# The holistic approach and one-one correspondence reco: with AI tools

CEPC WS@Barcelona

Manqi

## Holistic approach



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

#### 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."

17/06/25 Hao Liang, Yongfeng, etc CEPC WS@Barcelona

https://arxiv.org/abs/2310.03440 https://arxiv.org/abs/2309.13231 3

## Analysis: vvH, H $\rightarrow$ 2 jet



- vvH,  $H \rightarrow bb/cc/gg/ss$  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!

## **Color Singlet Identification**



#### at full hadronic ZH event

# CSI: bottleneck for measurement at full hadronic events



Published for SISSA by 2 Springer

Received: March 11, 2022 Revised: September 9, 2022 Accepted: November 11, 2022 Published: November 16, 2022

JHEP11(2022)100

The Higgs  $ightarrow b ar{b}, c ar{c}, gg$  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 combination 0.27% 4.03% 1.56%

1.57%

1.06%

0.35%

0.49%

 $H \to c\bar{c}$ 

14.43%

10.16%

7.74%

5.75%

 $H \rightarrow qq$ 

10.31%

5.23%

3.96%

1.82%

Z decay mode  $H \rightarrow b\bar{b}$ 

 $Z \rightarrow e^+ e^-$ 

 $Z \to \mu^+ \mu^-$ 

 $Z \to q\bar{q}$ 

 $Z \rightarrow \nu \bar{\nu}$ 

 Table 3. The signal strength accuracies for different channels.

*E-mail:* ruanmq@ihep.ac.cn

- 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



17/06/25

#### Scaling...



# 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

#### Holistic approach + ACSI



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

ACSI makes a leap even from Holistic, but still has quite some margin to improve...

 $H \rightarrow ss$  within the reach...

https://arxiv.org/pdf/2506.11783

CEPC WS@Barcelona

#### From PFA to 1-1 correspondence

1

 $\checkmark$ 

 $\checkmark$ 

Х

X

×

 $\checkmark$ 

**Final state** particles



OMPUTER PHYSIC

Check for updates



https://arxiv.org/abs/2411.06939

Computer Physics Communications 314 (2025) 109661



Computational Physics

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

Yuexin Wang<sup>a,b,0</sup>, Hao Liang<sup>a,c,d</sup>, Yongfeng Zhu<sup>e</sup>, Yuzhi Che<sup>a,f</sup>, Xin Xia<sup>a,c</sup>, Huilin Qu<sup>g</sup>, Chen Zhou<sup>e</sup>, Xuai Zhuang<sup>a,c</sup>, Manqi Ruan<sup>a,c</sup>,

#### 17/06/25



#### **Boson Mass Resolution**



Higgs factory: need BMR < 4% (critical for qqH & qqZ separation using recoil mass to di-jet) Strongly motivated to improve BMR to 3% or even lower, especially for NP & Flavor CDR baseline (left plot): BMR = 3.75% 17/06/25 CEPC WS@Barcelona 12

## BMR decomposition @ CDR



- 1<sup>st</sup> HCAL resolution dominant the uncertainties from intrinsic detector resolution: need better HCAL → R & D of GSHCAL
- 2<sup>nd</sup> Leading contribution: Confusion from shower Fragments (fake particles), need better Pattern Reco.

• CDR baseline - GRPC HCAL

#### **GSHCAL:** simulation



Peng Hu <sup>a,b</sup>, Yuexin Wang <sup>a,c</sup>, Dejing Du <sup>a,b</sup>, Zhehao Hua <sup>a,b</sup>, Sen Qian <sup>a,b,\*</sup>, Chengdong Fu <sup>a,b</sup>, Yong Liu <sup>a,b</sup>, Manqi Ruan <sup>a,b</sup>, Jianchun Wang <sup>a,b</sup>, Yifang Wang <sup>a,b</sup>

<sup>a</sup> Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing 100049, China <sup>b</sup> University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100049, China <sup>c</sup> China Center of Advanced Science and Technology, Beijing 100190, China

#### ARTICLE INFO

ABSTRACT

Keywords: Higgs factory CEPC HCAL Glass scintillator 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

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  $v\bar{v}H \rightarrow v\bar{v}gg$  channel. The distribution is fitted with a Gaussian function extended to  $\pm 2$  standard deviations.



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

Table A.1		
AURORA detector	geometry	parameters.

Sub-detector	Thickness (mm)	Inner radius (mm)	Outer radius (mm)	Length (mm)	Volume (m <sup>3</sup> )	Transverse cell size	#Layers	#Channels
Vertex	-	-	16-60	125-250	-	$25 \times 25 \ \mu m^2$	6	$5.3 \times 10^{8}$
			155	736				
Si-strip	-	-	300	1288	-	$20 \ \mu m \times 2 \ cm$	3	$3.0 \times 10^{7}$
Tracker			1810	4600				
TPC	-	300	1800	4700	47	$1 \times 6 \text{ mm}^2$	220	$2.9 \times 10^{6}$
ECAL	173	1845	2018	5250	15	1 × 1 cm <sup>2</sup>	30	$2.5 \times 10^{7}$
HCAL	1145	2072	3250	7590	180	$2 \times 2 \text{ cm}^2$	48	$1.8 \times 10^{7}$
Solenoid	700	3275	3975	7750	120	-	-	-
Yoke	1200	4000	5200	10500	470	-	-	-

Computer Physics Communications 314 (2025) 109661

# Cluster splitting: the most severe confusions



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):  $6.3/7.6 \text{ GeV} \rightarrow 0.7/1.4 \text{ GeV}$ 

CEPC WS@Barcelona



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

Al enhanced reconstruction:  $3.4\% \rightarrow 2.8\%$ .

Recent update: further optimization + Pid, etc, current value ~2.68% 17/06/25 CEPC WS@Barcelona

# 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

#### **BMR:** perspectives

- Resolutions: assume improved by 50%
  - Crystal ECAL: With efficient control of confusion
  - Detector optimization + Innovative Estimator (Energy, Time, Spatial...) with 5d calorimeter (ToF) & AI: ToF could determine very precisely the energy of low-E hadron – Giving its type identified...
- Charged w/o track: improved by 20% via Improve tracking efficiency, etc
- Double Counting: improved by 60% via Improve matching in the core PFA, i.e., Arbor
- Lost: improved by 15% (mainly at Mis vetoing & Merging, both improving by 30%)
- Need to better understand, identify & control the impact of secondary particles... (those generated in interactions between primary V.S. Upstream material, plus back-scattering)



#### BMR: from CDR to possible future...



#### Pid: differential performance





17/06/25 Neutral Hadron ID: 5d Calorimetwys with client capability (δt~100 ps/hit)

True

21

#### Perspectives with 1-1 correspondence



- 5d calo is critical: ToF for all visible particle, thus Pid...
- ToF enhanced energy measurement: expecting BMR:  $2.8 \rightarrow 2.2-2.4$ , Strongly Boost the light quark ID.
  - Need excellent CALO + ToF ~ o(10 ps)/Cluster
  - Assume Low energy neutrons & secondary particles can be tamed... still challenge...

#### 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...

## Summary

- ... Higgs factory should and could have excellent performance...
- AI as the trend...
  - 1-1 correspondence reconstruction: excellent PID + BMR of 2.7%
    - AURORA (CDR detector + GSHCAL), started to evaluate other concepts.
    - Roadmap to 2% BMR demonstrated,
    - 5-d calo is the key
  - Holistic approach
    - Reco: Jet origin id, highly relevant to 1-1
    - Analysis: Holistic + ACSI: enhance the discovery power by ~3 folds
- *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
  - Lots more to explore, with unsupervised, LLM, ... rich interplay & synergies.

- ...

#### Future: From leaves to the trees

- The hadronization process is ~ tree like
  - PFA & 1-1 corresponding committed to reconstruct well the leaves the final state particles that actually interacts with detector/calorimeter
  - Possible to identify the entire tree: reco parenting info of final state particles
    - Pi-0,
    - K\_short, Lambda, EPJP (2020) 135:274
    - Phi, PRD 105, 114036 (2022)
    - ...
    - Tau, D, B...
- Impact:
  - Essential for Flavor & New Physics
  - Enhance Jet Origin Identification
- Methodology: Comparative analysis
  - Conventional + Al
- Synergies with Event building Trigger + On line + Off line...



## Al era: Holistic approach

- Feed all reconstructable info. to the classifier in principle free of human intervene (no need to find Cut variables, etc..). Require excellent detector & reconstruction, where 1-1 serves as a benchmark & standard
- Supervised Learning Systematic uncertainty control is the challenge, esp. for precision measurements. Relies strongly on accurate simulation
  - Theoretical: need dedicated efforts on theoretical framework, For the Higgs factory, the challenges include high precision perturbative calculation, the hadronization models, and potentially QCD effect like color-reconnection effects
  - Experimental: need profound understanding of the detector response requires innovative Calibration & Monitoring, plus Digitization & Validation. For which, the 1-1 provides much more observable and ways...
  - Need comparative analysis over the relevant phase space, to control & to understand the scaling behavior, which will also shed light on AI development.
  - Exploration just started
- Longer term... non-supervised learning, or even migrate to LLM/General models...
- Even longer term: Data stream + information compressing using reco + analysis + interpretation... Al is essential, plus we need to set check points & mile stones to quantify and understand its behavior

#### Back up

#### Performance requirements

- To reconstruct all Physics Object, especially Jets
  - Z & W: ~ 70% goes to a pair of jets
  - Higgs: ~97% final state with jets (ZH events)
  - Top:  $t \rightarrow W + b$





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

CEPC WS@Barcelona

#### BMR dependence to its components



17/06/25

CEPC WS@Barcelona

#### BMR dependence on Cut...



... If the High Values tails could be tamed...

#### BMR: receipt & comparison to JER

Chinese Physics C Vol. 43, No. 2 (2019) 023001

#### The Higgs signatures at the CEPC CDR baseline\*

Hang Zhao(赵航)<sup>1,2,3</sup> Yong-Feng Zhu(朱永峰)<sup>1,4</sup> Cheng-Dong Fu(傅成栋)<sup>1</sup> Dan Yu(于丹)<sup>1</sup> Man-Qi Ruan(阮曼奇)<sup>1,2,1)</sup>

<sup>1</sup>Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China <sup>2</sup>CAS Center for Excellence in Particle Physics, Beijing 100049, China <sup>3</sup>Collaborative Innovation Center for Particles and Interactions, Hefei 230026, China <sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** As a Higgs factory, the CEPC (Circular Electron-Positron Collider) project aims at precision measurements of the Higgs boson properties. A baseline detector concept, APODIS (A PFA Oriented Detector for the Higgs factory), has been proposed for the CEPC CDR (Conceptual Design Report) study. We explore the Higgs signatures for this baseline design with  $\nu\bar{\nu}$  Higgs events. The detector performance for reconstructing charged particles, photons and jets is quantified with  $H \rightarrow \mu\mu, \gamma\gamma$  and jet final states, respectively. The resolutions of reconstructed Higgs boson mass are comparable for the different decay modes with jets in the final states. We also analyze the  $H \rightarrow WW^*$  and  $ZZ^*$  decommends where a decomparation between different decay and an explore the decomparation between different decay modes.

Table 3. Higgs boson mass resolution (sigma/Mean) for different decav modes with jets as final state particles, after event cleaning.

$H \rightarrow bb$	$H \rightarrow cc$	$H \rightarrow gg$	$H \to WW^*$	$H \rightarrow ZZ^*$
3.63%	3.82%	3.75%	3.81%	3.74%





inst

PUBLISHED BY IOP PUBLISHING FOR SISSA MEDIALAB

RECEIVED: January 15, 2021 ACCEPTED: April 13, 2021 PUBLISHED: July 21, 2021

#### Jet performance at the circular electron-positron collider

#### P.-Z. Lai,<sup>a</sup> M. Ruan<sup>b</sup>,\* and C.-M. Kuo<sup>a</sup>

<sup>a</sup>Department of Physics and Center for High Energy and High Field Physics,

National Central University, No. 300, Zhongda Rd., Taoyuan City 32001, Taiwan

<sup>b</sup>Experimental Physics Division, Institute of High Energy Physics,

19B Yuquan Road, Beijing, China

E-mail: Manqi.ruan@ihep.ac.cn



# Idea 3: global measurements with multi-class classification

- Extend the holistic approach to sufficiently large group of event classification ~ o(100) classes of Higgs events + critical SM backgrounds.
  - Higgs signal: {IIH, vvH, qqH}\*Br(H->X)
  - SM backgrounds
    - {WW, ZZ}\*{non, semi, full hadronic} + Single Z/W\*{non, semi hadronic} + ...
    - 2 fermion
    - Di-photon (mostly pile up), more relevant to Linear Collider
    - Beam induced background + Noise...
    - ...
- Impact
  - Simultaneously measure large number of channels
  - Benchmarks for optimization, in a end-end way: express the entire detector + recon capability in terms of, i.e., global migration matrix

#### Idea 3: high value measurements

- H->ss:
  - Good enough det + ~o(20) iab will probably leads to 5-sigma confirmation
- Higgs width: separate W fusion from ZH (i.e., with vvqq final state)
- qqH, H->inc, for the g(HZZ) measurements
- Higgs self coupling
- •
- Could be well extended, especially with EW sectors...

#### Impact on physics benchmarks...



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

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

### At ILD: Preliminary

- BMR (wo PU) & Pid
- PU study

Need to further confirm the det. Para + PU condition

#### Fake particle identification and BMR



m<sub>visible</sub> [GeV]

140

160

120

120

m<sub>visible</sub> [GeV]

140

160

#### **Energy fraction**

#### Increased fractions in ILD

- ≻ Charged w/o cluster
- > Lost (need further decomposition... using 3-stage particle mapping)



Partly due to Pandora create more mult-track PFO

Preliminary! One-one framework needs further polish to be more precise & descriptive

CEPC WS@Barcelona



CEPC

ILD (preliminary)



 $e^{\pm}$  0.999 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

True

17/06/25

#### pile-up subtraction

• BMR

0.08

0.06

A.U. / 0.50 GeV

0.02

• initial ~5.14%

initial

80

100

- rm pred pile-up 3.85%
- rm truth pile-up 3.29%
  - using MCP-clu/trk link
- rm truth pile-up 3.15% using MCP-PFO link



60

CEPC WS@Barcelona

#### PU pfo identification...

eff = 57.97%, pur = 62.23%



#### Challenges & needed actions...

- Relevant uncertainties, recommend quantification & amelioration A.S.A.P
  - Beam induced background need PFA in the space time (POST)
  - Event building with Trigger system
  - Geometry acceptance: MDI & FWD design etc
  - Detector stability mechanic design simulation & aging study
  - Tracker Noise, B-Field mapping
  - Calorimeter Noise, dead zone, inhomogeneity (i.e., attenuation)...
  - Calibration & Monitoring: could be partly addressed by 1-1

#### Preliminary diagnosis of TRD performance

- In the standard of 1-1 correspondence
  - Visible energy decomposition
  - BMR decomposition
  - 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.

#### 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

CEPC WS@Barcelona

## 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)

CEPC WS@Barcelona

#### Photon reco. efficiency





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



#### **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).

## 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. Muon det. Info not available in ECORTENT (PBFA: (both Cyber & Arbor), to be improved.

17/06/25

47

# Jol at TRD, CDR & AURORA (ideal)



Using truth Pid, TRD has better JoI than CDR detector, as it uses longer Barrel + stitching VTX Pid at TRD is limited, will degrade the H $\rightarrow$ ss measurement... (software version 0401, not 1-1) Neutral Hadron ID has strong impact on Light Quark ID: highly appreciated in H $\rightarrow$ ss

## 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.





### Thoughts on the Det. optimization

- To minimize the shower volume of incident particles
- To share the task, if necessary, between different det. Technologies
- Propose several concepts, para. to be optimized.

ECAL cost breakdown	Table 7.17	
System	Cost (kCHF)	
Electromagnetic Calorimeter	114,968	
Scintillating Crystal	105,915	
SiPM	714	
Electronics (FEE)	1,099	
Mechanics	3,796	
Cooling	96	
Installation (3%)	3,349	
Extra cost for back-end electronics	2,780	$\supset$

#### **Crystal ECAL: cost estimation**

Readout occupies 4% of the ECAL construction cost

Shall we considering ECAL with more readout Channels, and re-optimize its Cell/Bar configuration?

#### Crystal ECAL: major cost drivers

- Crystals: discussions with SIC-CAS; reaching out more vendors

- SiPMs: experiences of the JUNO-TAO detector (~10m<sup>2</sup> SiPMs)

- Electronics (FEE/BEE): inputs from the CEPC electronics team
- Carbon-fiber mechanical structures: inputs from CF manufacturer(s)

22

# Design-2: Crystal bar + Mesh

- Geometry
  - Total Crystal Volume: 24 m<sup>3</sup>
  - Single Crystal Bar Dimension:
     2.67cm \* 2.67cm \* 40cm =
     291 cc, In total 80k bars
  - Inner Area: 80 m<sup>2</sup>
  - Total Readout Channel:
    - 80000\*2 = 160k (Crystal)
    - 800000\*4 = 3.2 M (Si)
- Comments

17/06/25

 Extra material budget ~ o(1%) of the total radiation length is tolerable for the EM resolution ~ 2-3 mm of Cu. per layer



### **EM** resolution



- Positioning layer: material budget of ~ 0.2 X0 (3 mm Cu) each, total fraction < 3%
- Compatible with CMS HGC Silicon layer wi cooling; which has much higher data rate & requirement on energy reco. -> further optimization is possible

CEPC WS@Barcelona