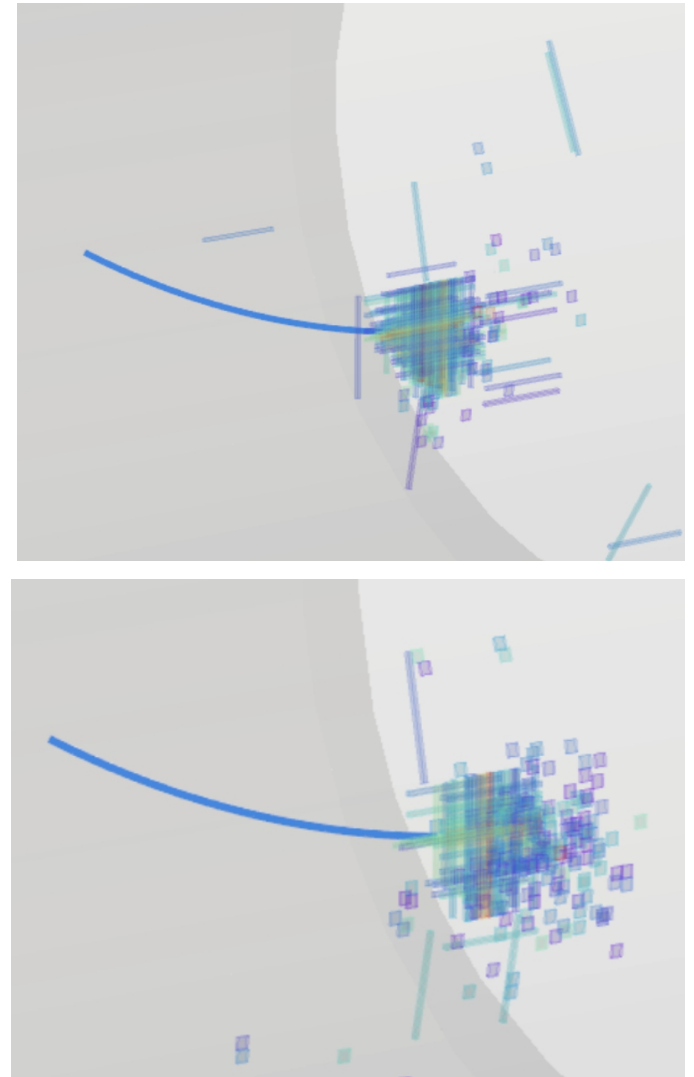


- Converted photon included,
- 10 GeV valley caused mainly by photon merging in pi-0

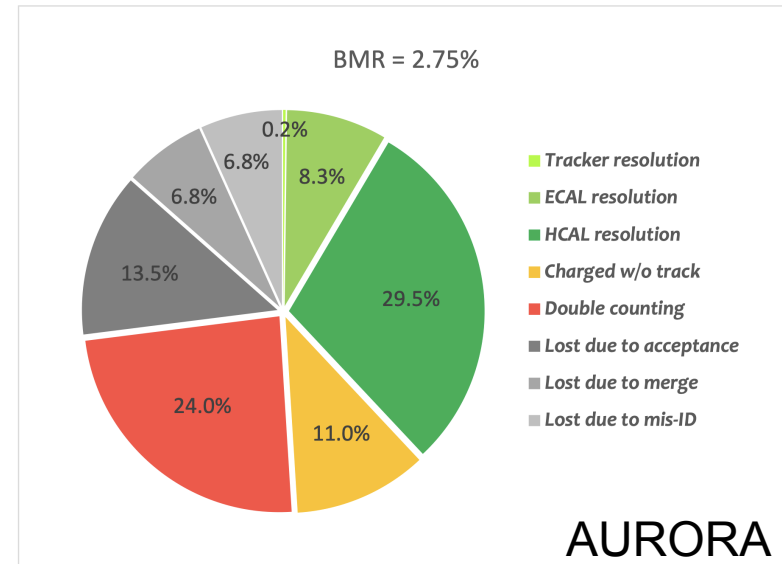
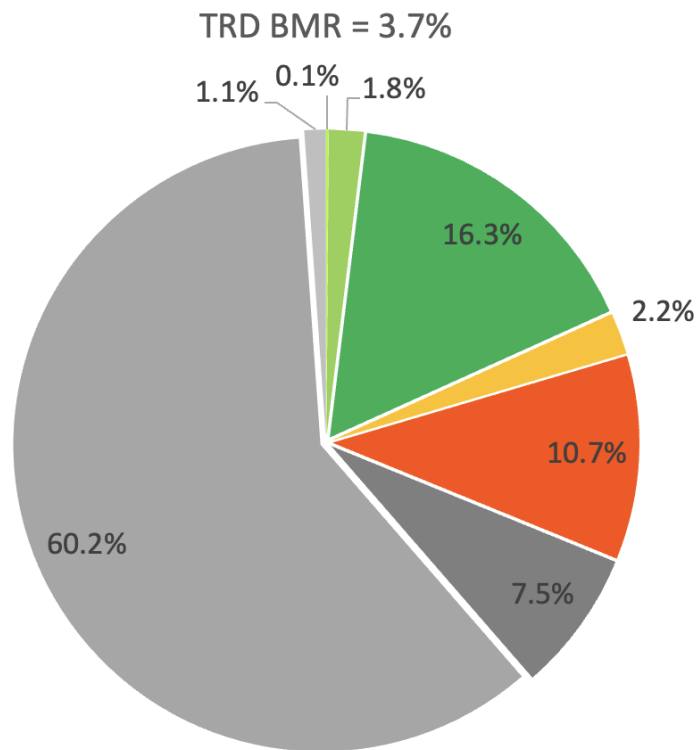


Thoughts on the Det. optimization

- Si-W ECAL: better BMR & Pid
- Xstal ECAL: excellent EM resolution
- 5-d calorimeter is appreciated
- In TRD, the bottleneck is the inefficiency of cluster reconstruction, esp. neutral particles in the jets. Primarily due to the fact that Xbar configuration has large shower volume, causing severer shower overlap – merging
- The current reco need to strength its ability neutral particle reco. While scaling behavior V.S. the bar length & B-Field could be a good starting point.



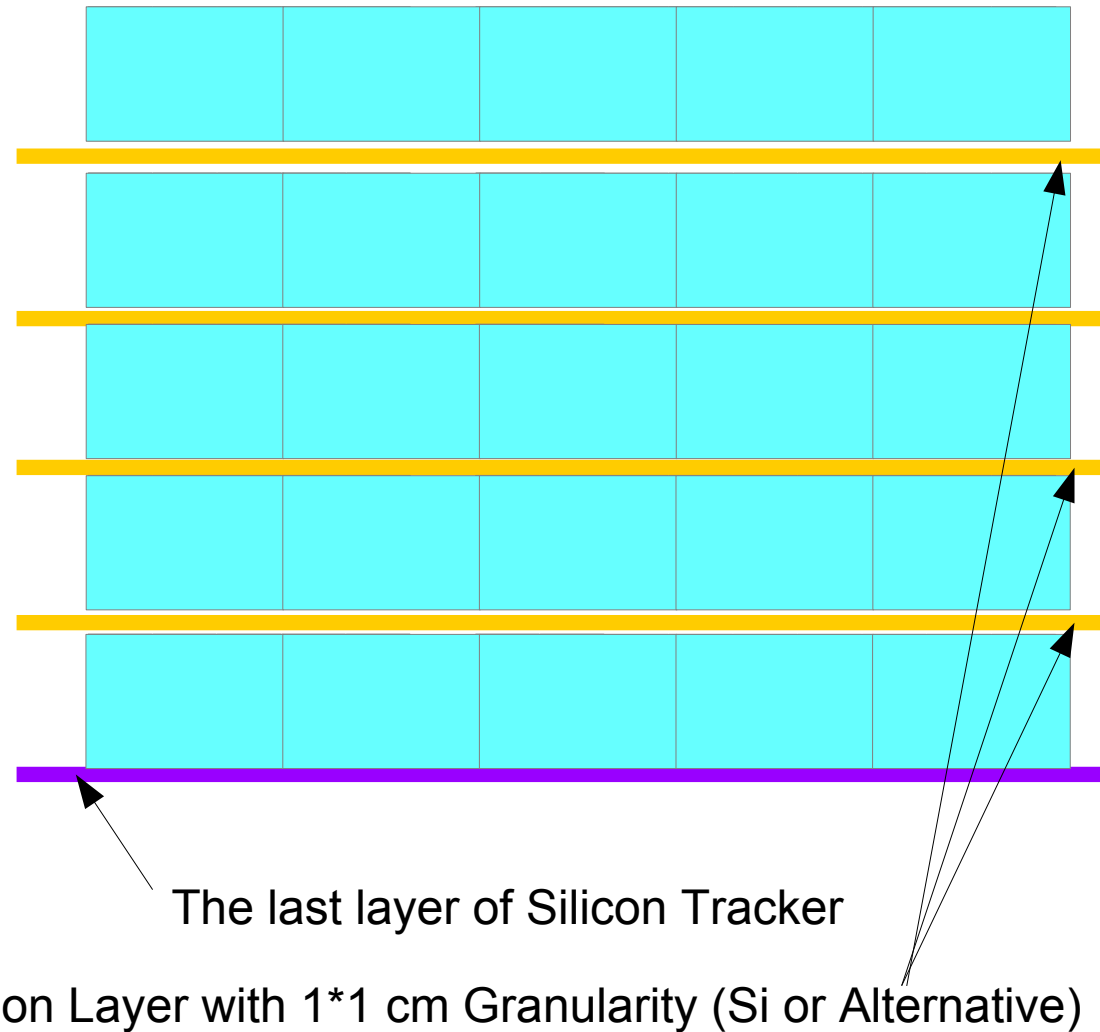
BMR decomposition



- TRD decomposition: Scaled from AURORA model
- Leading item: Lost due to merge & inefficiency - estimated from two independent methods.
 - Lost Truth Level Particle
 - Lost Total Energy (in taking into account the Double Counted ones).

Design-3: Crystal Tile + Mesh

- Geometry
 - Single Crystal Tile Dimension:
 $6\text{cm} * 6\text{cm} * 6\text{cm} = 216\text{ cc}$
 - Number of Tiles $\sim 110\text{ k}$
 - Inner Area: 80 m^2
 - Total Readout Channel:
 - $110\text{k} * (1, 2, 4)?$ (Crystal)
 - $800000 * 4 = 3.2\text{ M}$ (Si)
- Comments
 - Should quantify the inhomogeneity response with SiPM couple to larger volume Tile

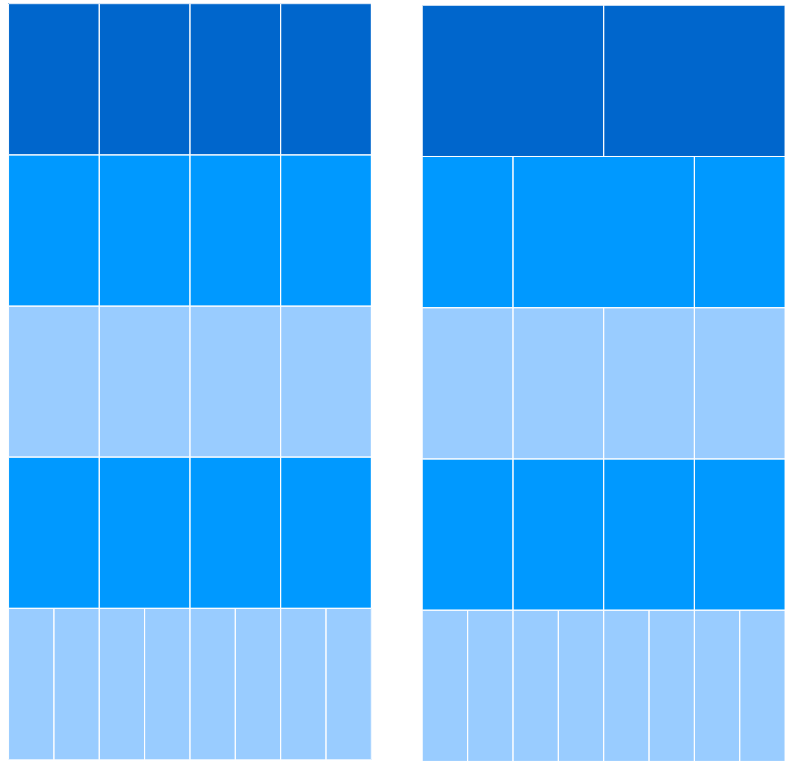


Summary

- We propose and realize the concept of 1-1 correspondence reconstruction
 - Within the reach of current technology & strongly boost the discovery power
 - A novel standard to quantify the global detector performance
- BMR achieved 2.7% at AURORA (CDR detector + GSHCAL), with a future perspective of 2%
 - Roadmap established & Needs lots of developments
- Diagnosis with 1-1, TDR Ref Det
 - Improves its BMR to 3.7% (relative. 6%)
 - The bottle neck is “LOST particle”: inefficiency to reconstruct neutral particles
 - Has limited Pid performance, but better tracking (esp. Low momentum ones). Comparing AURORA with TRD provides quite some insights & possible synergies.
 - The granularity & configuration of calo, especially ECAL, should be re-optimized, iterated with Reco. Tool development
- Possible scenarios: share the Positioning + timing, and the energy measurements task between Si/Xstal: worth quantitative study

Back up

Design-1: Crystal/Glass pillars



TDR Total Xstal Volume $\sim 24 \text{ m}^3$.

Conceptual para: 5 layers:

First Layer: $1 \times 1 \times 6 \text{ cm}^3$

Last 4 layers: $2 \times 2 \times 6 \text{ cm}^3$

Total readout channels 1.6 Million

Compared to $\sim 570 \text{ k}$ SiPM readout in Xstal Bar: 7% increase of the ECAL cost.

Needed study: EM resolution evolution with increasing of longitudinal seg (gap, mech, cooling...)

Similar structure could be also used in HCAL, significantly reducing the channel num.

Full absorption HCAL could tolerant much larger cell size, at the cost of glass & total volume increase $\sim 10\%$ of the cost.

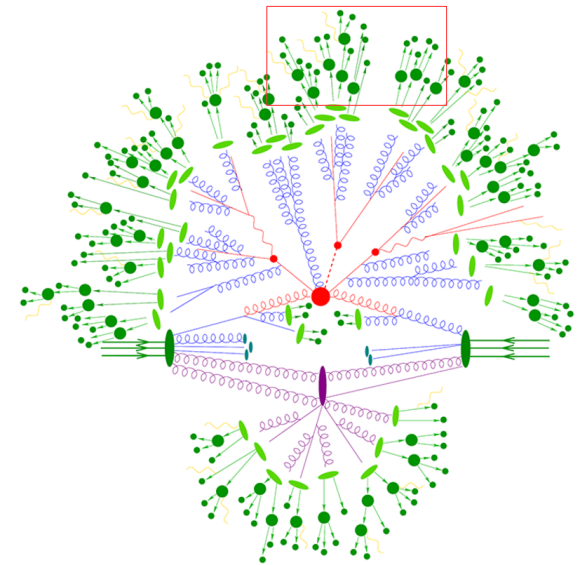
IP Full Glass ECAL should also be explored, with much more readout channels.

Warning: No free lunch!

- **Key Problem: Data – MC Discrepancy**, especially for supervised learning
- Solution
 - Better Simulation (inject more attention & information!)
 - Robust AI tools less dependent on Simulation: From supervised learning, to non/weakly-supervised, enhanced, LLM...
- **Hadronization & impacts...**
- **Beam induced backgrounds & machine protection**
- 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 & 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 Noise: rates, intrinsic, and radiation relevant ones
- Calibration Procedure & Monitoring methodologies...
- ...

Targeted studies

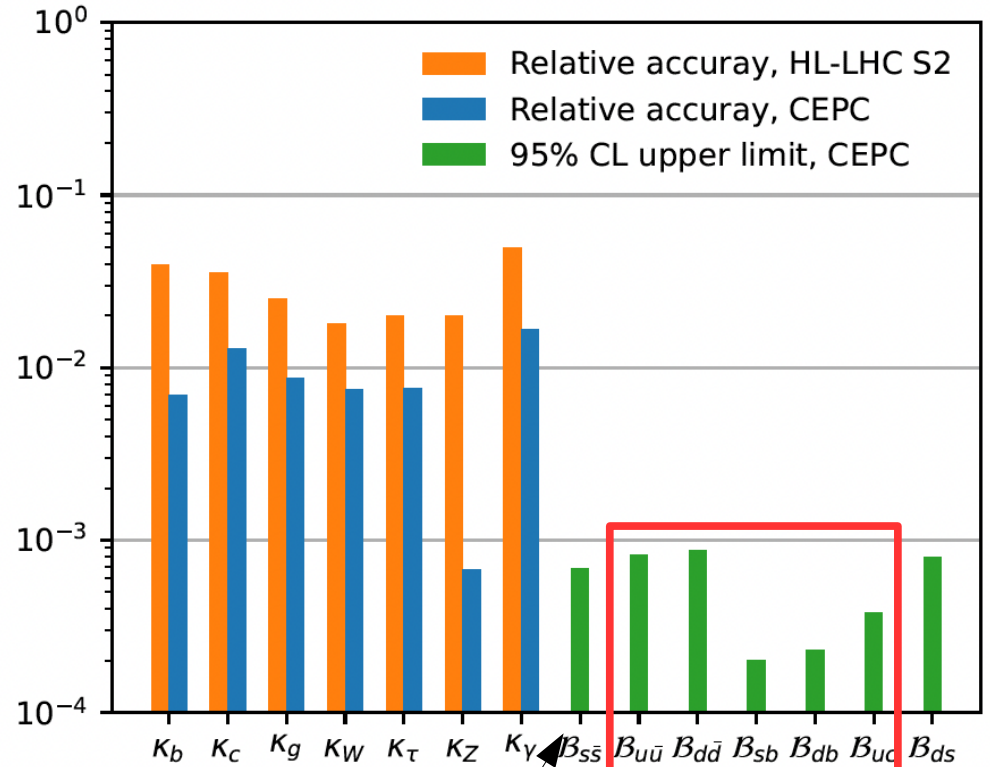
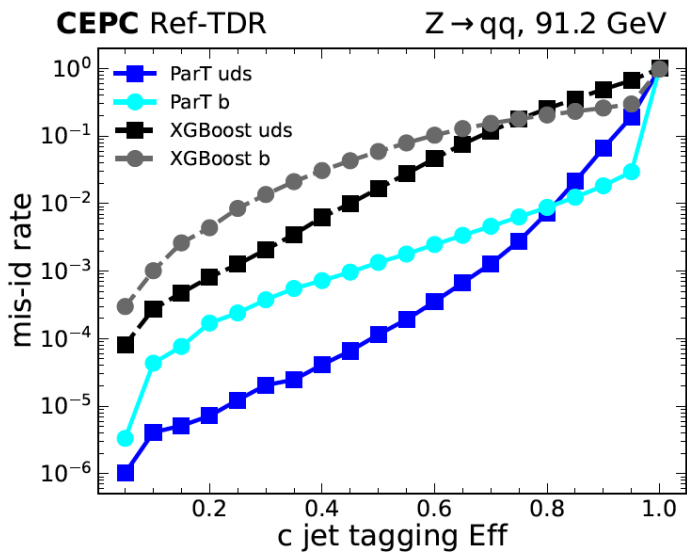
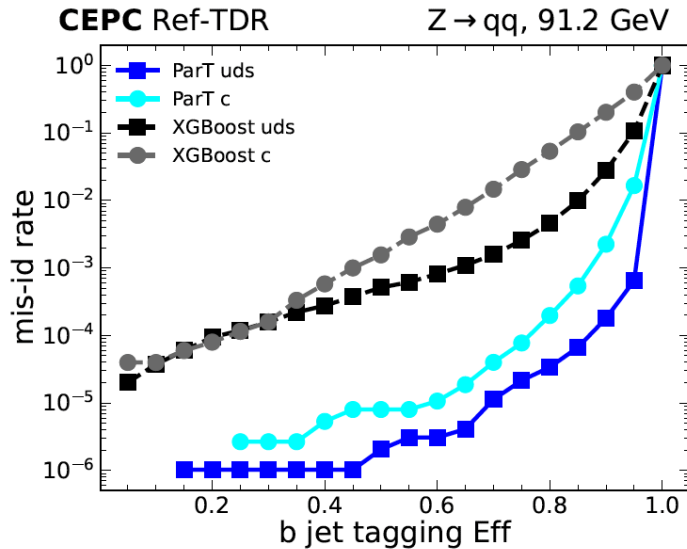
- **Usage of Timing**, especially at 5d calorimetry
 - **Clustering with time**
 - **PFA of the Space Time**... to reconstruct all final state particles in the data stream continuum, and correctly associate them with different VTX & sources, including Beam Induced Background
- **Holistic approach with sufficient categories (Multi-class identification)**
 - $\sim o(100)$ categories is sufficient... to identify almost all physics events at a Higgs factory
 - $\sim o(10)$ is probably sufficient for Z factory
 - Free of human intervene – in principle
 - In progress & to understand the scaling/optimization
- **Color singlet identification & iteration with QCD studies**
- **From leaves to the Tree:**
 - Id the parenting info of final state particles; i.e., $B \rightarrow D \rightarrow K \rightarrow \pi^0 \rightarrow \gamma \dots$
- **Longer term: Learning from all particle \rightarrow hits**
- ...



Summary

- ... Higgs factory **should and could** have excellent detector & reconstruction...
- AI enables us to **learning from all the particles**: enhance the discovery power ~ 3 times...
 - Holistic approach
 - Reco: Jet origin id, 'see' the quark & gluons...
 - Analysis: Processing in principle free from Human intervene.
 - + ACSI for full hadronic events
 - 1-1 correspondence reconstruction: excellent PID + BMR of 2.7%
 - 5-d calo is the key
 - Long term: learning from all hits
- **Multiple challenges, intriguing prospects, and significant synergies.**
 - Precise Simulation is critical, need profound understanding & development
 - To explore other methodologies: non/weakly-supervised, enhanced, LLM...
 - Lots more to explore, with unsupervised, LLM, ... rich interplay & synergies.
 - ...

Impact on Physics

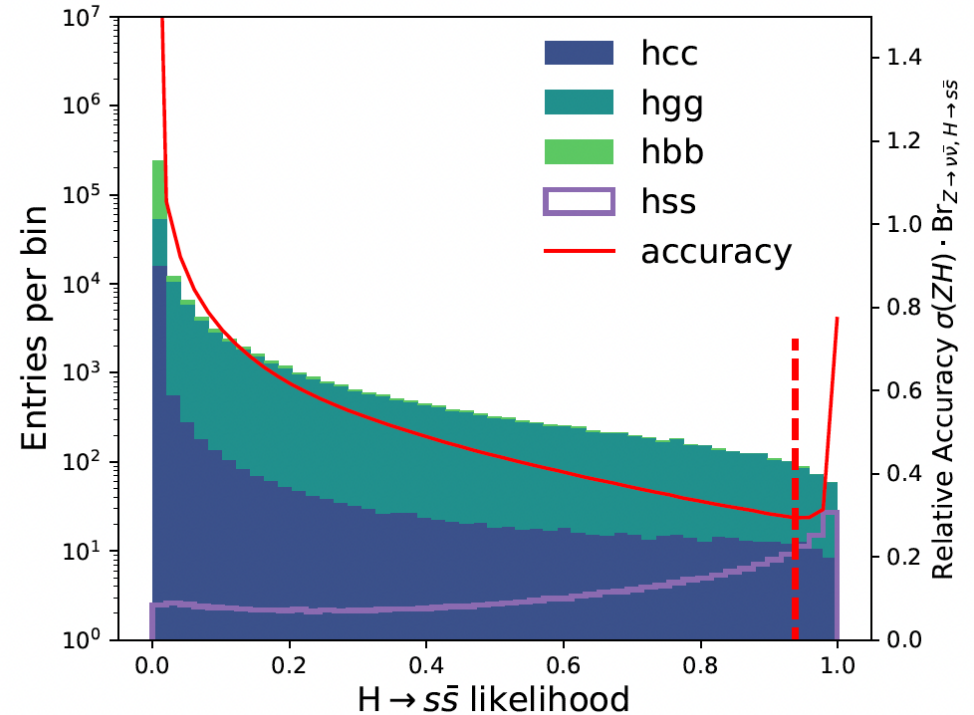
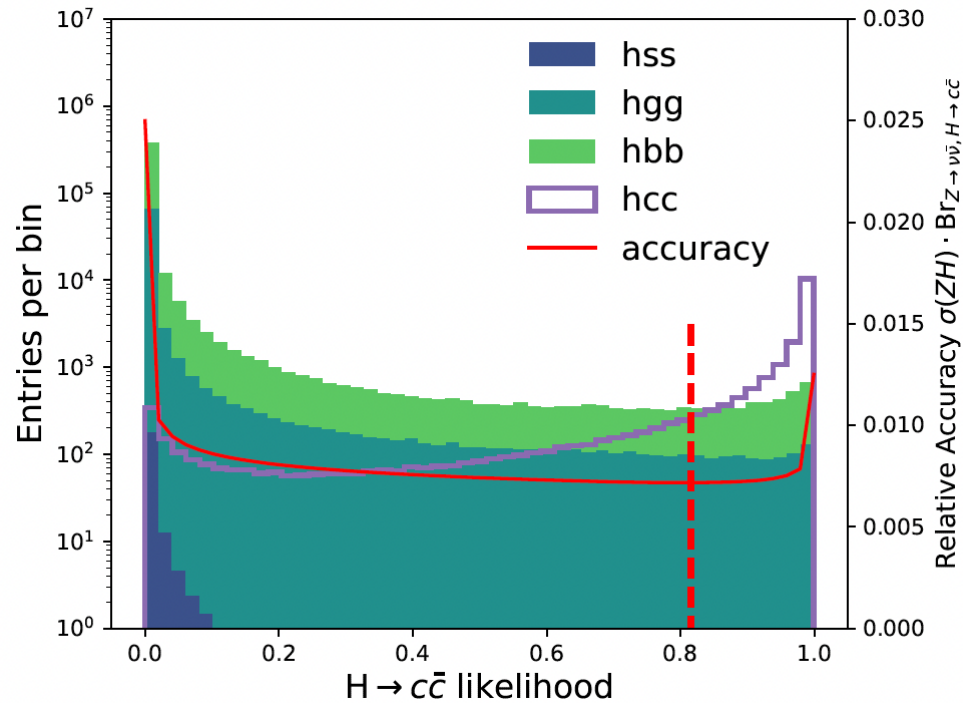


Improved by ~ 3 times

Improved by 1-2 orders of magnitudes
 ~ 2 folds improvement at $H \rightarrow cc$ & V_{cb}

Presumably... firstly quantified

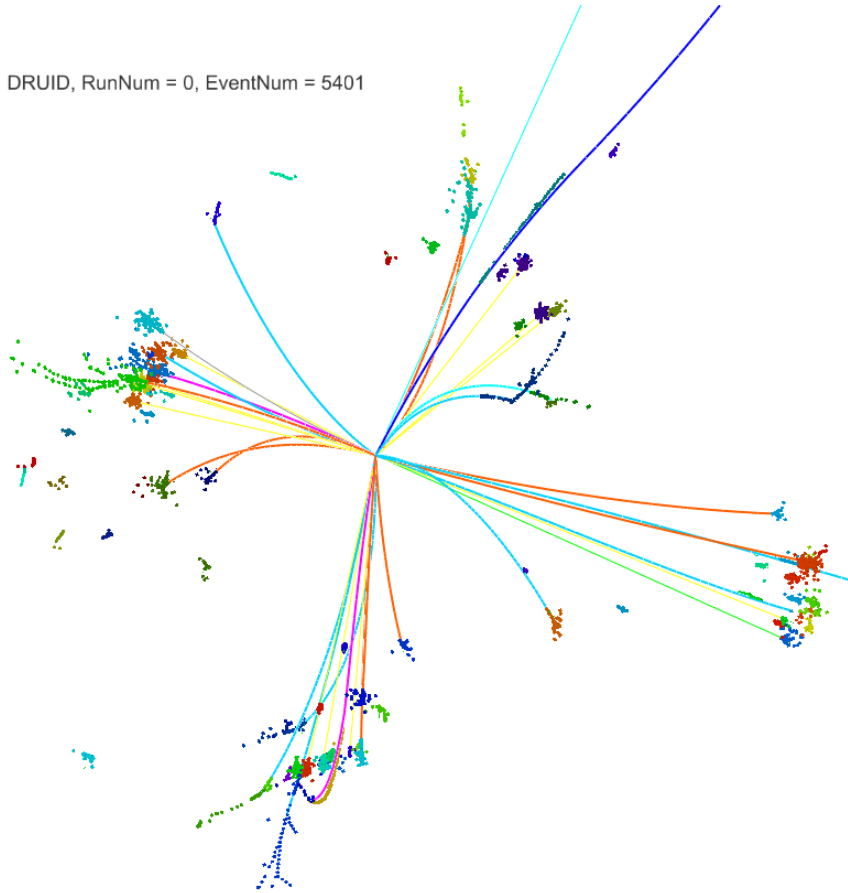
Holistic Analysis: $\nu\nu H$, $H \rightarrow 2 \text{ jet}$



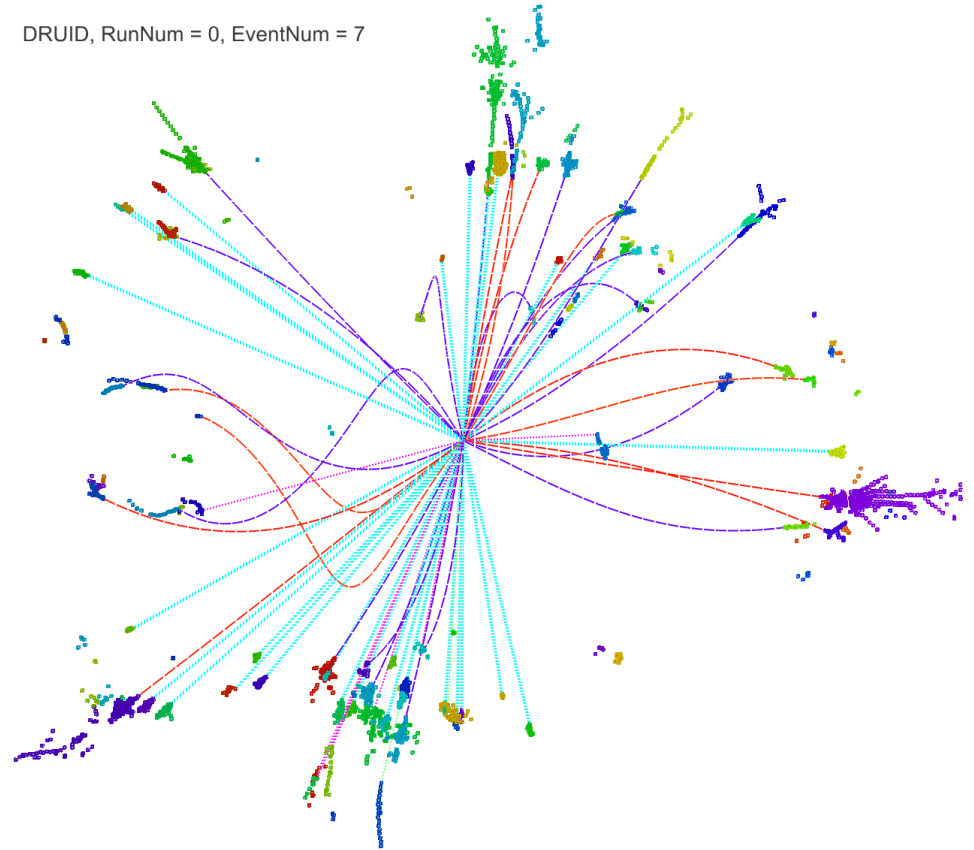
- $\nu\nu H$, $H \rightarrow bb/cc/gg/ss$ measurements: 4 kinds classification
- Simplified analysis with irreducible background...
- $H \rightarrow bb/cc/gg$: close to the statistic limits - 2-6 times better than previous studies (include other bkgrd, BDT based, etc)
- $H \rightarrow ss$: close to confirmation!

Color Singlet Identification

DRUID, RunNum = 0, EventNum = 5401

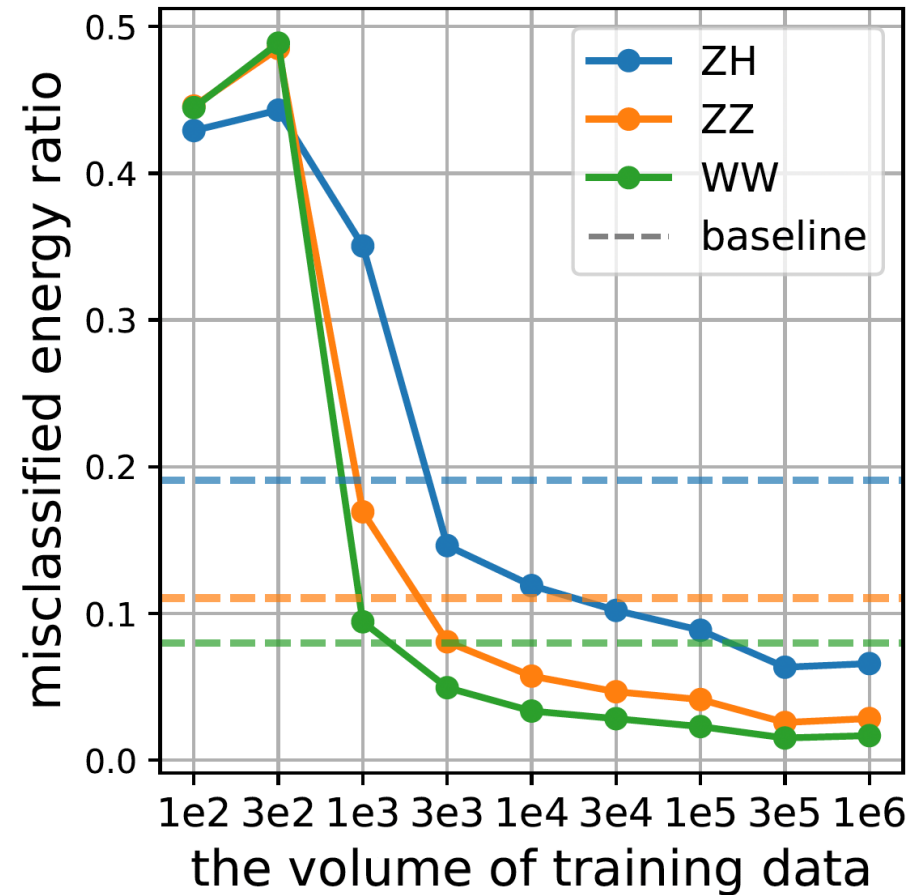
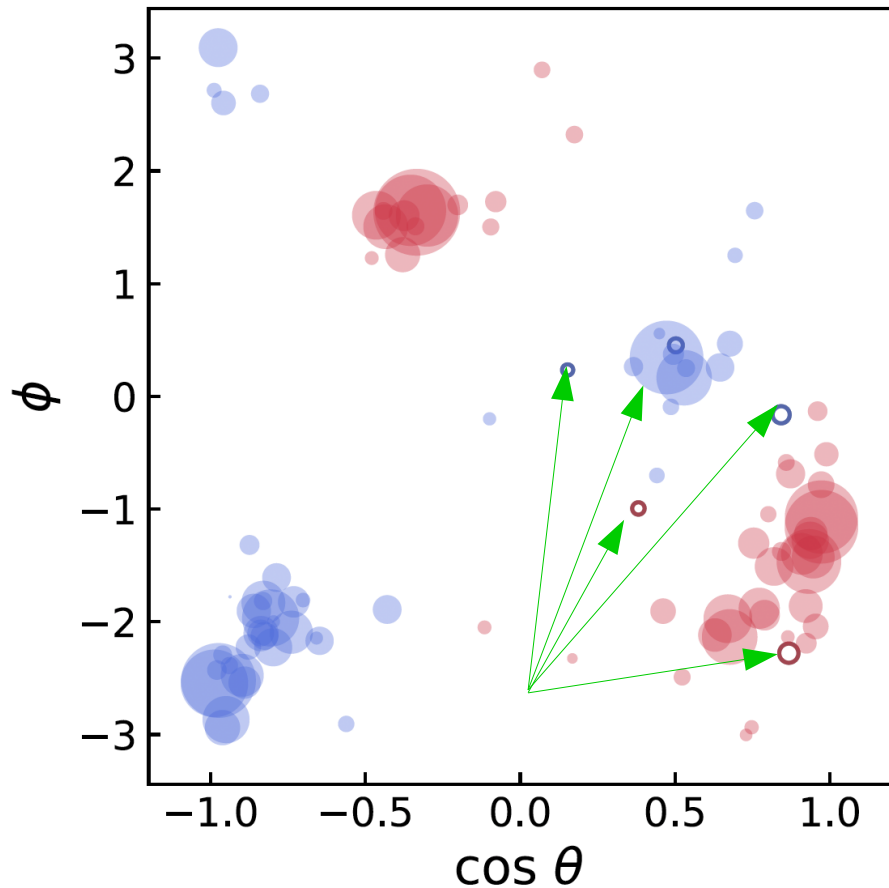


DRUID, RunNum = 0, EventNum = 7



at full hadronic ZH event

Advanced CSI: A coloring game

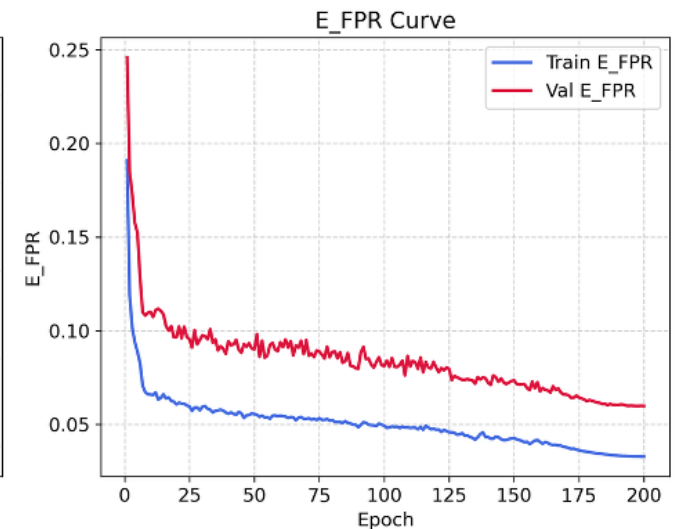
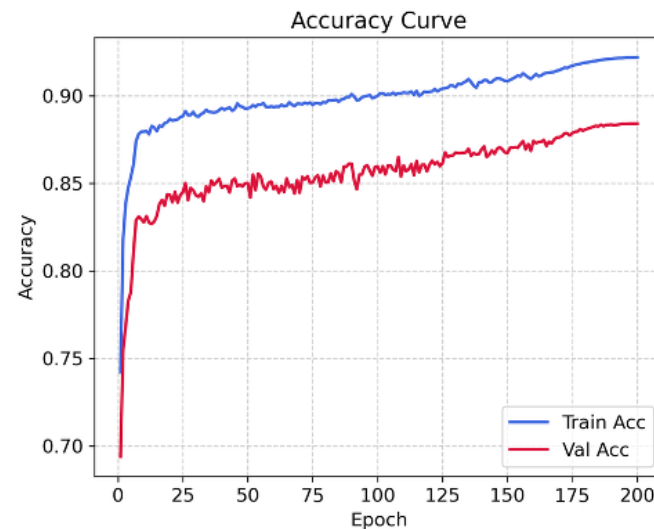
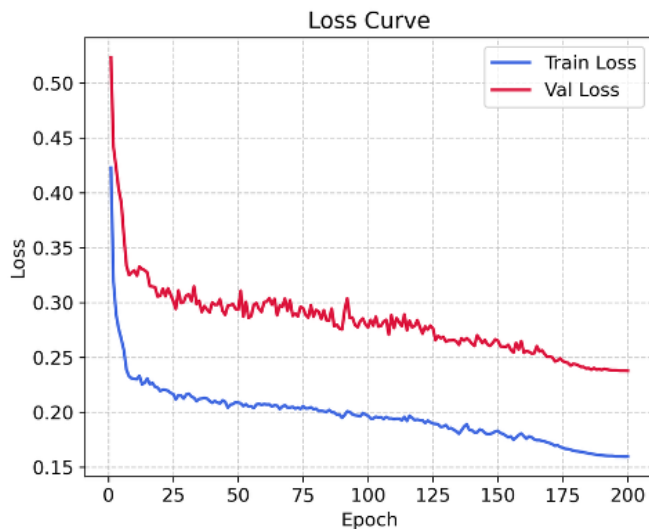
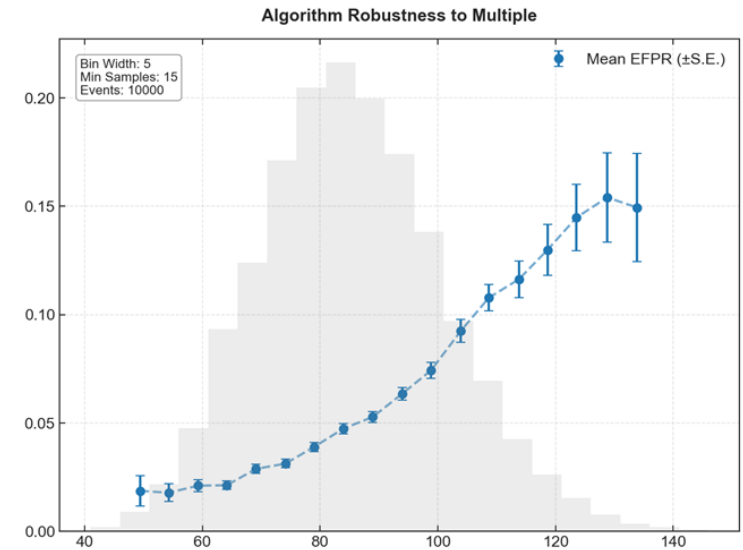


Mid classified energy ratio: reduced by 3-5 folds compared to conventional (jet clustering + matching)

Grouping: Applicable to Higgs self coupling, ... Pile up veto, and hit clustering, etc.

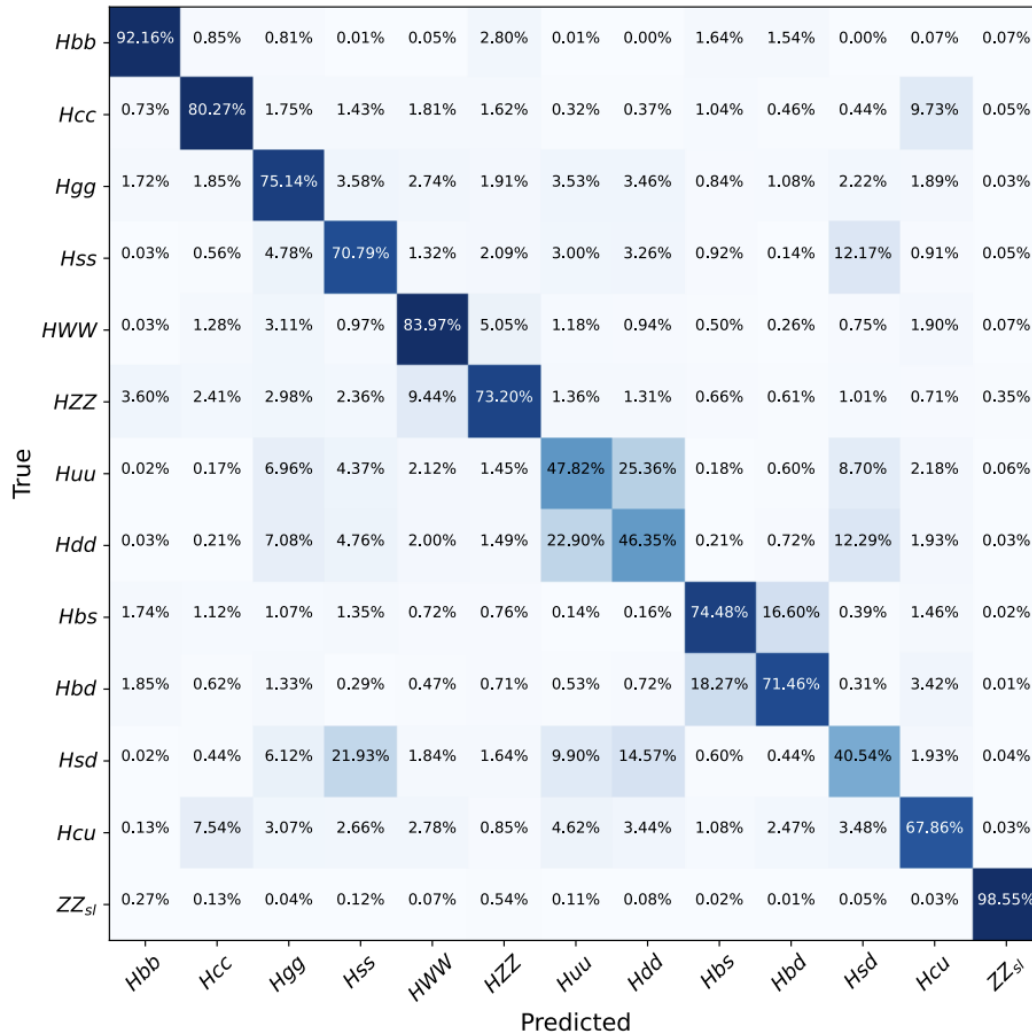
ACSI: exploration & optimization

Mis-classified energy ratio	Classical	Original Version	Mamba
WW	7.9%	1.7%	1.36%
ZZ	11%	2.8%	2.35%
ZH	19%	6.6%	5.97%

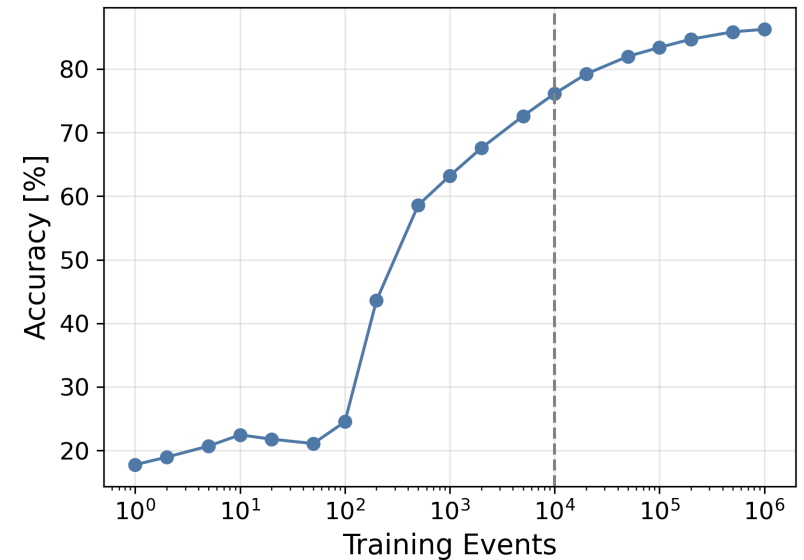


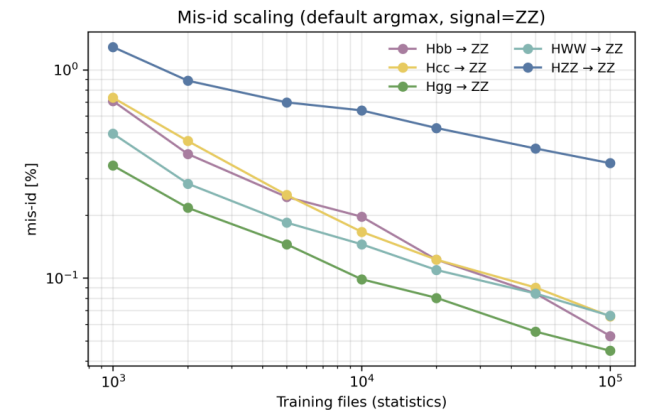
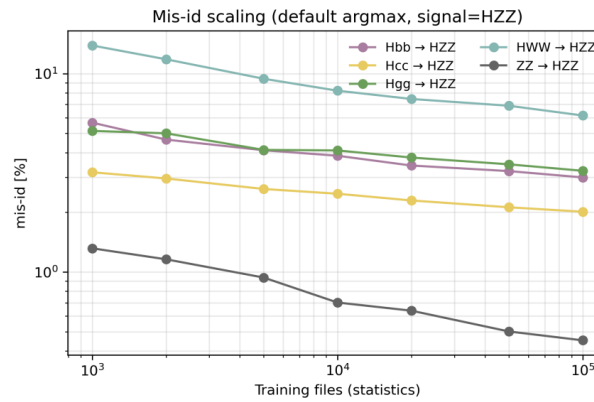
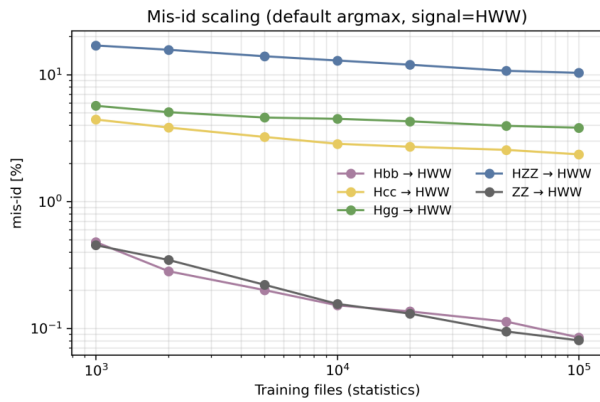
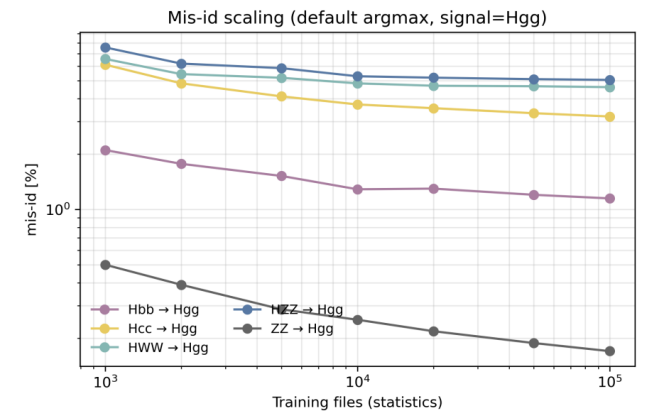
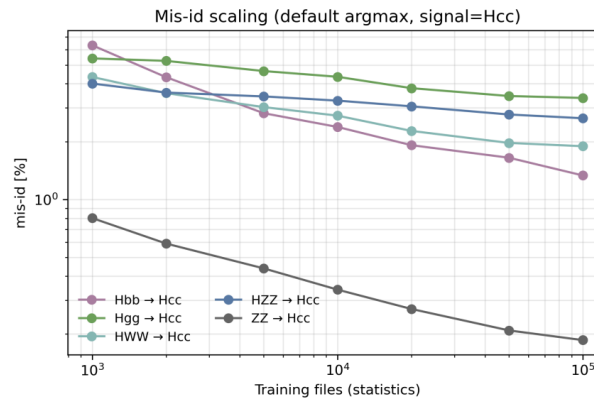
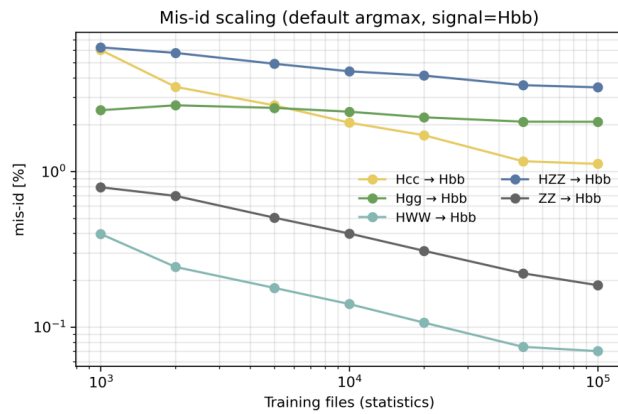
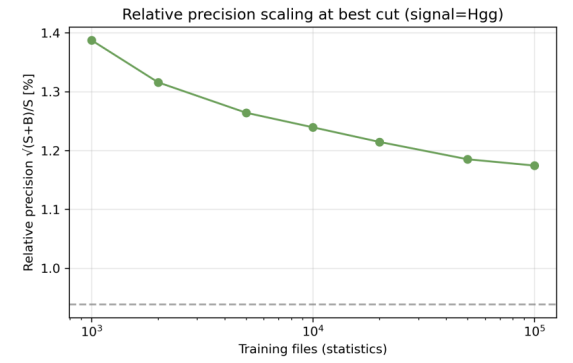
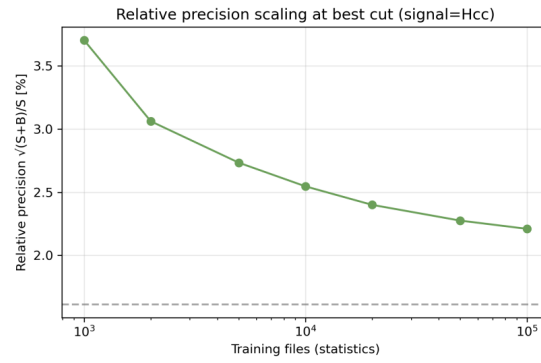
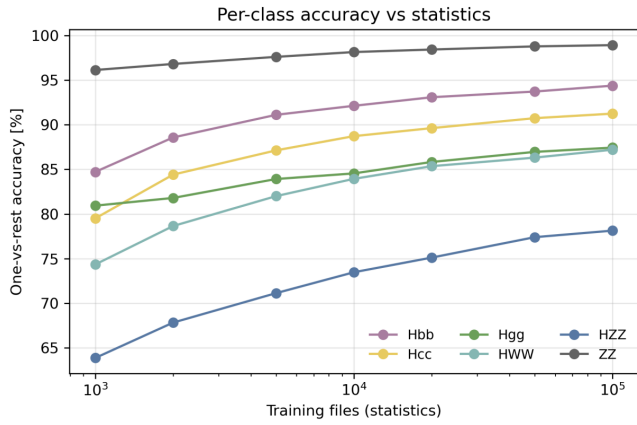
Trained on WW+ZZ+ZH & tested on ZH event

Multi-heads: Training in progress...

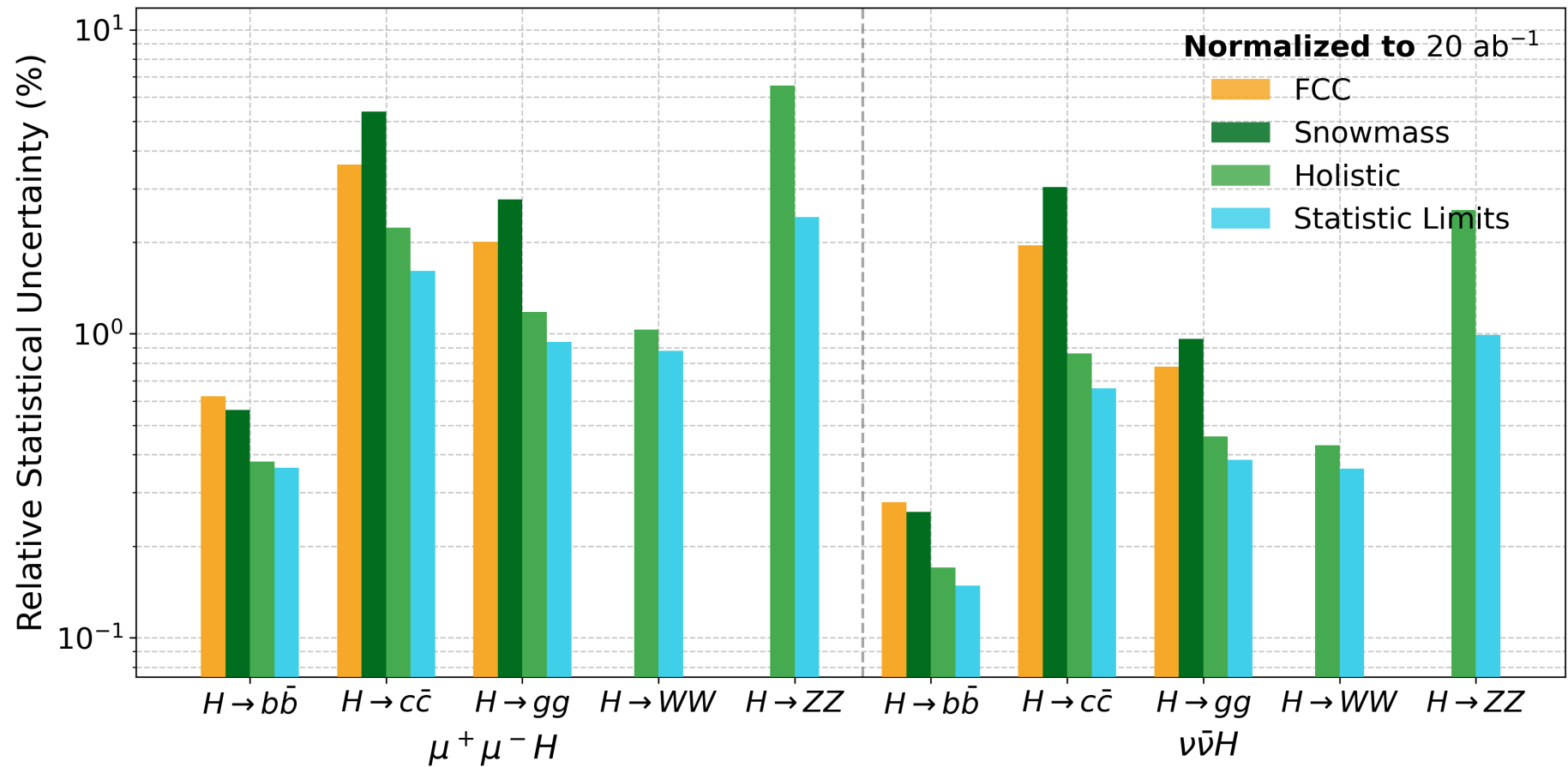


- $bb, cc, gg, WW/ZZ \rightarrow 4q$
- $H \rightarrow ss$
- $H \rightarrow uu, dd, FCNCs$





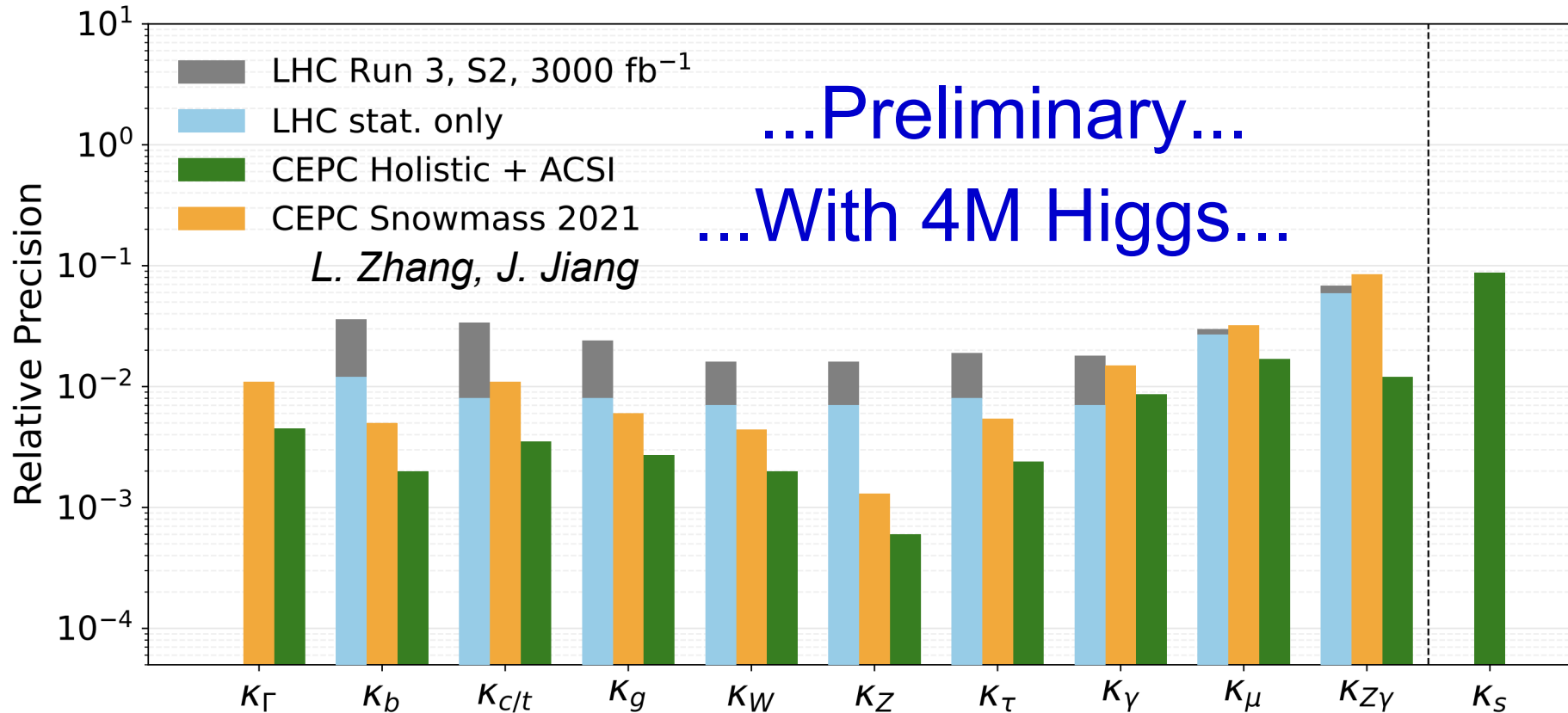
Holistic @ IIH + $\nu\nu H$



- Snowmass \rightarrow holistic : improvement of 2-4 folds

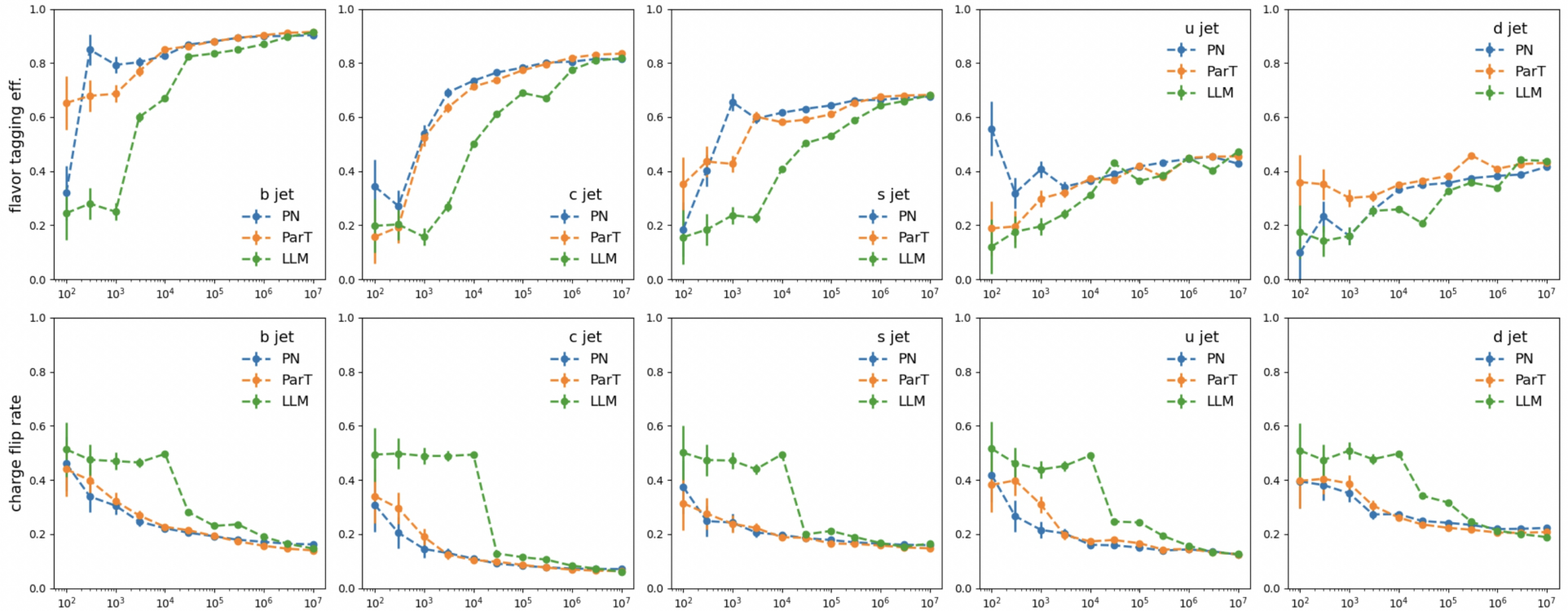
Comparison with HL-LHC

10-kappa framework precision



- LHC mainly dominated by theoretical/systematic...
- CEPC Holistic/ACSI approach the statistic limit – more luminosity is appreciated
- CEPC numbers not include theoretical uncertainties...
- CEPC numbers are based on fast simulation: **modeling an 1-1 correspondence Detector.**

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>

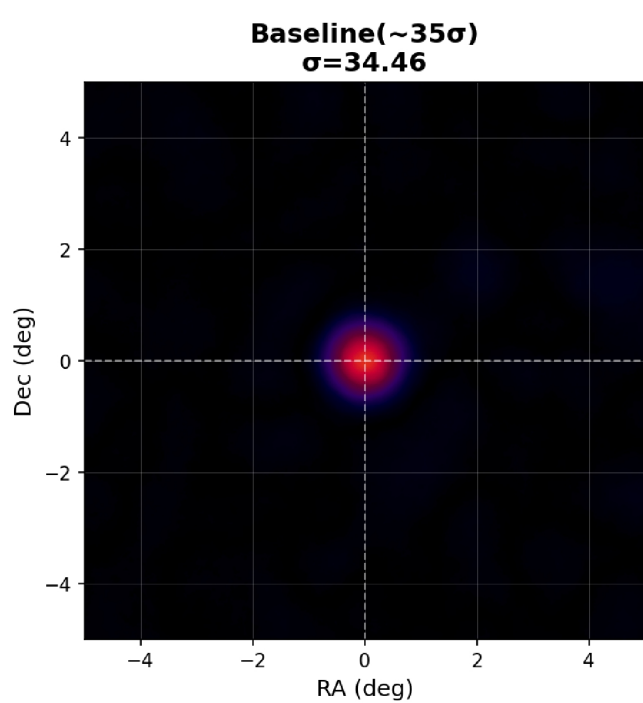


超对称
Super Symmetry
Technologies

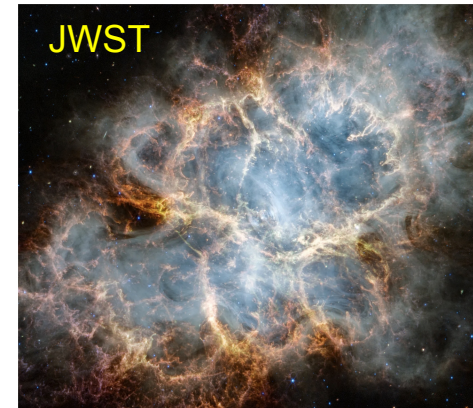
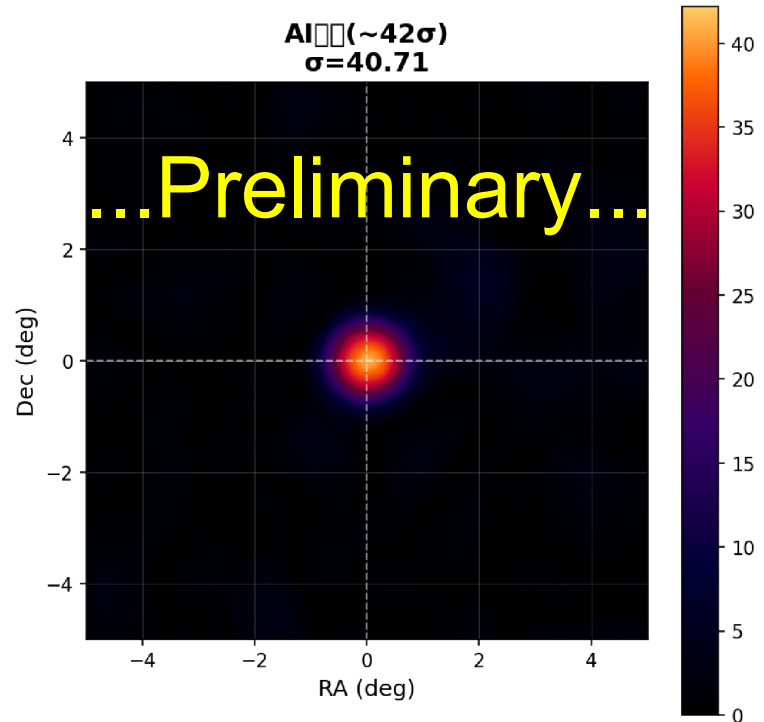
LHASSO: Learning from all hits

Colormap: cosmic | Energy: 25-100TeV

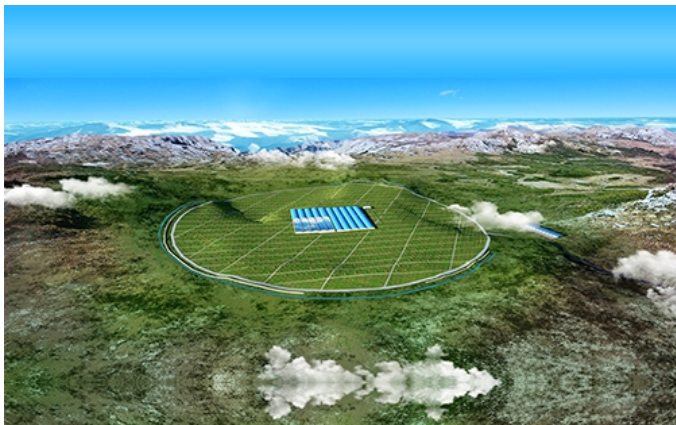
Baseline (~35 σ)
 $\sigma=34.46$



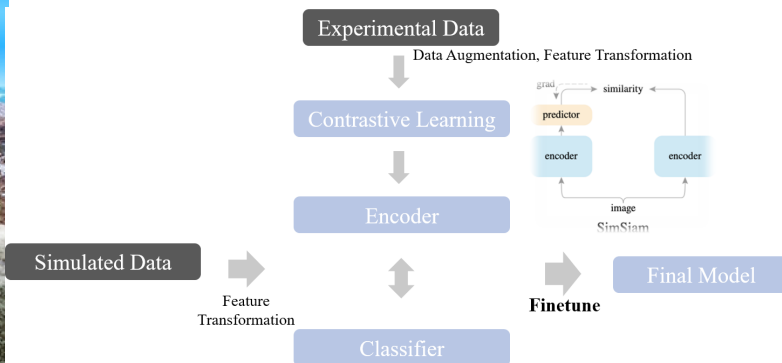
AI (~42 σ)
 $\sigma=40.71$



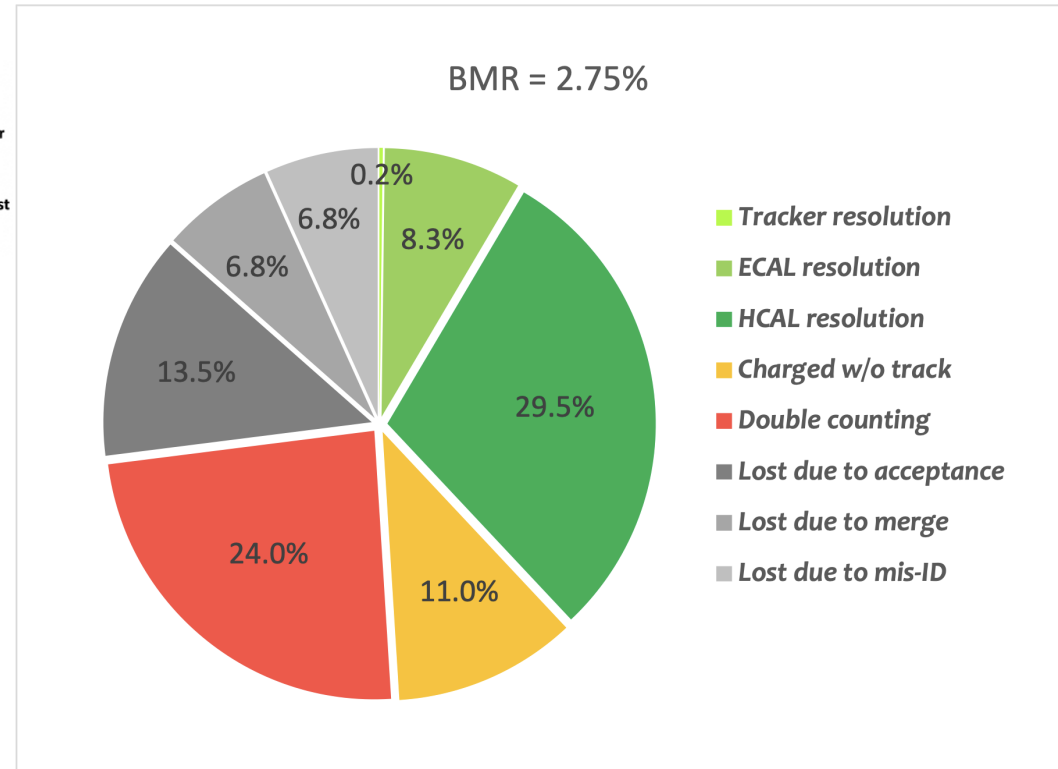
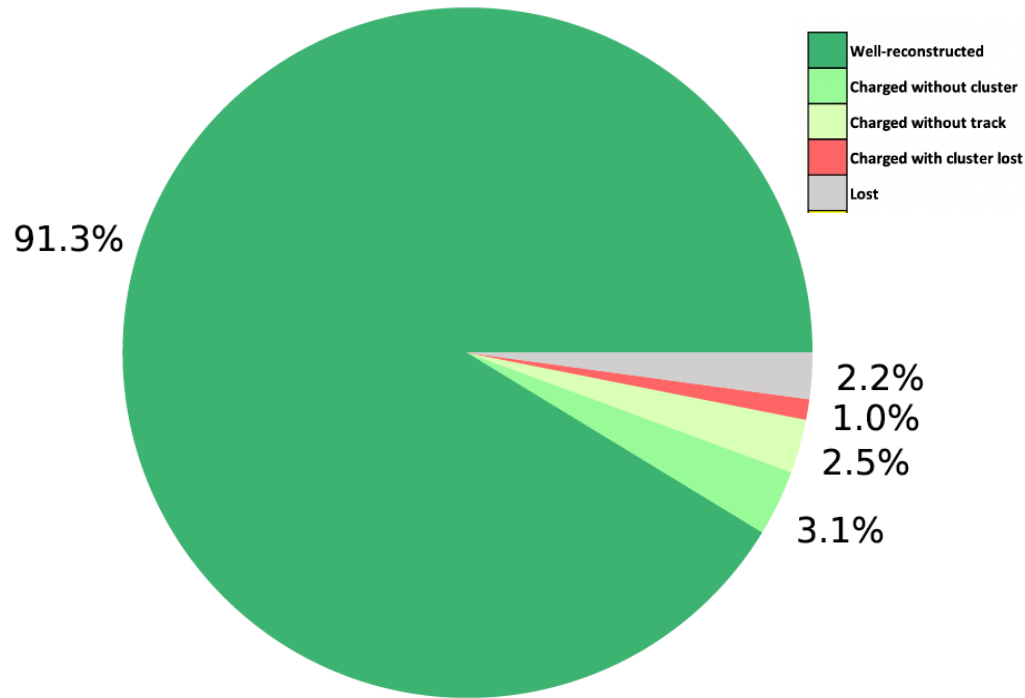
1 month of data



04/03/26



BMR decomposition @ AURORA

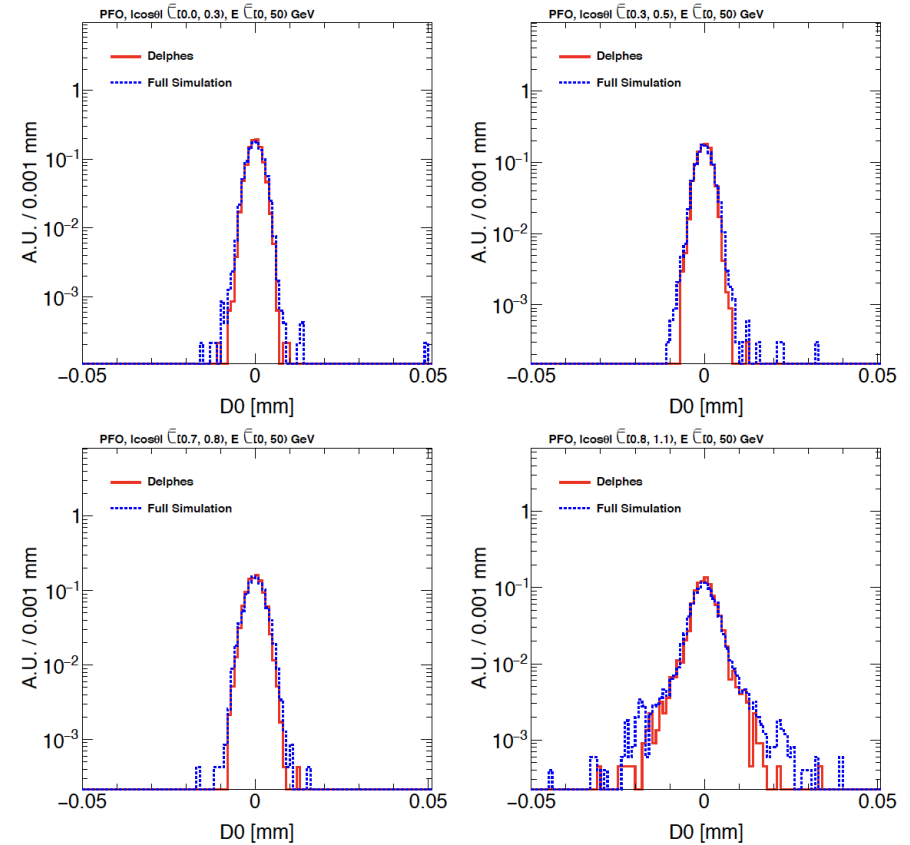
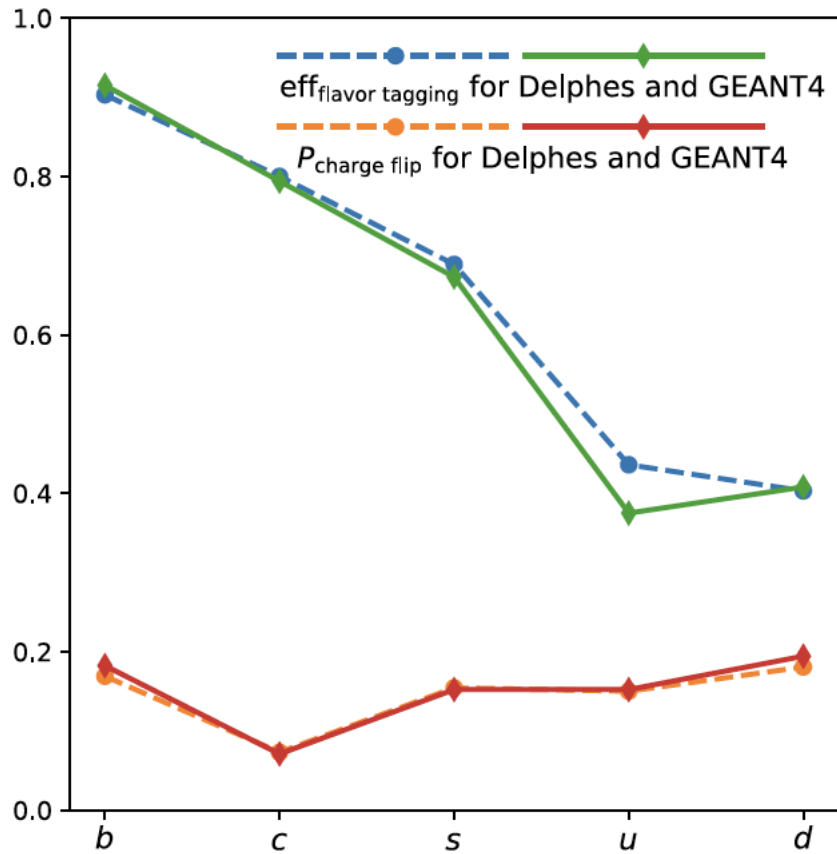


1-1 corresponding type: contributing to the BMR via resolution: $\sim o(0.1 - 0.001)$ of its mean value

Double Counting & Lost type: contributing to the BMR $\sim o(1)$ to its mean value

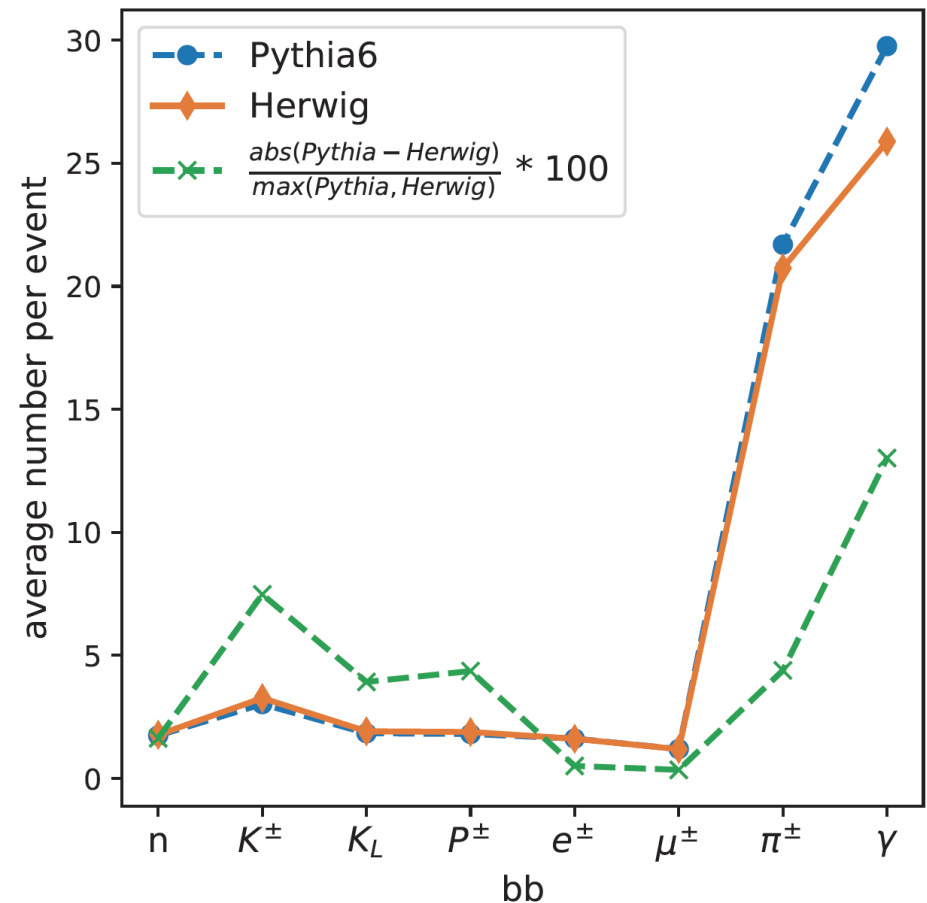
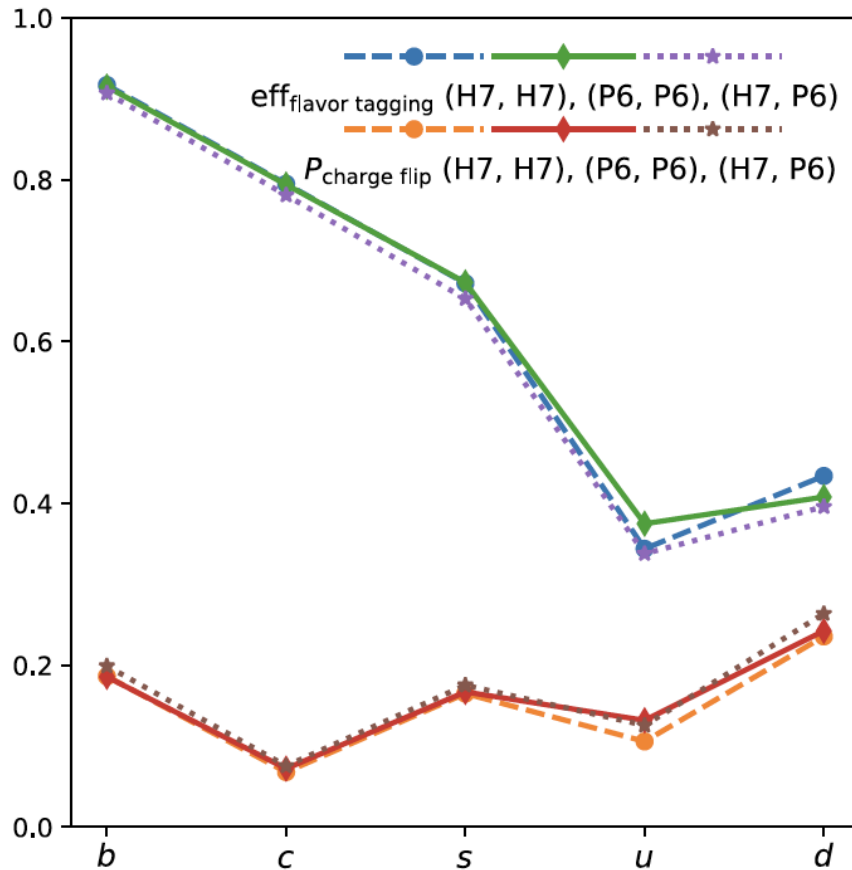
Fast/Full Simulation

Z \rightarrow $\mu\mu$ (91.2 GeV)



- Delphes \sim Perfect PFA (1 – 1 correspondence..)

V.S. Hadronization models



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

GSHCAL: simulation

Nuclear Instruments and Methods in Physics Research A 1059 (2024) 168944



Contents lists available at ScienceDirect
Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima



Full Length Article

GSHCAL at future e^+e^- Higgs factories

Peng Hu ^{a,b}, Yuexin Wang ^{a,c}, Dejing Du ^{a,b}, Zhehao Hua ^{a,b}, Sen Qian ^{a,b,*}, Chengdong Fu ^{a,b},
Yong Liu ^{a,b}, Manqi Ruan ^{a,b}, Jianchun Wang ^{a,b}, Yifang Wang ^{a,b}

^a Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing 100049, China

^b University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100049, China

^c China Center of Advanced Science and Technology, Beijing 100190, China

ARTICLE INFO

Keywords:
Higgs factory
CEPC
HCAL
Glass scintillator

ABSTRACT

The excellent jet energy resolution is crucial for the precise measurement of the Higgs properties at future e^+e^- Higgs factories, such as the Circular Electron Positron Collider (CEPC). For this purpose, a novel design of the particle flow oriented hadronic calorimeter based on glass scintillators (GSHCAL) is proposed. Compared with the designs based on gas or plastic scintillators, the GSHCAL can achieve a higher sampling fraction and more compact structure in a cost-effective way, benefiting from the high density and low cost of glass scintillators. In order to explore the physics potential of the GSHCAL, its intrinsic energy resolution and the contribution to the measurement of the hadronic system was investigated by Monte Carlo simulations. Preliminary results show that the stochastic term of hadronic energy resolution can reach around 24% and the Boson Mass Resolution (BMR) can reach around 3.38% when the GSHCAL is applied. Besides, the key technical R&D of high-performance glass scintillator tiles is also introduced.

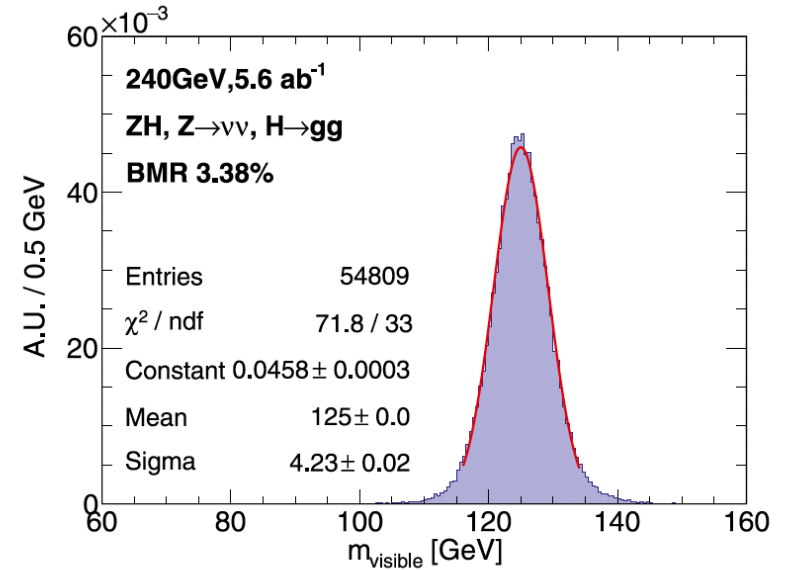


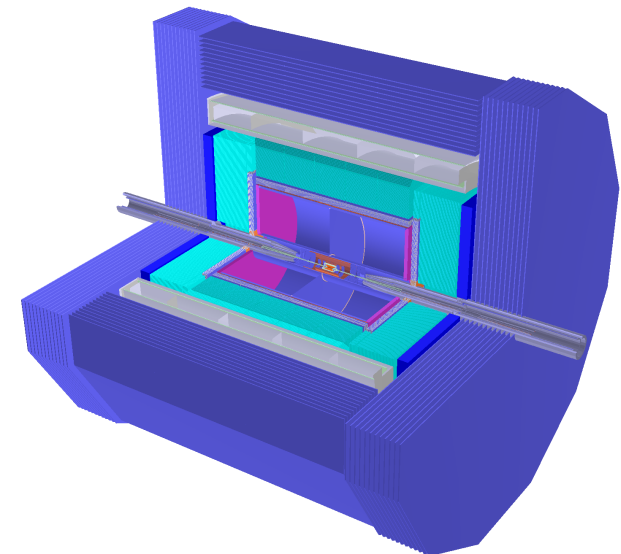
Fig. 5. Distribution of the reconstructed total visible invariant mass for $\nu\nu H \rightarrow \nu\nu gg$ channel. The distribution is fitted with a Gaussian function extended to ± 2 standard deviations.

Y. Wang, H. Liang, Y. Zhu et al.

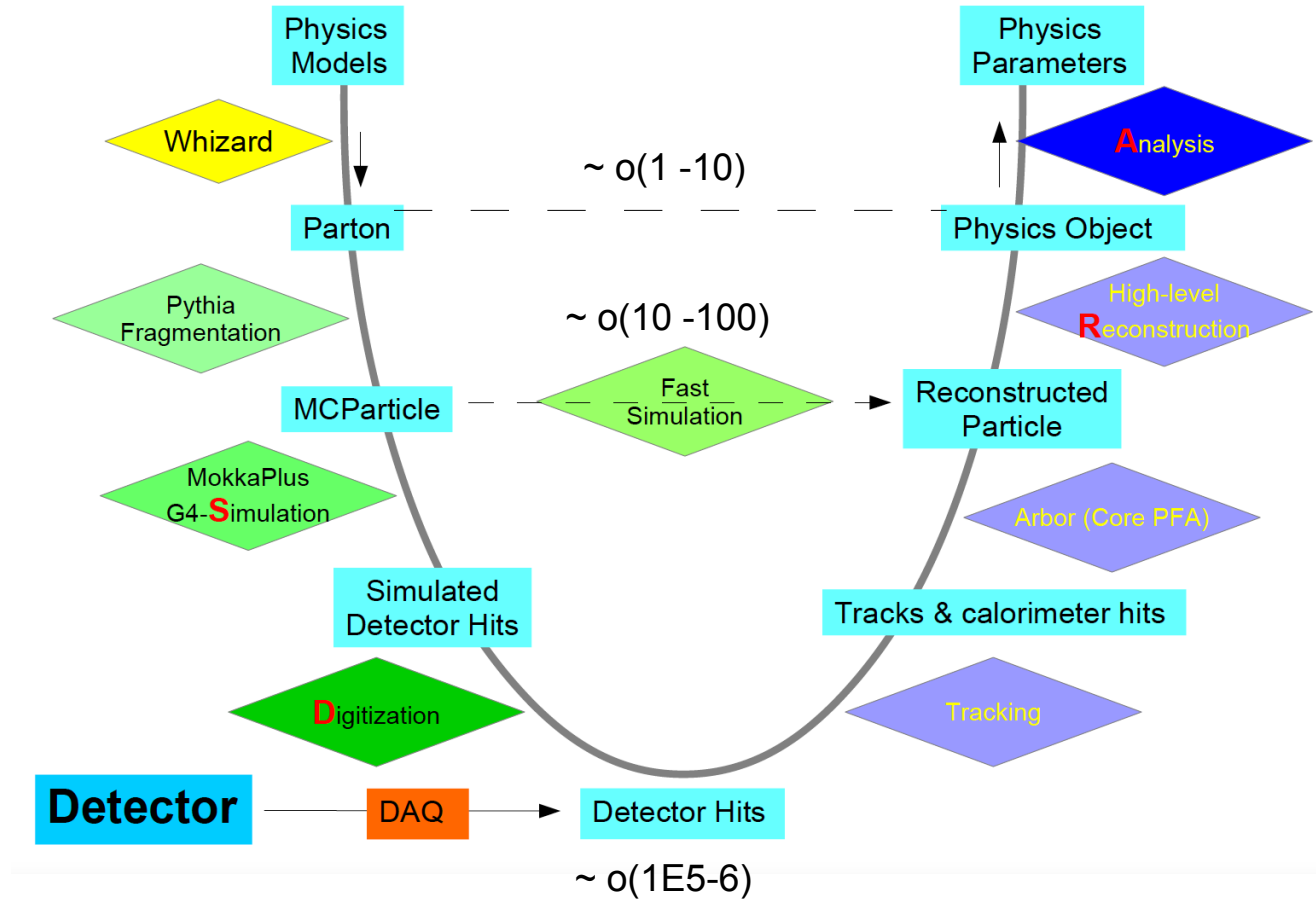
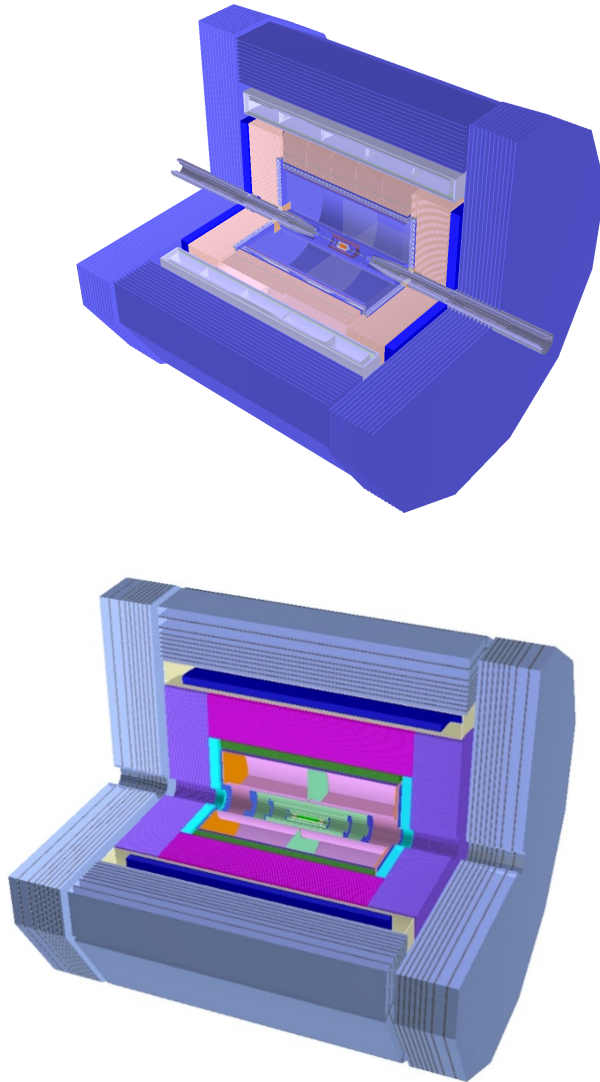
Computer Physics Communications 314 (2025) 109661

Table A.1
AURORA detector geometry parameters.

Sub-detector	Thickness (mm)	Inner radius (mm)	Outer radius (mm)	Length (mm)	Volume (m ³)	Transverse cell size	#Layers	#Channels
Vertex	-	-	16–60	125–250	-	25 × 25 μm ²	6	5.3 × 10 ⁸
Si-strip	-	-	300	1288	-	20 μm × 2 cm	3	3.0 × 10 ⁷
Tracker	-	300	1810	4600	-	-	-	-
TPC	-	300	1800	4700	47	1 × 6 mm ²	220	2.9 × 10 ⁶
ECAL	173	1845	2018	5250	15	1 × 1 cm ²	30	2.5 × 10 ⁷
HCAL	1145	2072	3250	7590	180	2 × 2 cm ²	48	1.8 × 10 ⁷
Solenoid	700	3275	3975	7750	120	-	-	-
Yoke	1200	4000	5200	10500	470	-	-	-



CEPC Detector & Reconstruction



PFA oriented Approach: **Arbor, etc**