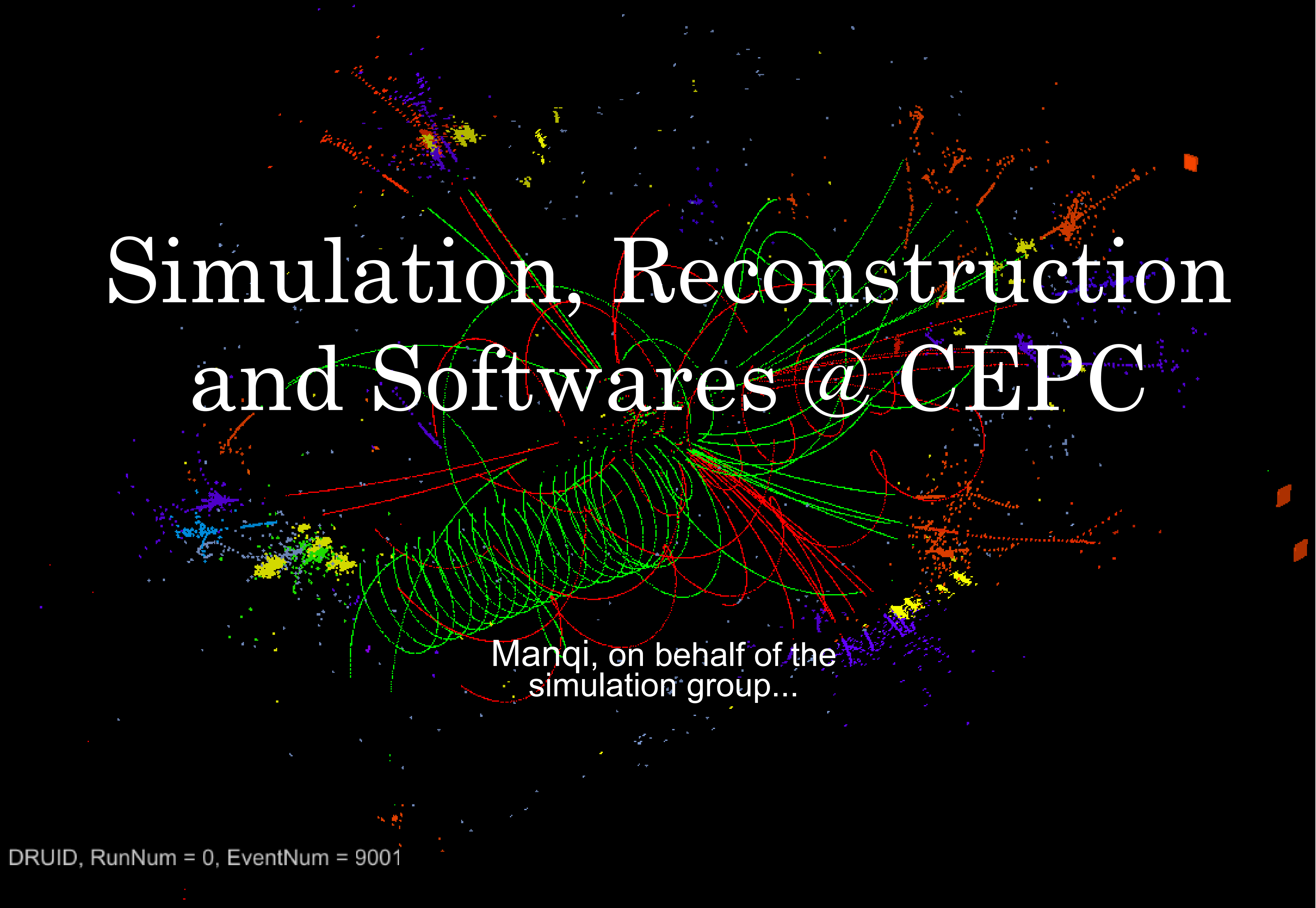
Simulation, Reconstruction and Softwares @ CEPC

Manqi, on behalf of the simulation group...



Outline

- SCRAC Software Status
- Simulation
- Reconstruction
    - Tracking
    - **PFA**
    - Flavor Tagging
- Detector optimization
- Perspective & Roadmap

SCRAC

Generator
Whizard, PYTHIA
(ref samples
Available)

Full Simulation
Mokka, DD4HEP
(req: Geant4,
Database...)

Digitization
MarlinReco, etc.
eg: **G2CD** for
Calorimeter

Tracking
MarlinReco,etc.

Stdhep | LCIO

GDML/Root

Event Display: **Druid**

PFA
Pandora/**Arbor**

Flavor Tagging
LCFIPlus

Analysis
Marlin, Root,
FSClasser

Physics result

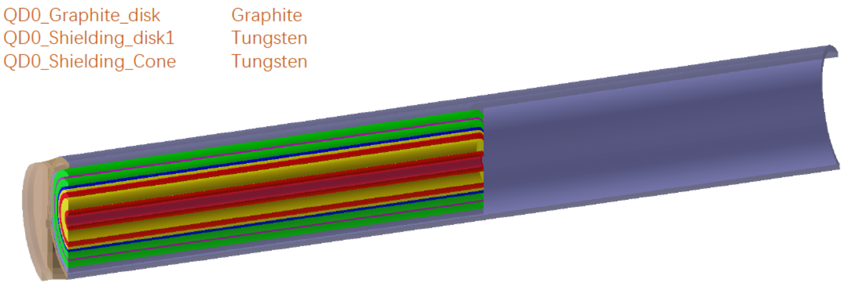
LCIO | root

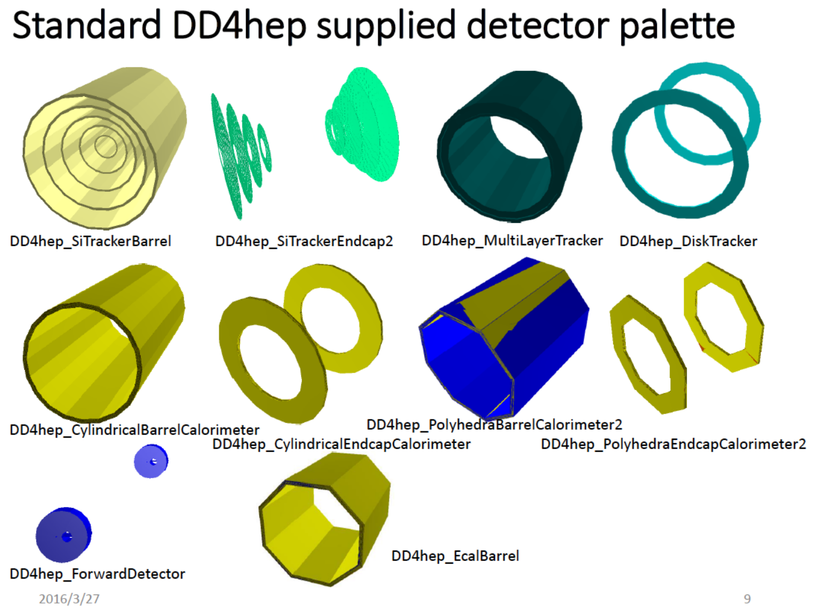
Simulation Calibration Reconstruction Analysis Chain

Simulation - geometries

- Nankai University
  - **Xu yin**, Qiuyang, Yanli, etc
  - Release of official CEPC detector models (CEPC_v1, CEPC_o_v2, ...) & Hit map – tracking level validation, etc
  - Irradiation study & MDI design implementation
- IHEP
  - **Fu Chengdong**, Shuzheng, etc
  - Extension of Mokka
    - General HGC toolkit
    - SLCIO input plugin for generators
  - DD4HEP

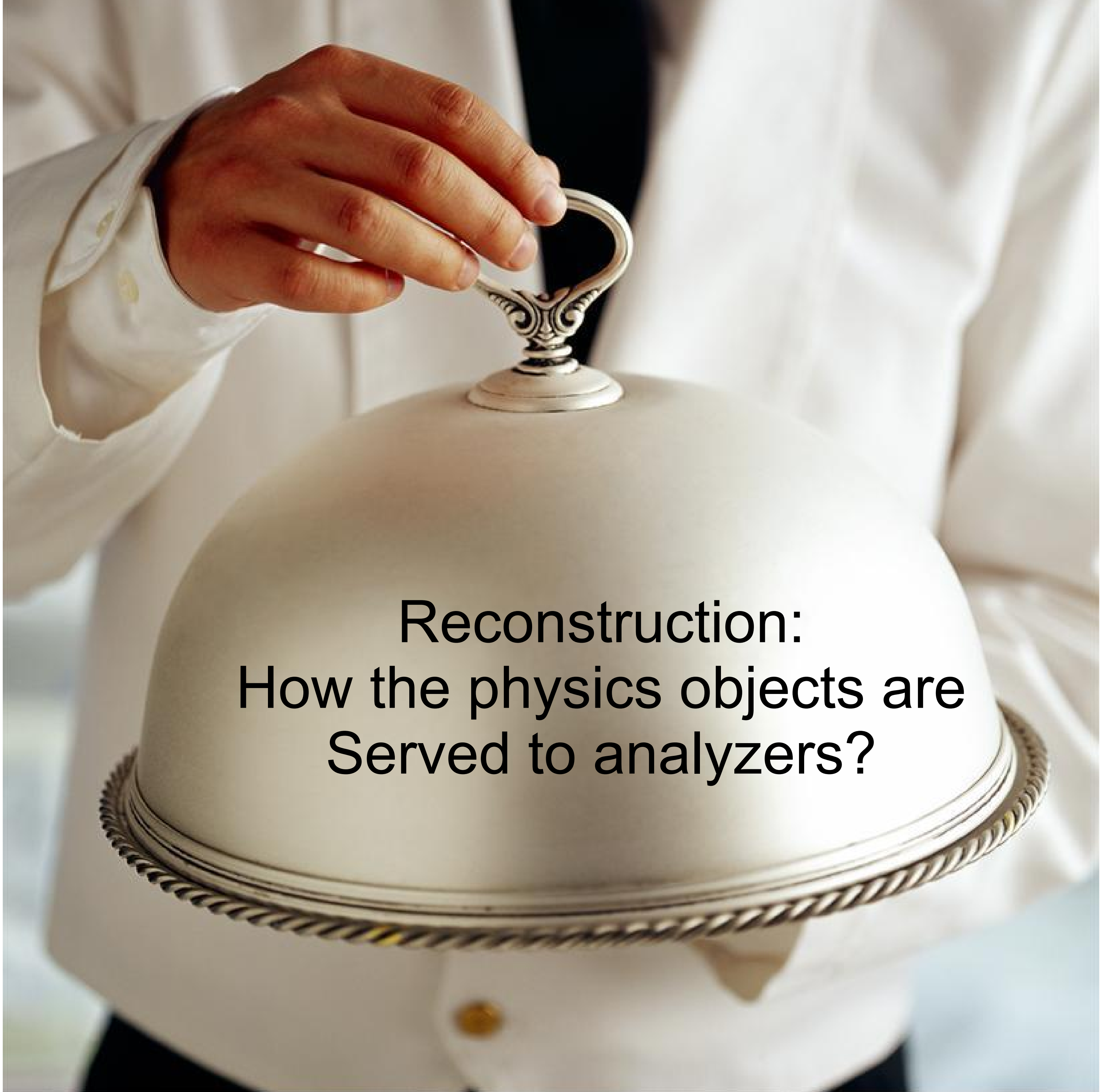
Detail of QD0, QF1



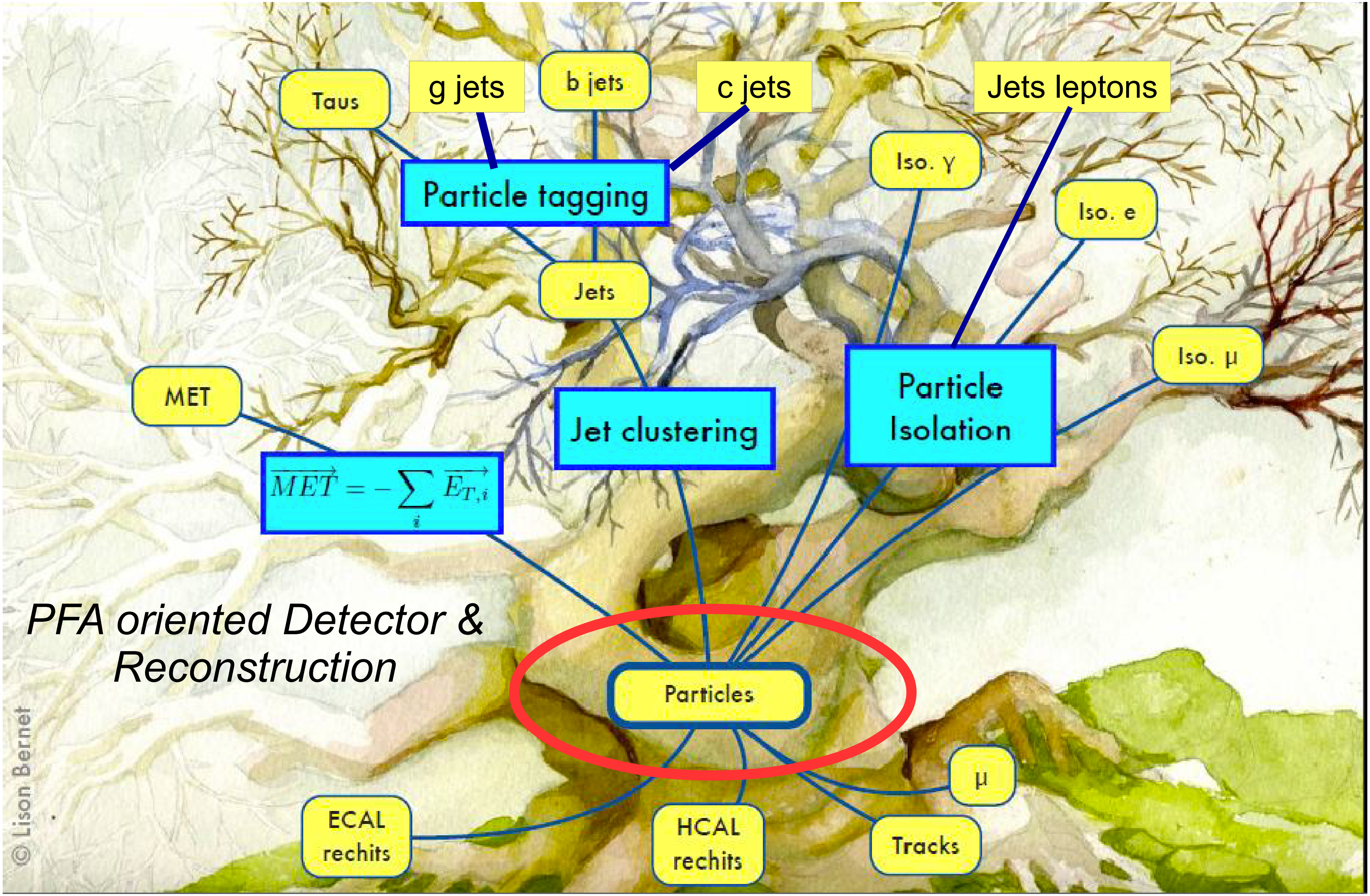


Standard DD4hep supplied detector palette

2016/3/27 9



Reconstruction:
How the physics objects are
Served to analyzers?



PFA oriented Detector & Reconstruction

Taus

g jets

b jets

c jets

Jets leptons

Particle tagging

Iso. γ

Iso. e

Jets

Iso. μ

MET

Particle Isolation

Jet clustering

$$\overrightarrow{MET} = -\sum_i \overrightarrow{E_{T,i}}$$

Particles

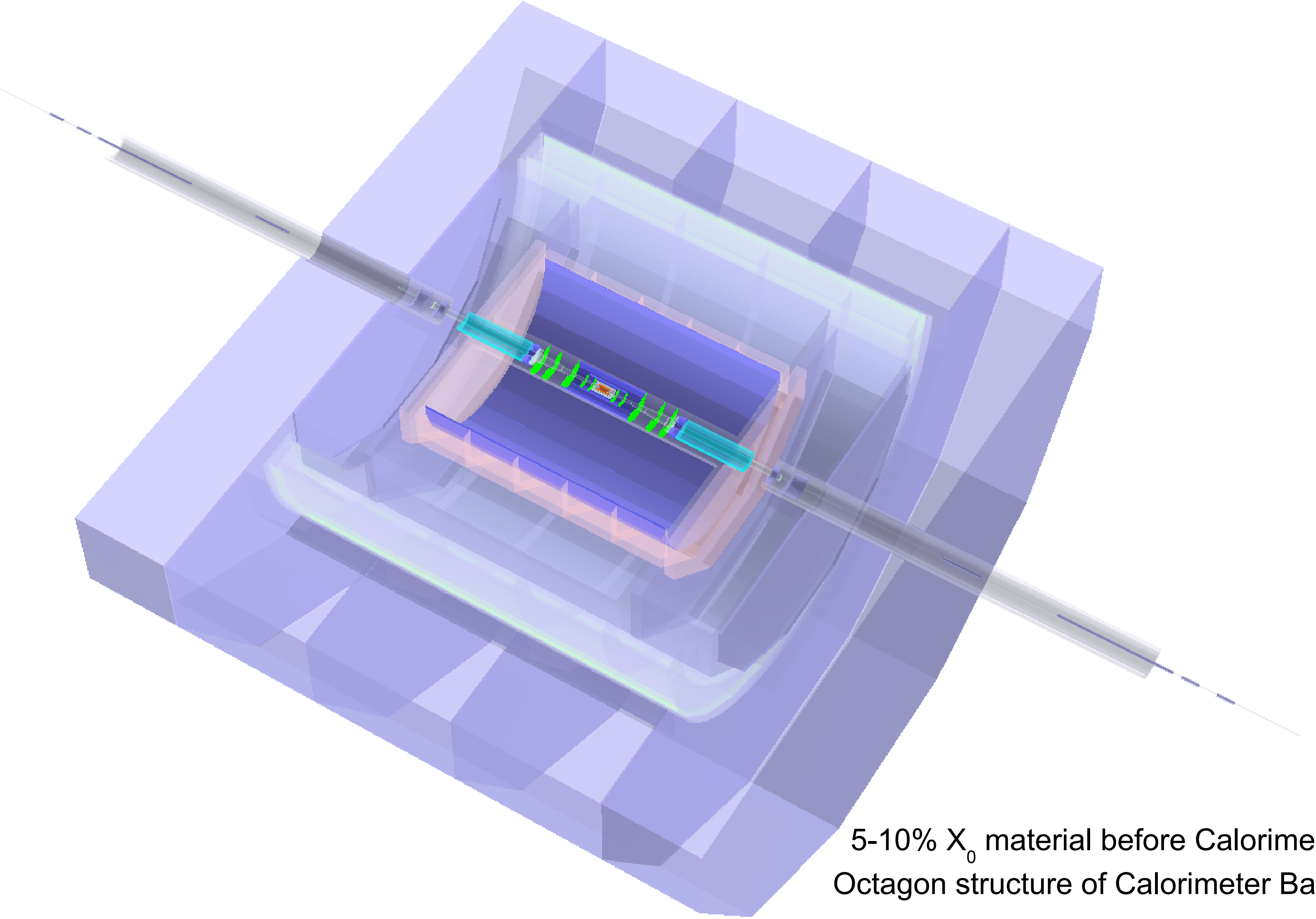
μ

ECAL rechits

HCAL rechits

Tracks

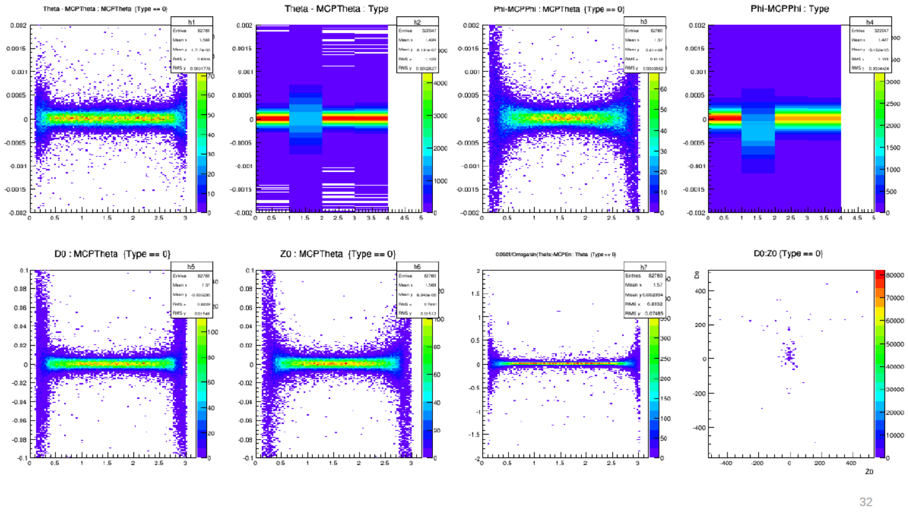
© Lison Bernet



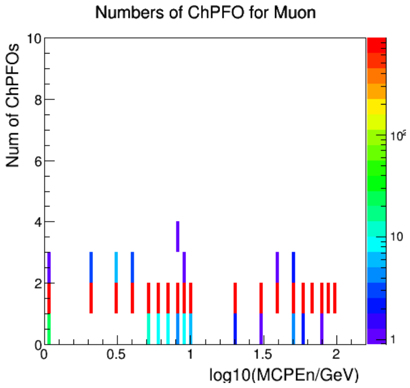
5-10% X_0 material before Calorimeter
Octagon structure of Calorimeter Barrel

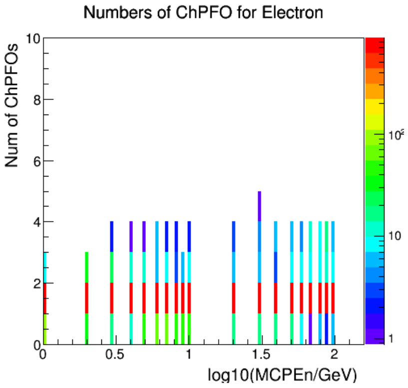
Tracking

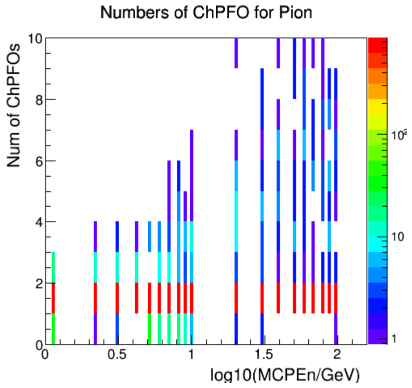
- Objective: for any geometry, produce & understand the track response:
  - Track efficiency & multiplicity
  - Differential resolution of D0, Z0, Phi, TanLambda(Theta), Omega(P_T)
- Team: Libo, Wu Linghui & Zhang Yao
  - Libo, Tracking Expert at ILC, voluntarily supporting CEPC study
- Status:
  - Kalman-Filter based technology
  - Good understand of TPC – Silicon based geometry;
  - Pure Silicon geometry – ongoing work
  - Digitization need profound understanding...

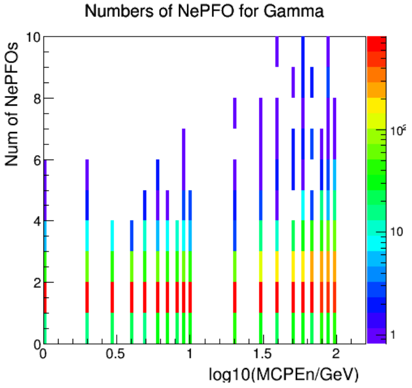


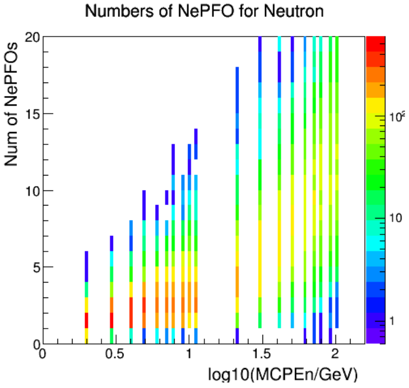
Arbor @ single particle

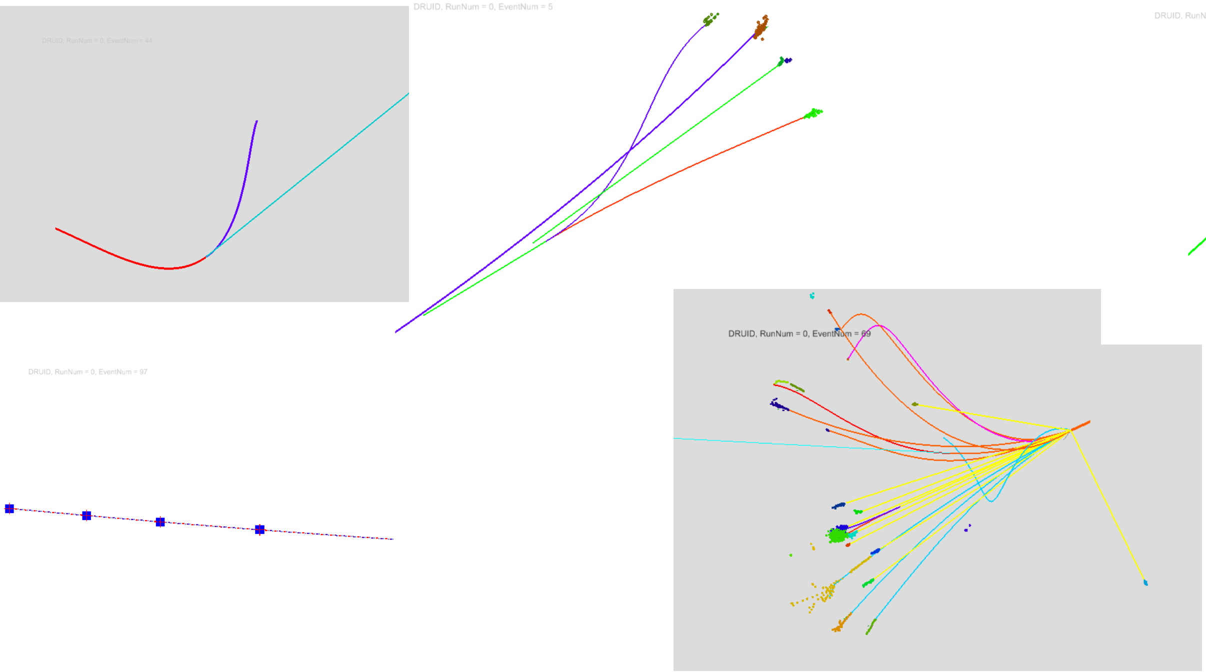


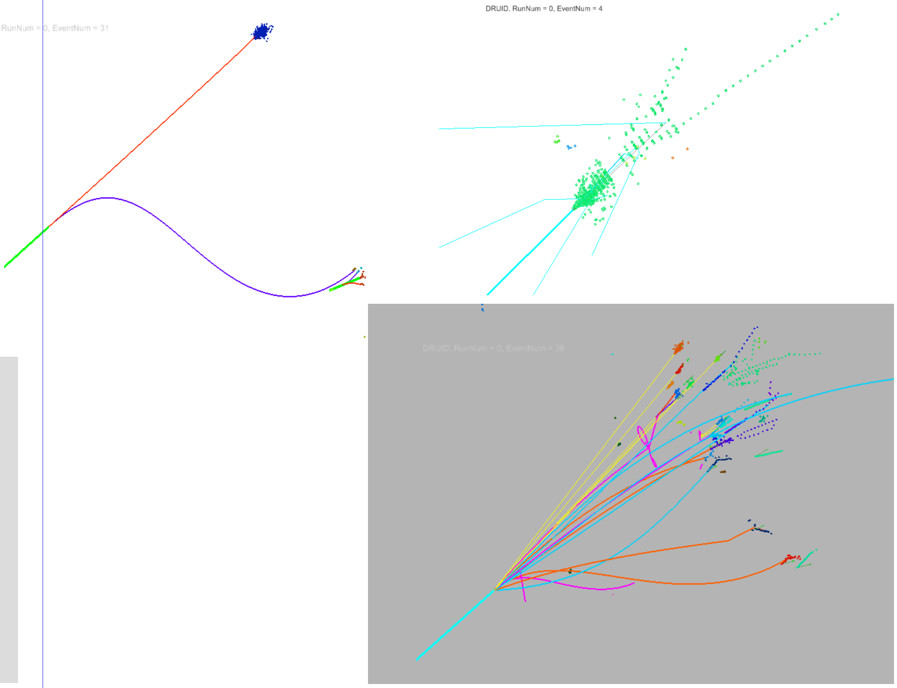






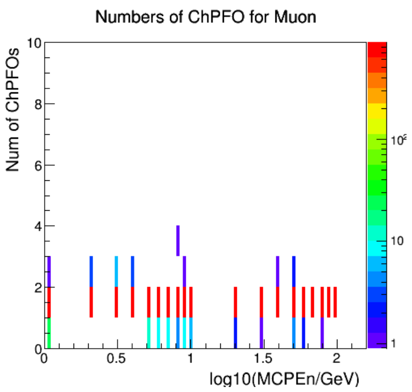


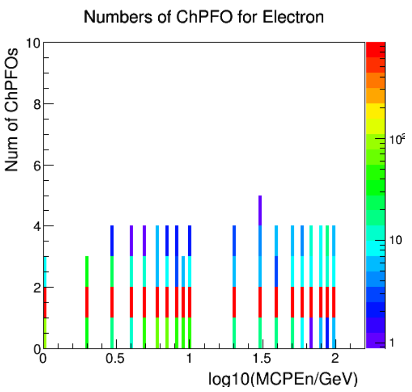


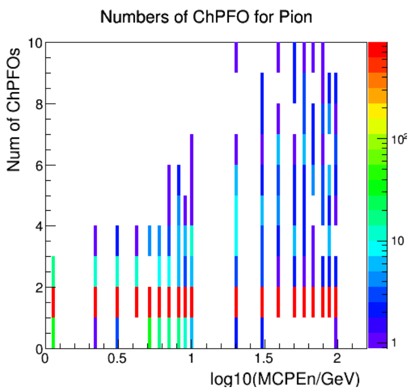


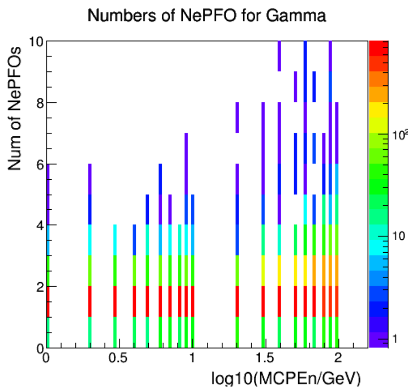
Charged *Neutral*

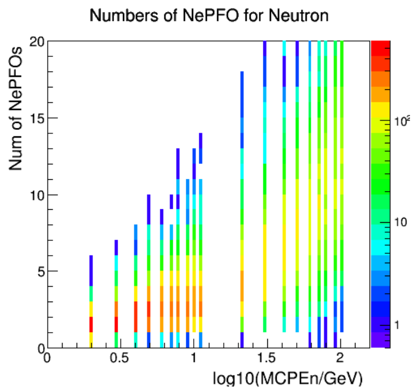
Arbor @ single particle

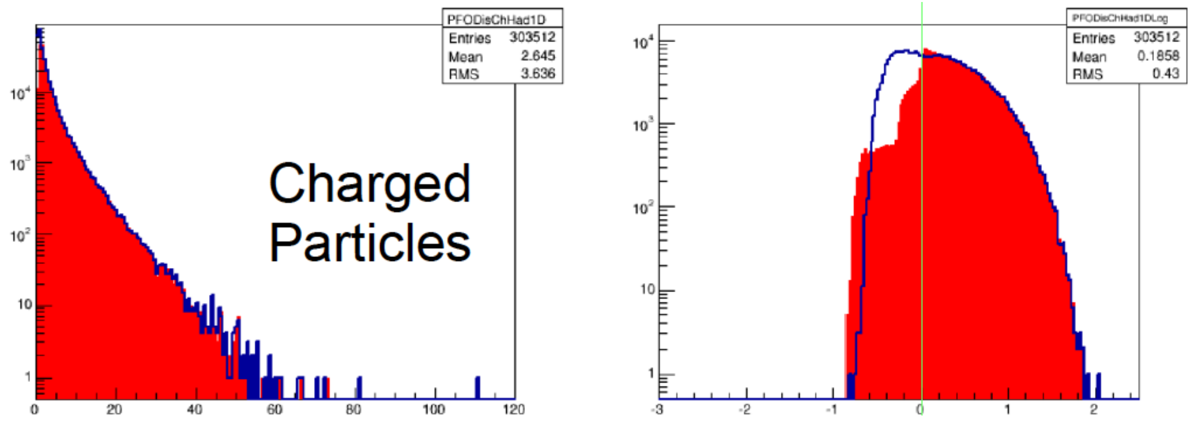

Numbers of ChPFO for Muon

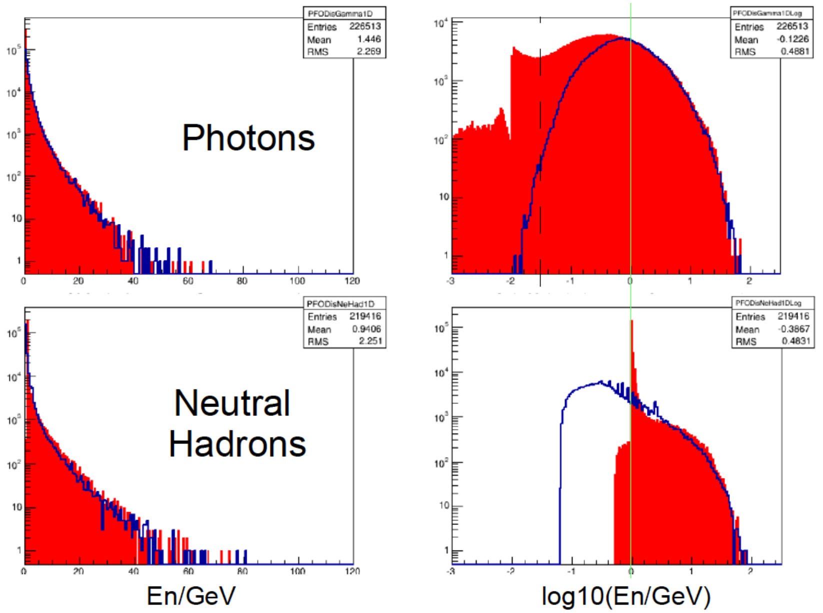

Numbers of ChPFO for Electron


Numbers of ChPFO for Pion


Numbers of NePFO for Gamma

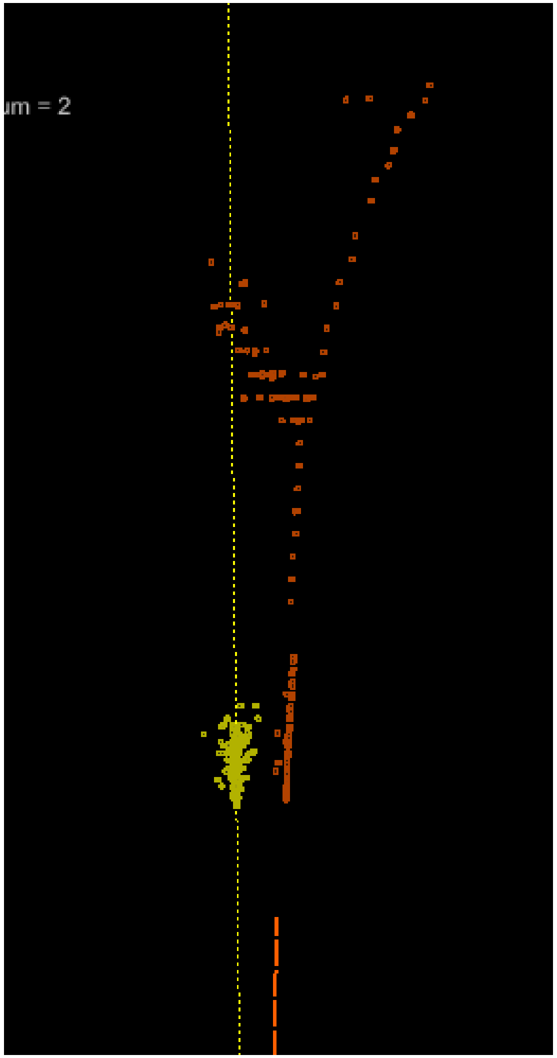

Numbers of NePFO for Neutron

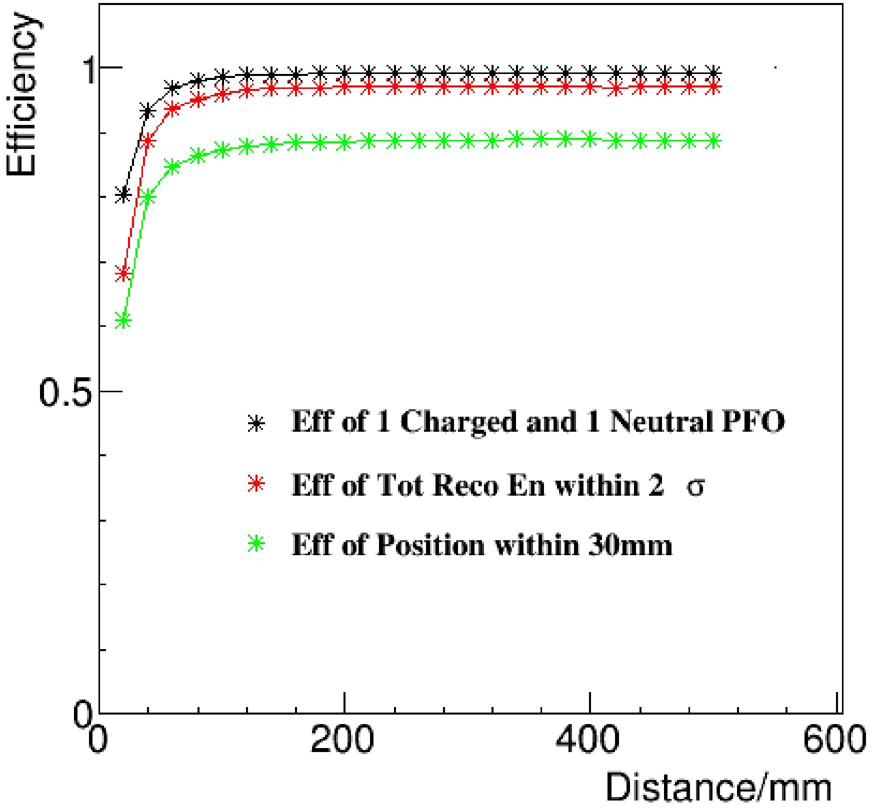

Charged Particles

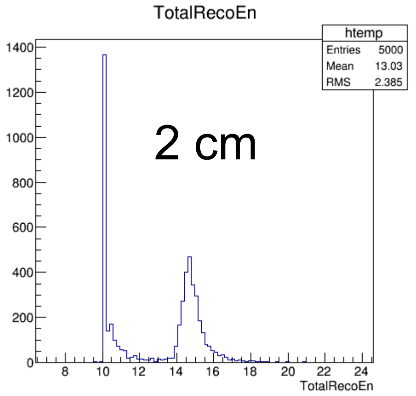

Photons

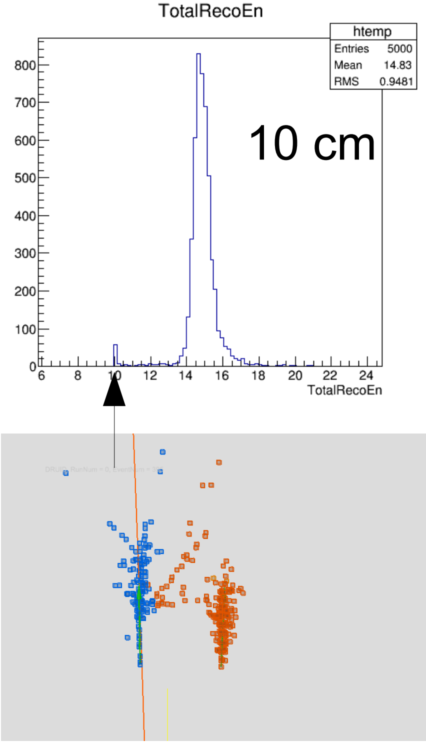
Neutral Hadrons

Separation



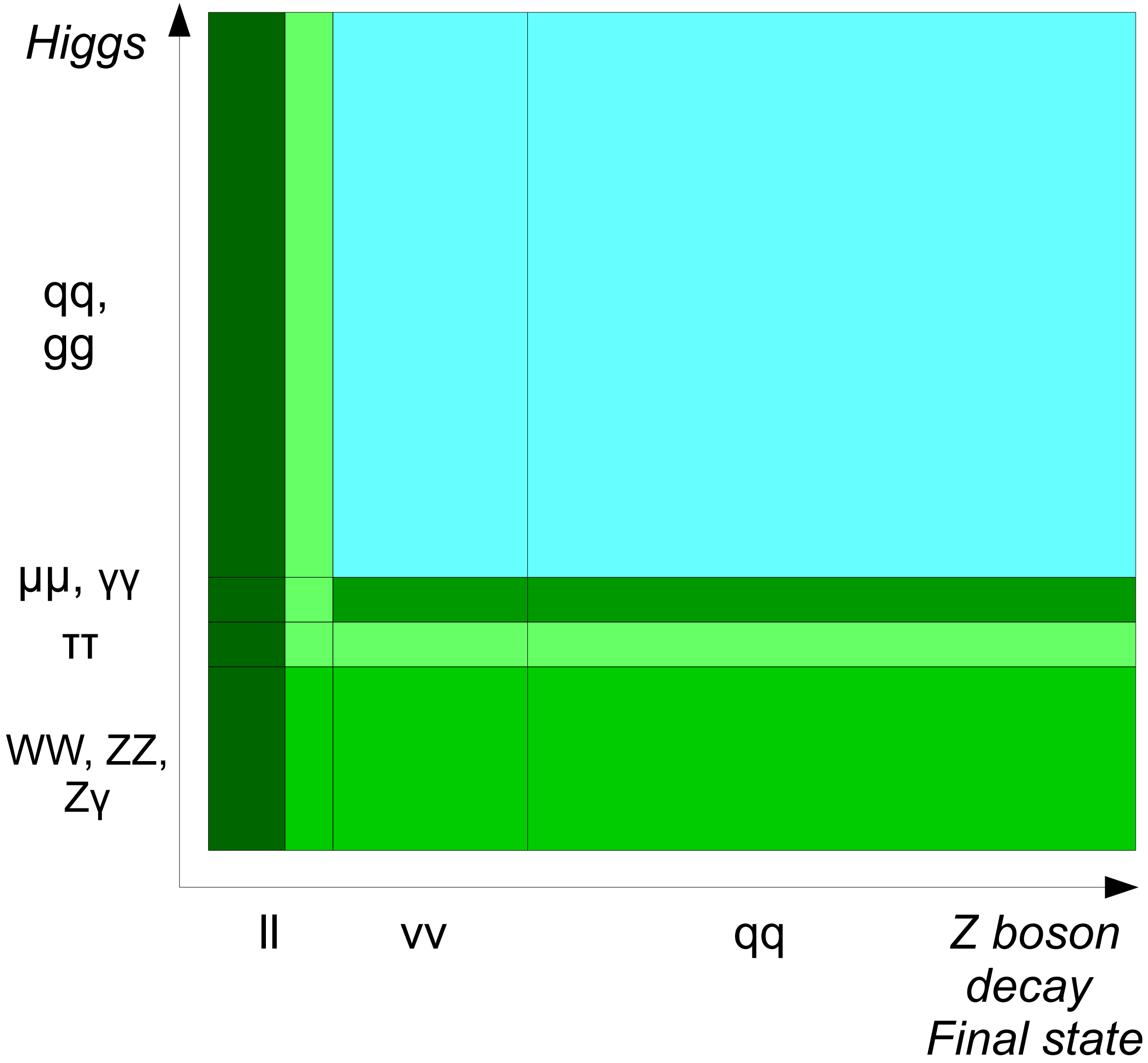






Tiny inefficiency: bridging effect by fragments

Lepton ID



Essential

Signal Classification & Background rejection.

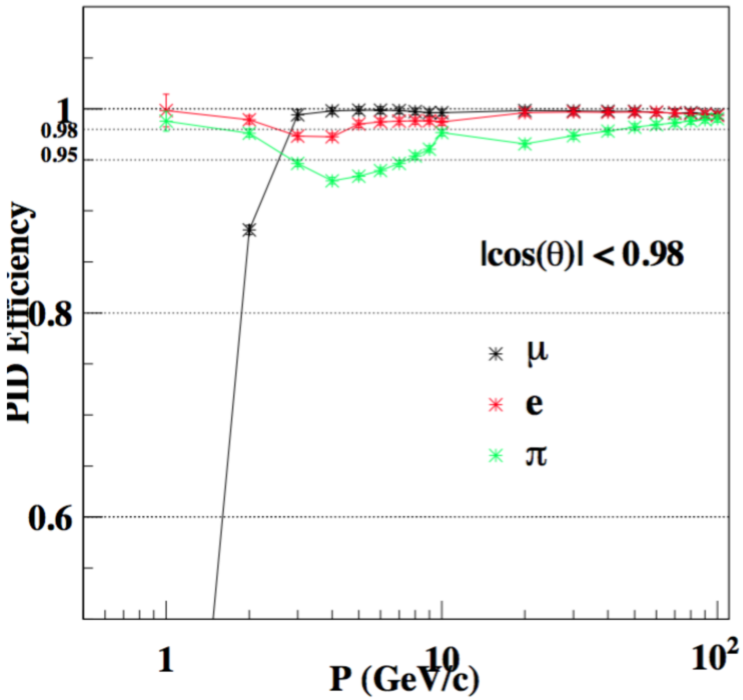
Almost everything you want to measure at electron-positron collider...

Higgs:

Even for H->bb, cc, gg measurement, the lepton number provides useful information for b/c tagging

Lepton ID @ Arbor V3

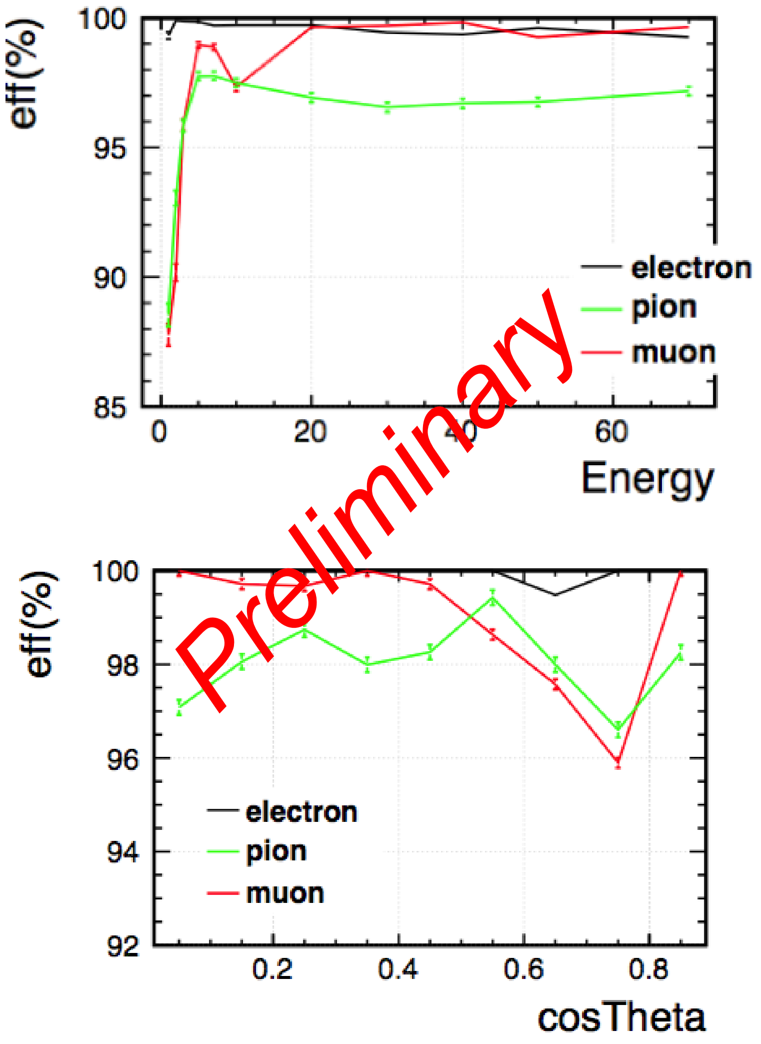
- Developed by **BINSONG**
- E > 10 GeV, Efficiency for Lepton ID > 99%...
- Leads to elegant
  - Higgs recoil analysis with Z -> leptons
  - Higgs->ZZ/WW->leptonic/semi-leptonic final states
    - We are advanced comparing to ILC studies
    - Improve signal efficiency by 2 times comparing to PreCDR
    - More details: See Gang & Yuqian's talk



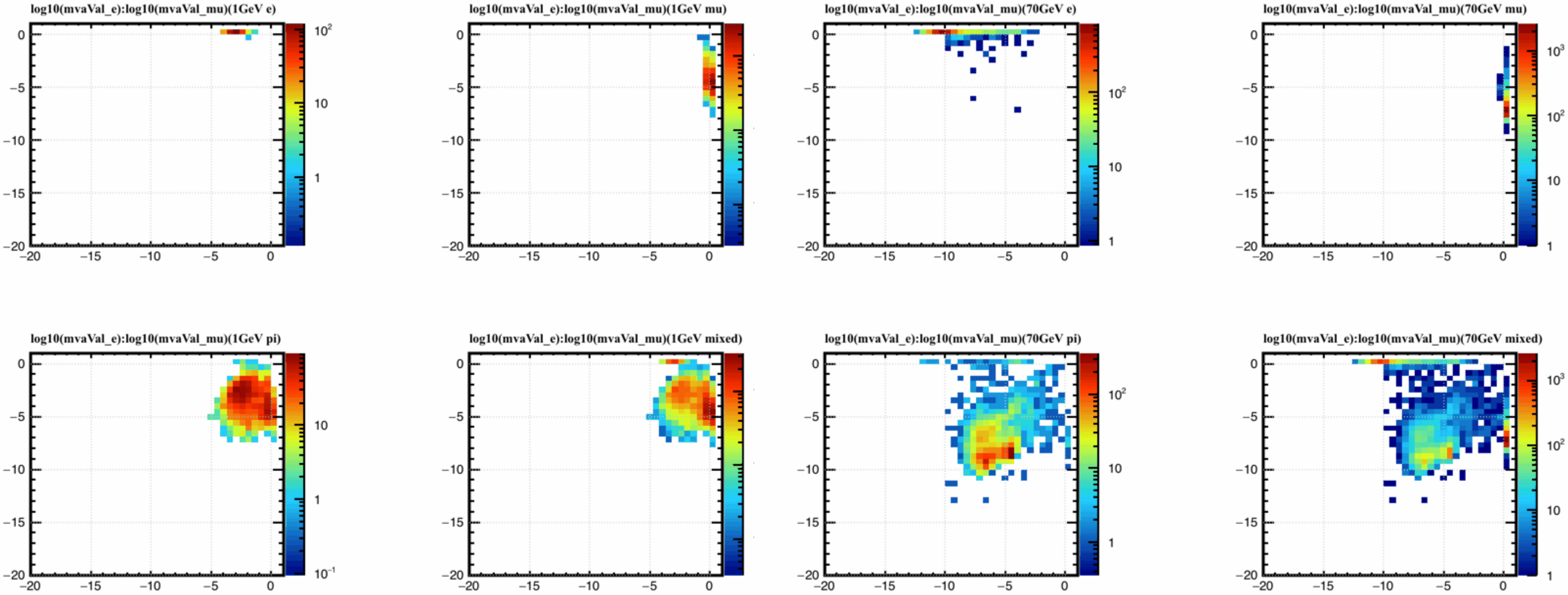
ε^e_e	P^e_μ	P^e_π	P^e_{udf}
P^μ_e	ε^μ_μ	P^μ_π	P^μ_{udf}
P^π_e	P^π_μ	ε^π_π	P^π_{udf}

Lepton ID, next step

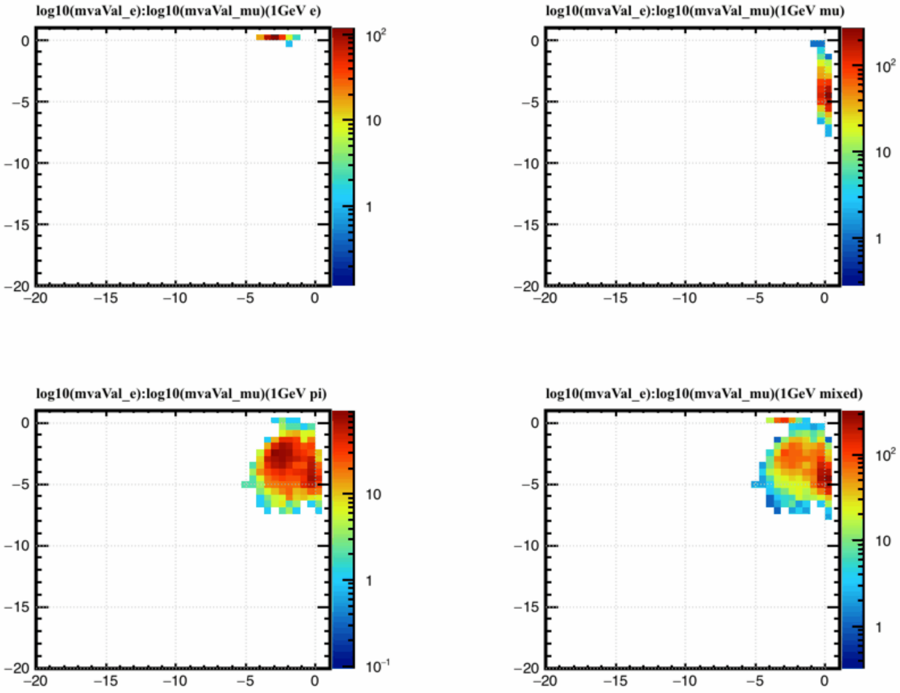
- Developed by **YUDAN** (Joint Ph.D)
- TMVA based
  - For each charged PFO, provide a electron-likeness and muon-likeness value
  - Typical working point:
  - eff > 99% for E > 20 GeV Lepton
  - eff > 97% for E > 5 GeV Pion
- To be polished, encapsulated & integrated
- More importantly: We understand
  - We know where/why the limit is;
  - We know how to improve



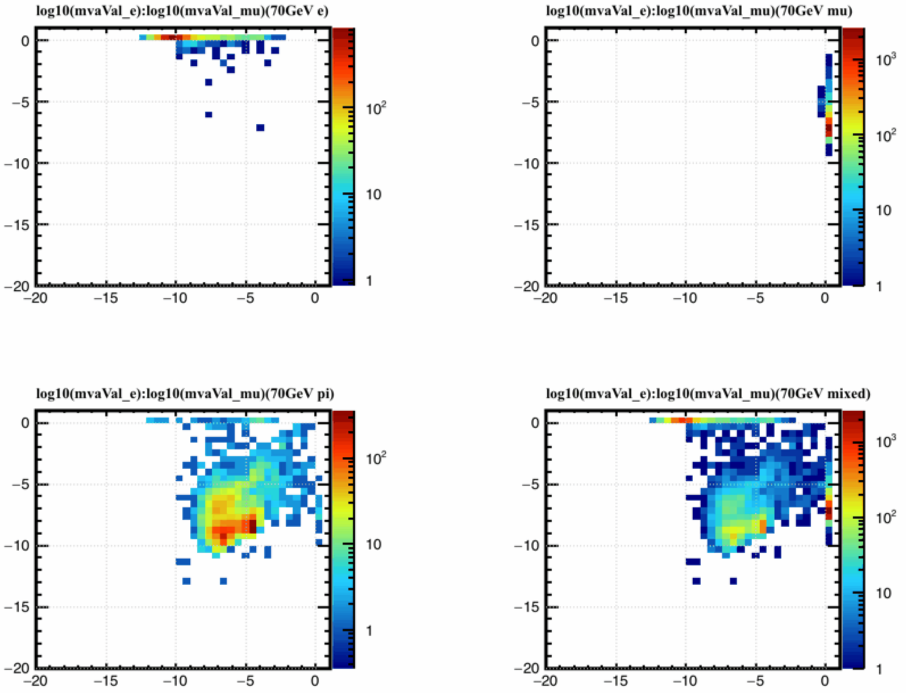
barrel **Endcap**



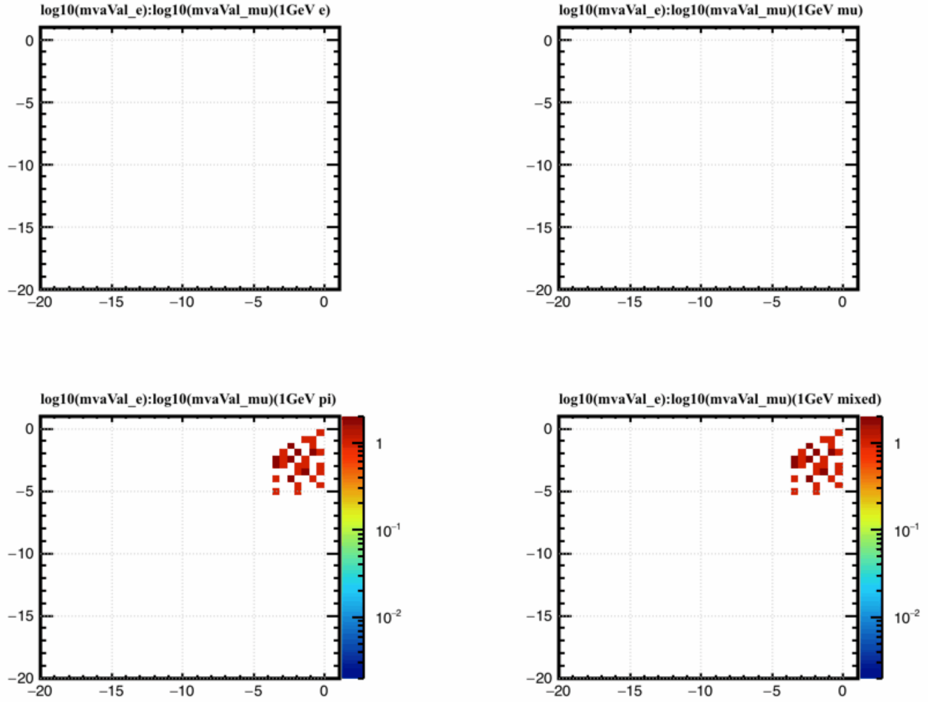
barrel



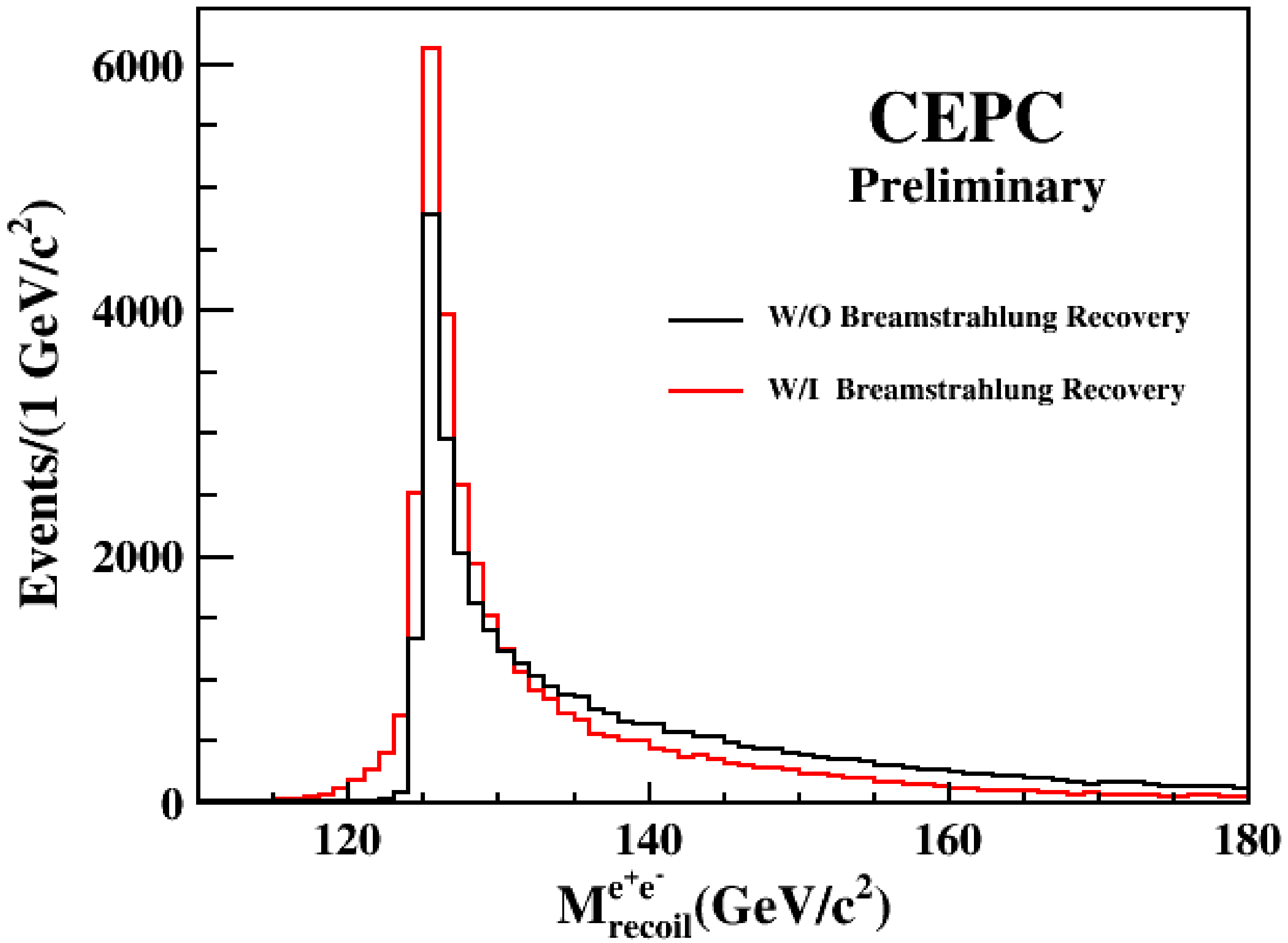
Endcap



Overlap

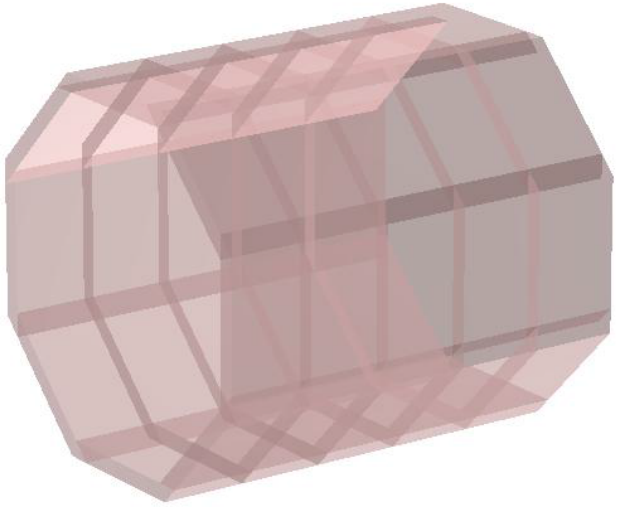


Bremsstrahlung photon recovery of electron/positron

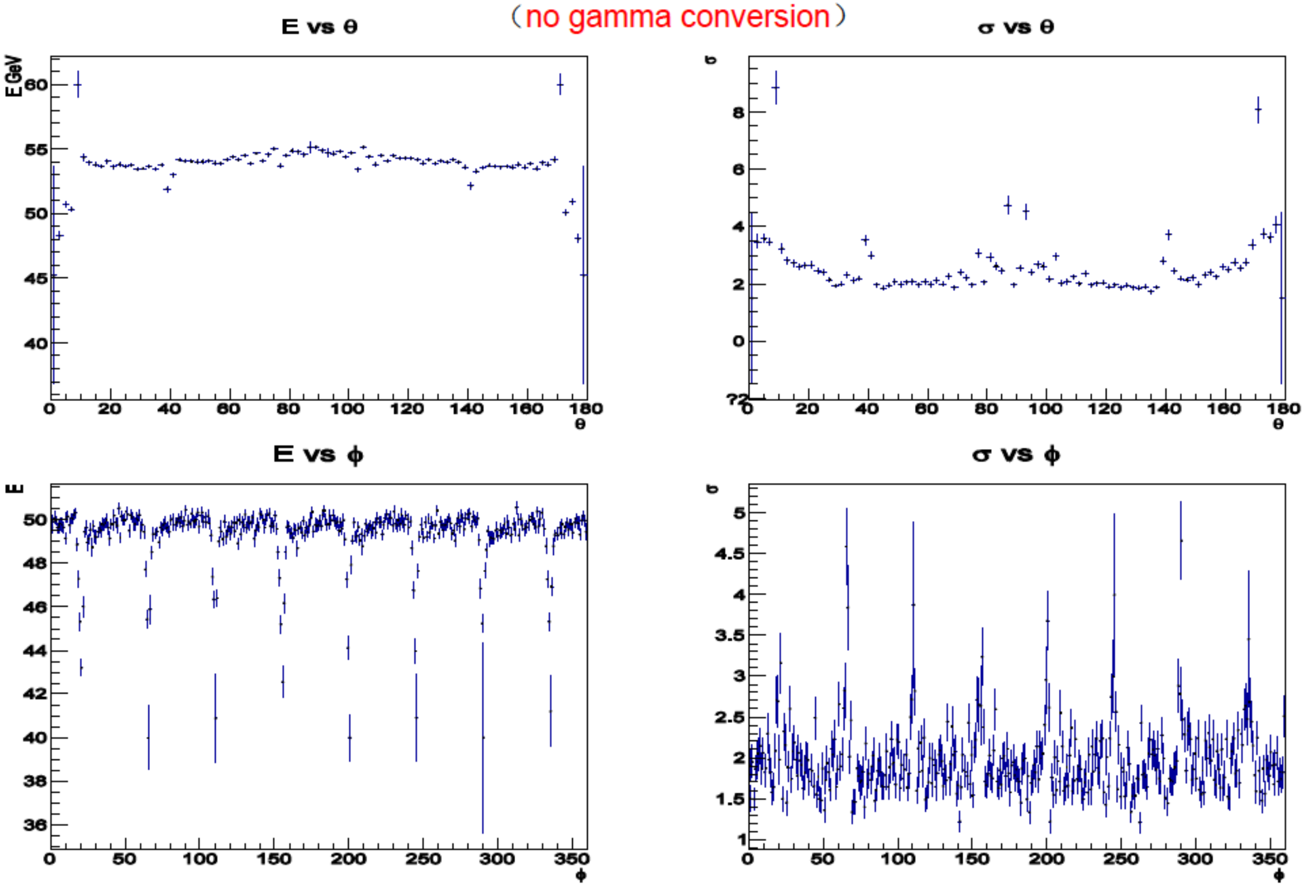


- *Developed by Zhenxing, Binsong, Wanglei, etc*

Arbor: photon reconstruction



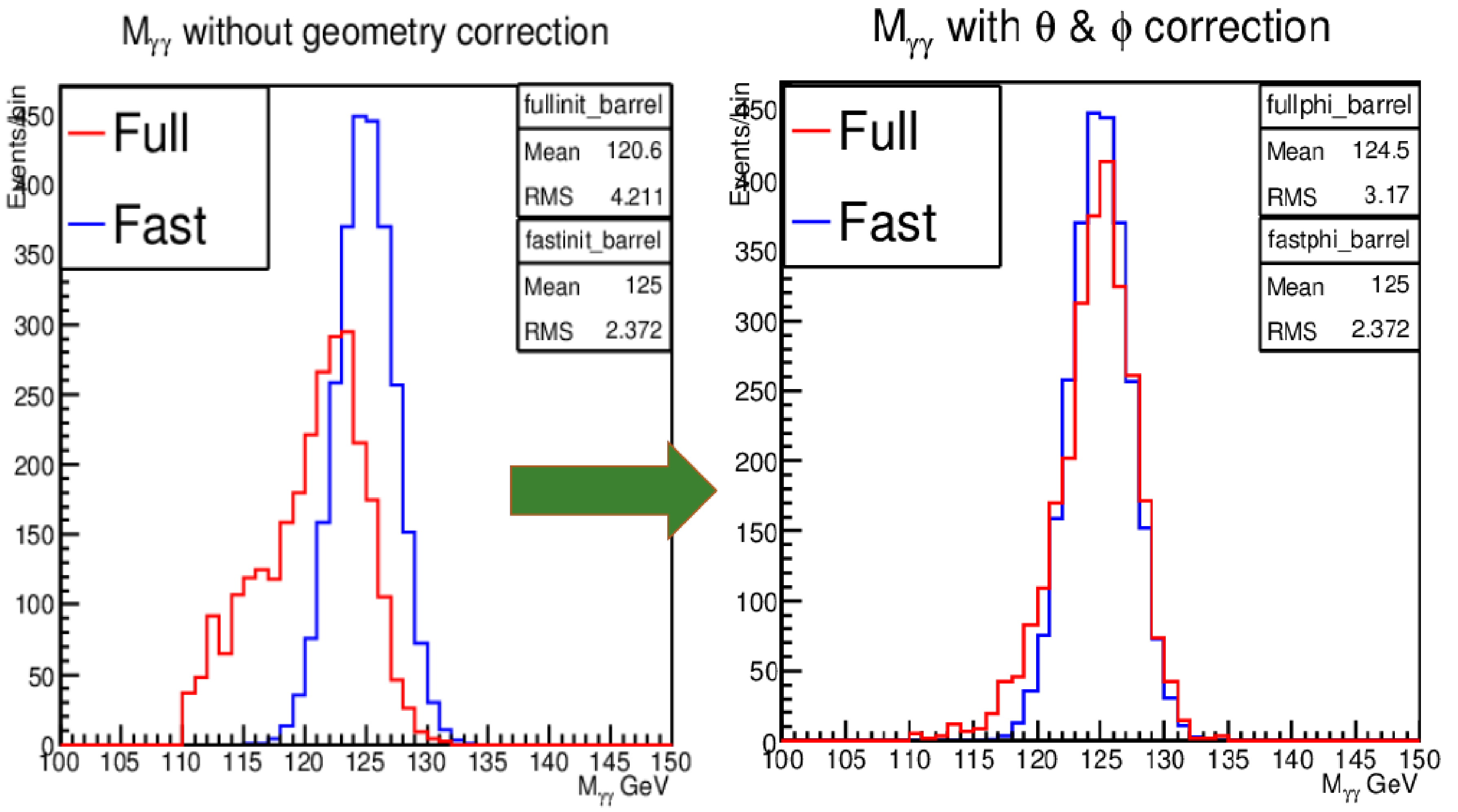
ECAL Barrel of ILD/CEPC_v1



（no gamma conversion）

Angular Correlation of EM Shower energy response

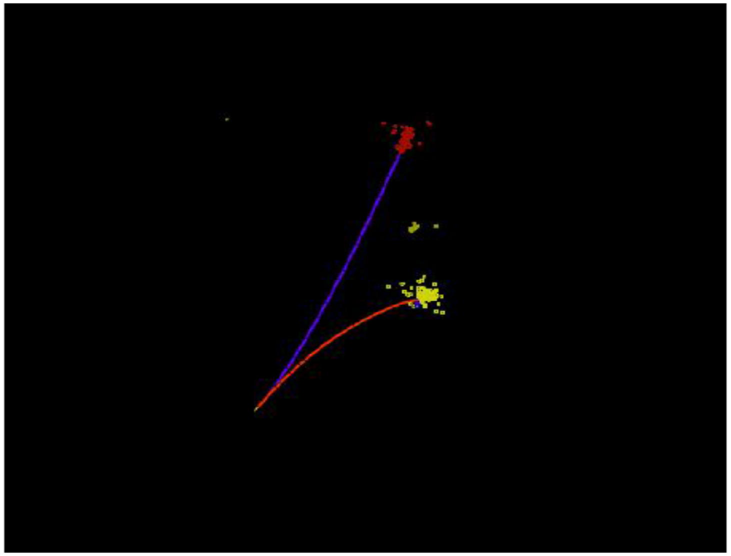
Arbor: photon reconstruction

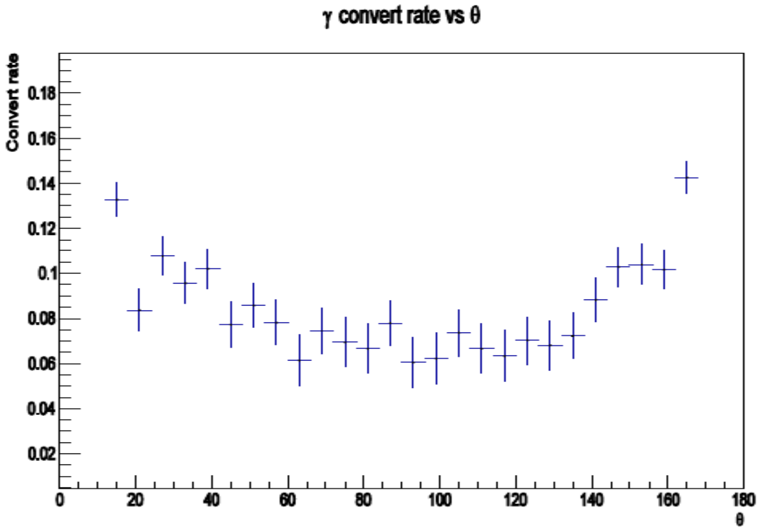


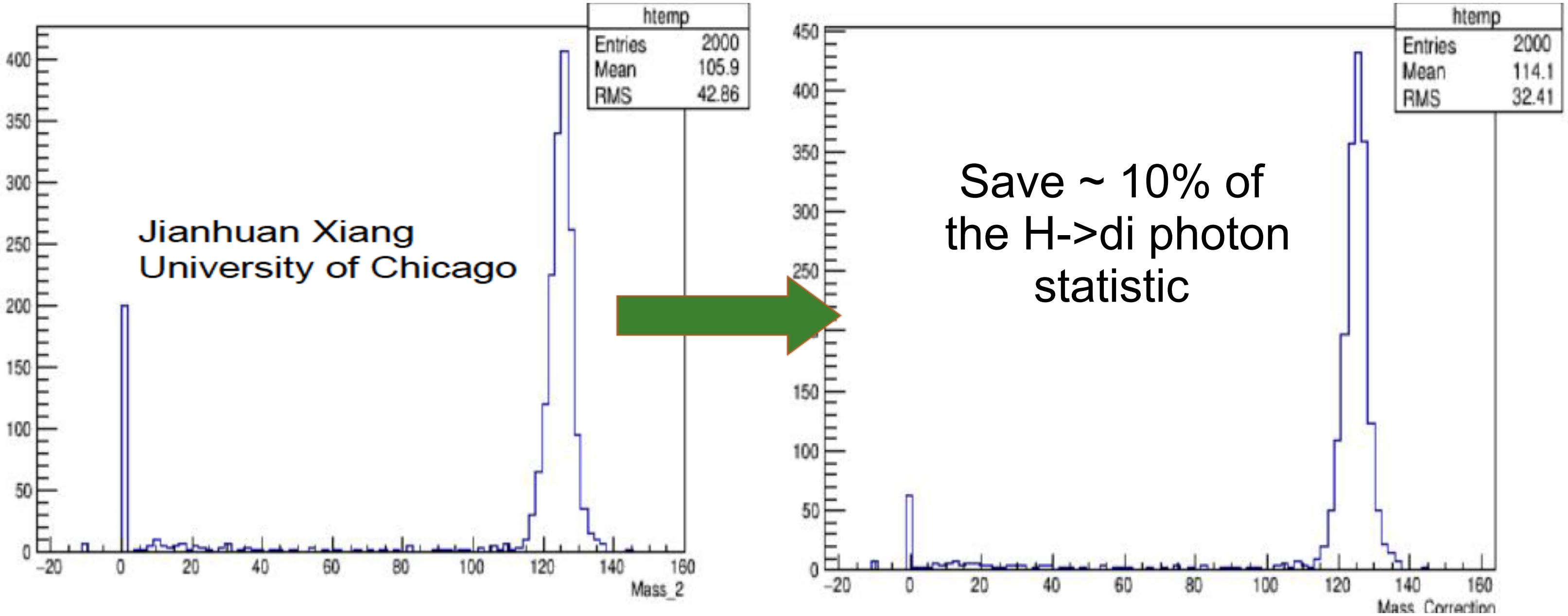
$M_{\gamma\gamma}$ without geometry correction

$M_{\gamma\gamma}$ with θ & ϕ correction

Photon conversion & recovery



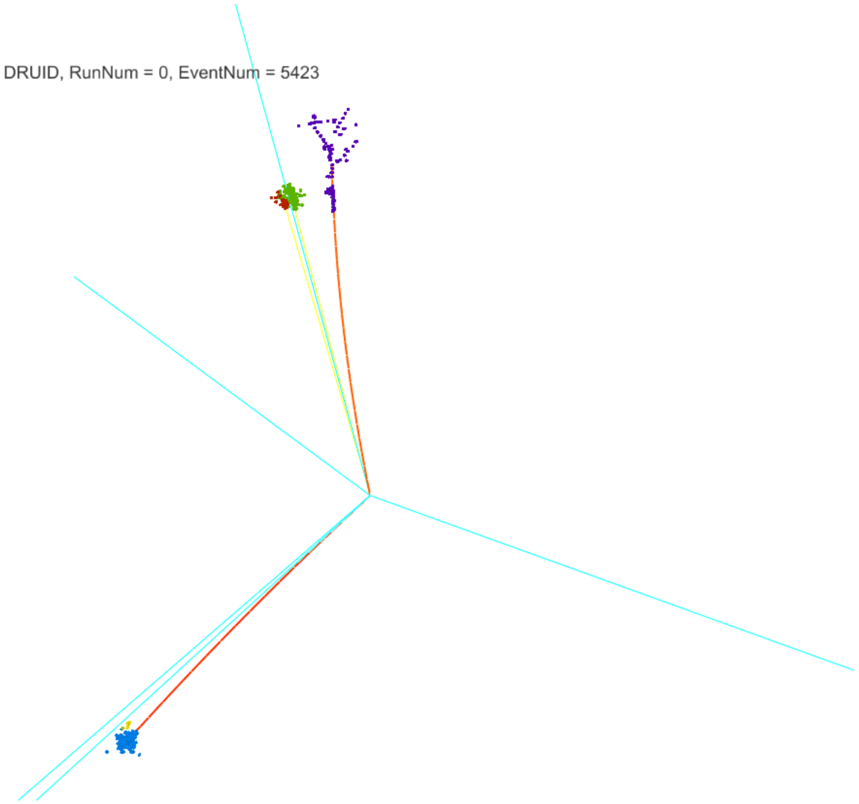


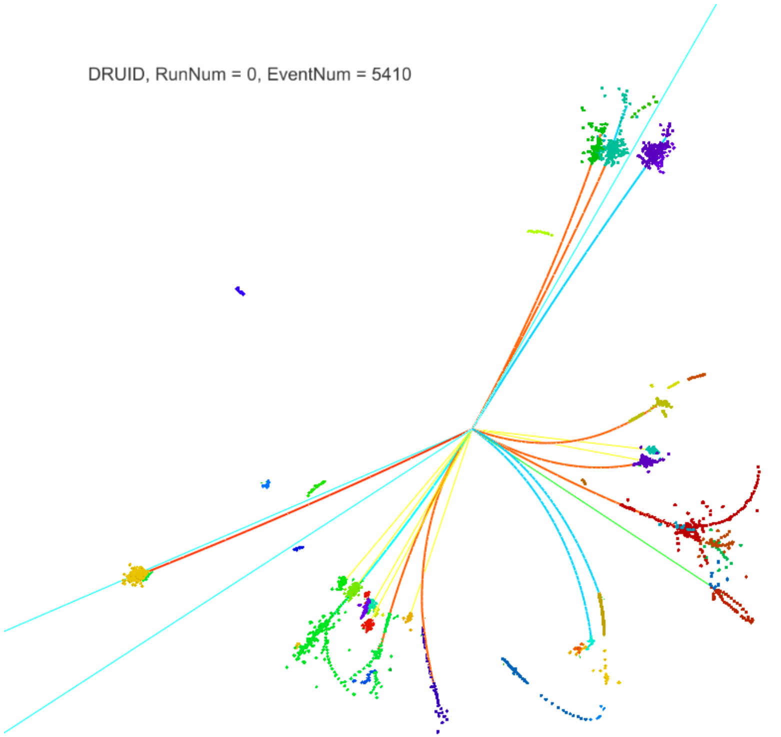


Jianhuan Xiang
University of Chicago

Save ~ 10% of the H->di photon statistic

Arbor: Tau reconstruction



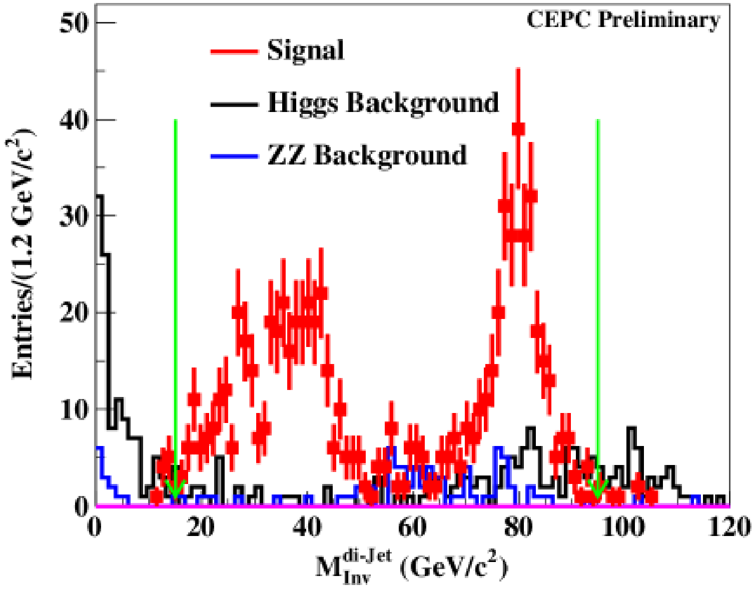


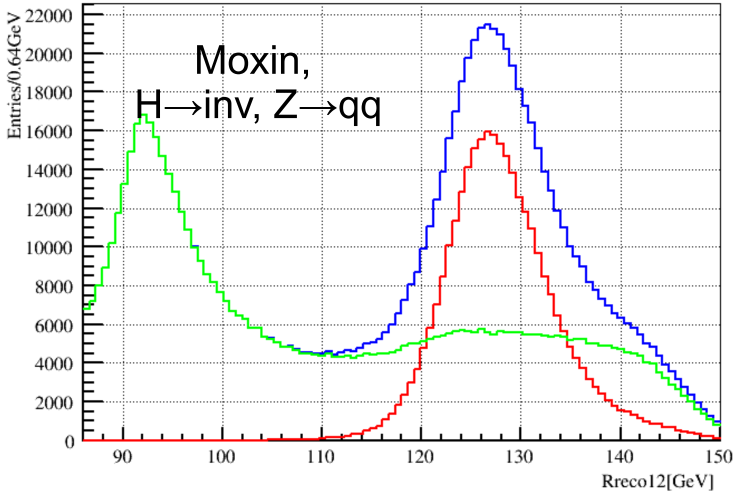
To be covered soon...

- In no-jet environment: counting number of charged particle – (pions & leptons), photons (pi0s) + restrict impact parameters leads to very high efficiency in Tau finding:
  - At inclusive Higgs decay sample: Efficiency ~ 98% for of H→tautau event finding, with llH and vvH final state. The remaining bkgrds are H→WW/ZZ→leptonic/tau final state
  - More detail: see Gang's talk

Arbor: JER/MET

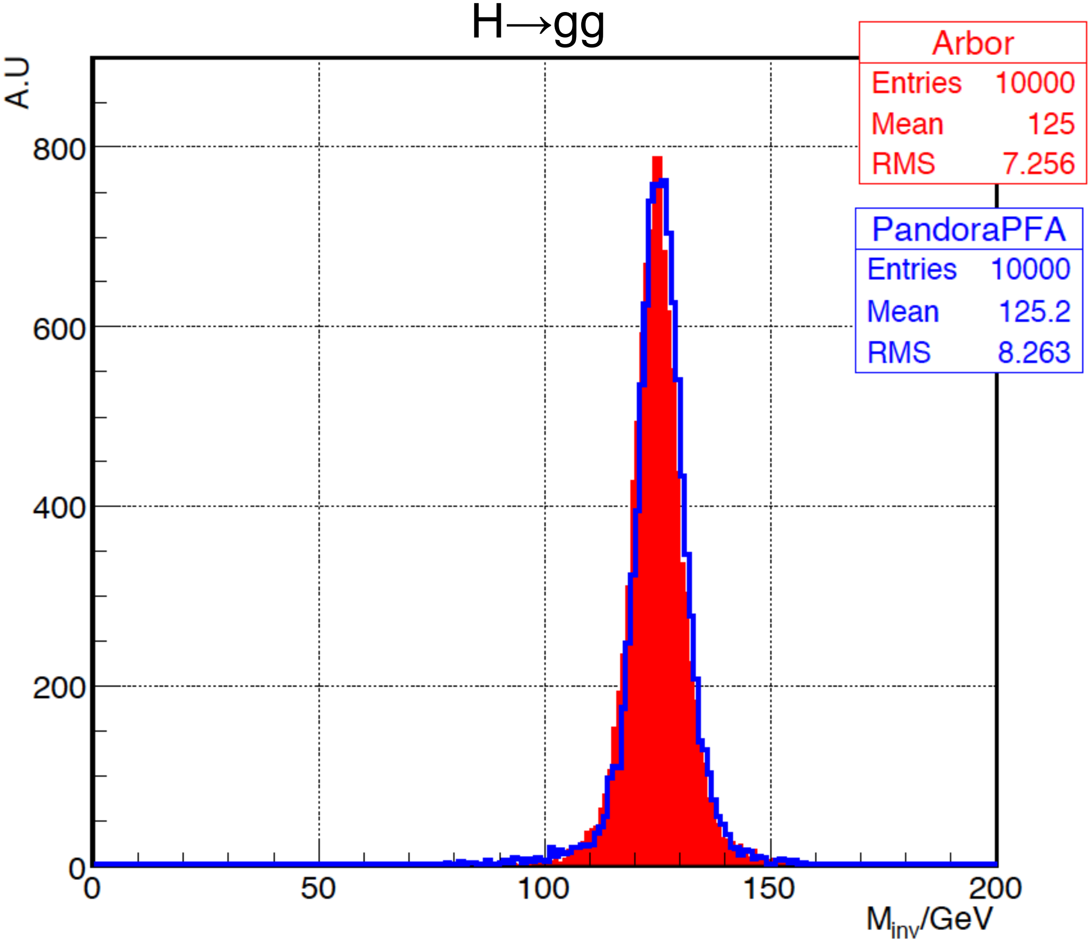
Liao libo, H→WW*→lvqq





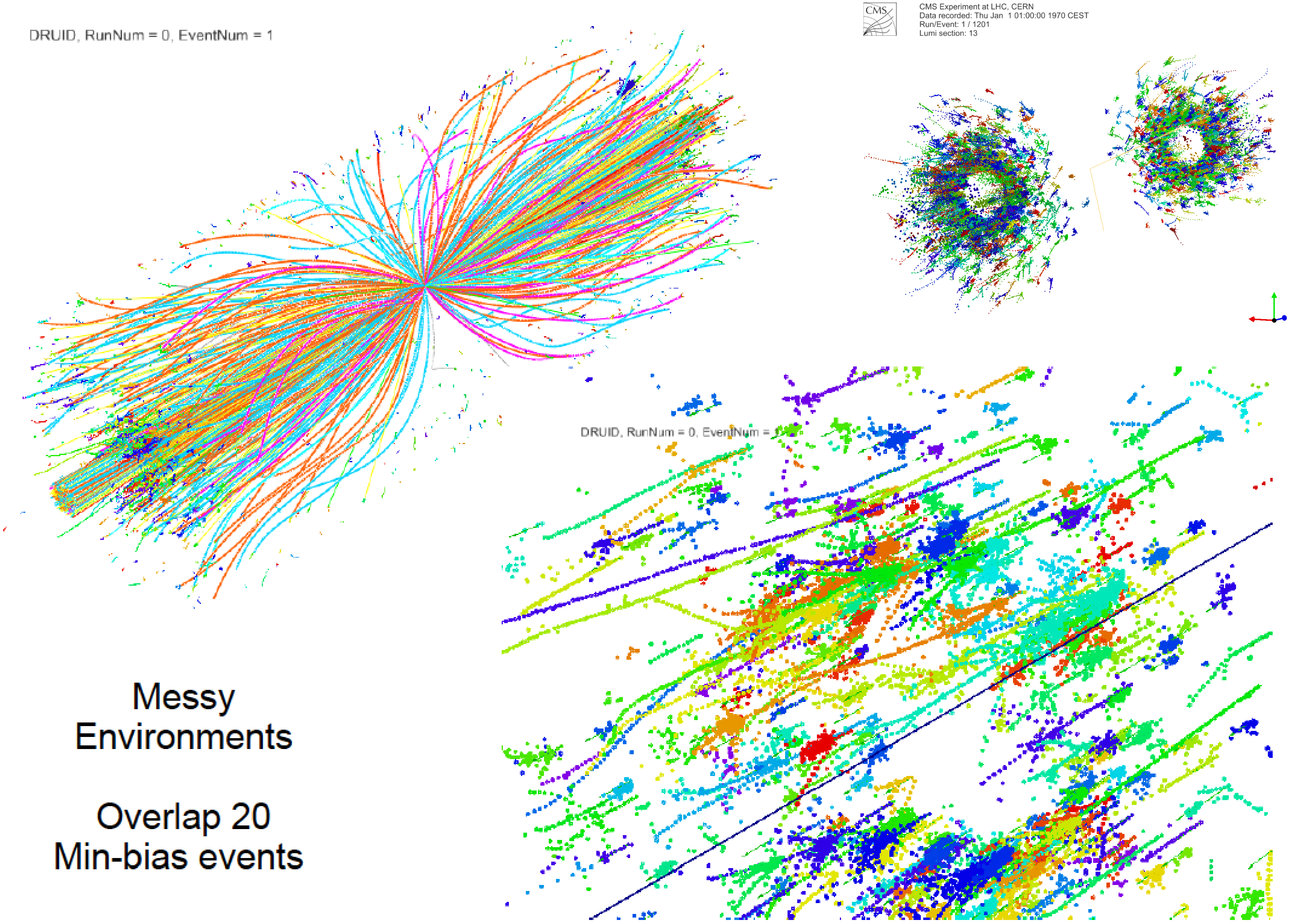
Moxin,
H→inv, Z→qq

H→gg



- MET: usually no ambiguity;
- Jet: Highly depending on Jet clustering if #Jet > 2...

KD algorithm boost: $N^2 \rightarrow N\log(N)$

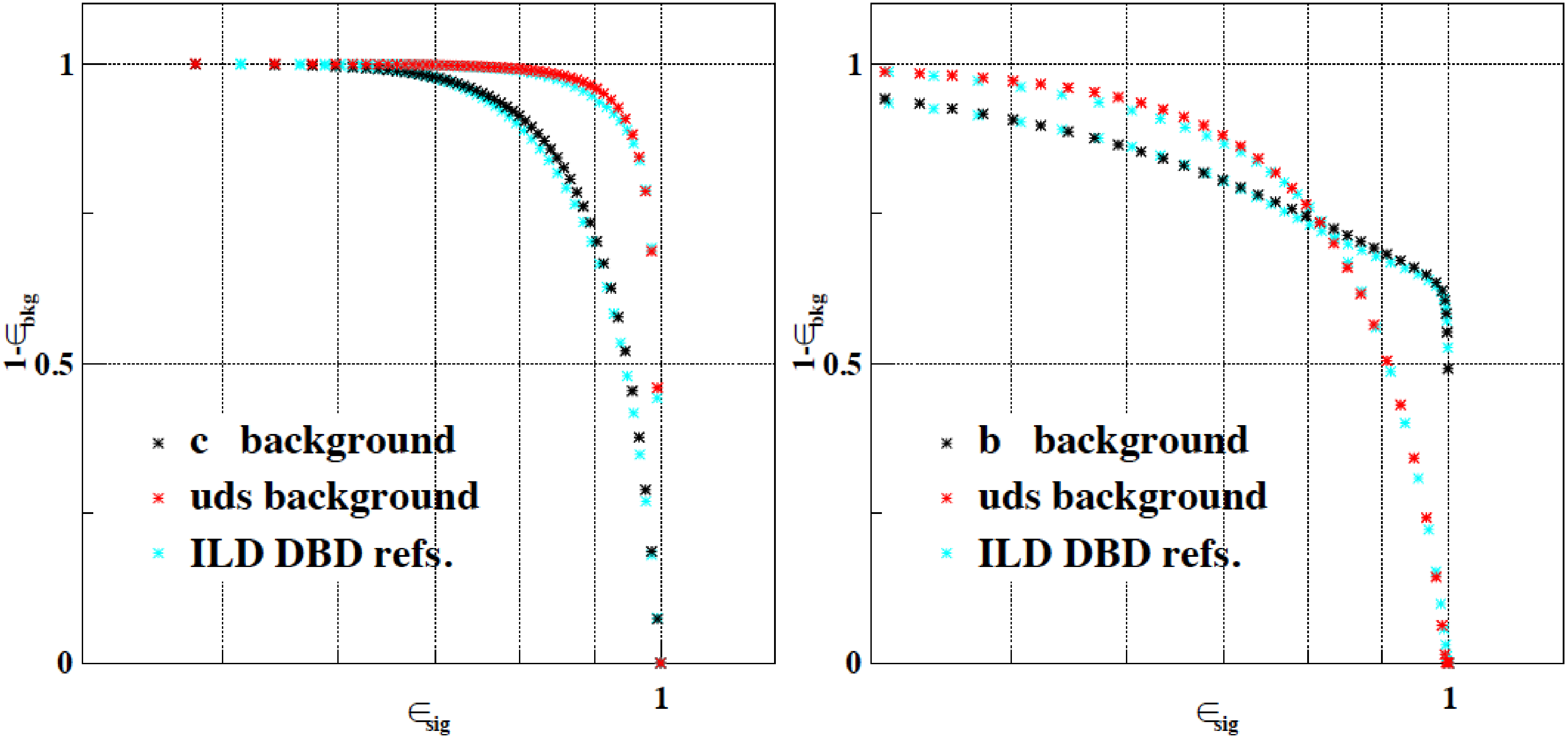


Messy
Environments

Overlap 20
Min-bias events

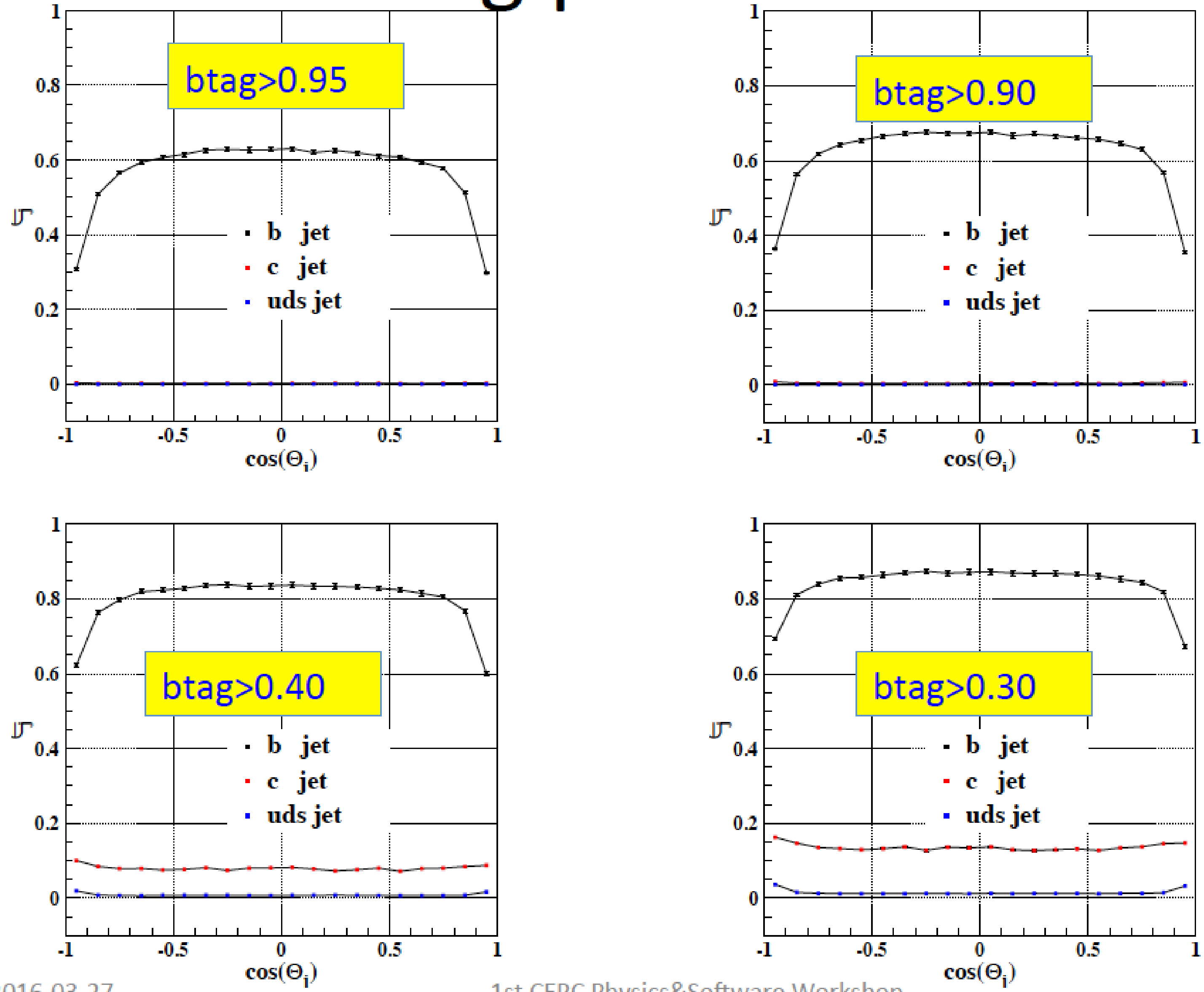
CEPC Event: speed up by 1 order of magnitude

Flavor Tagging

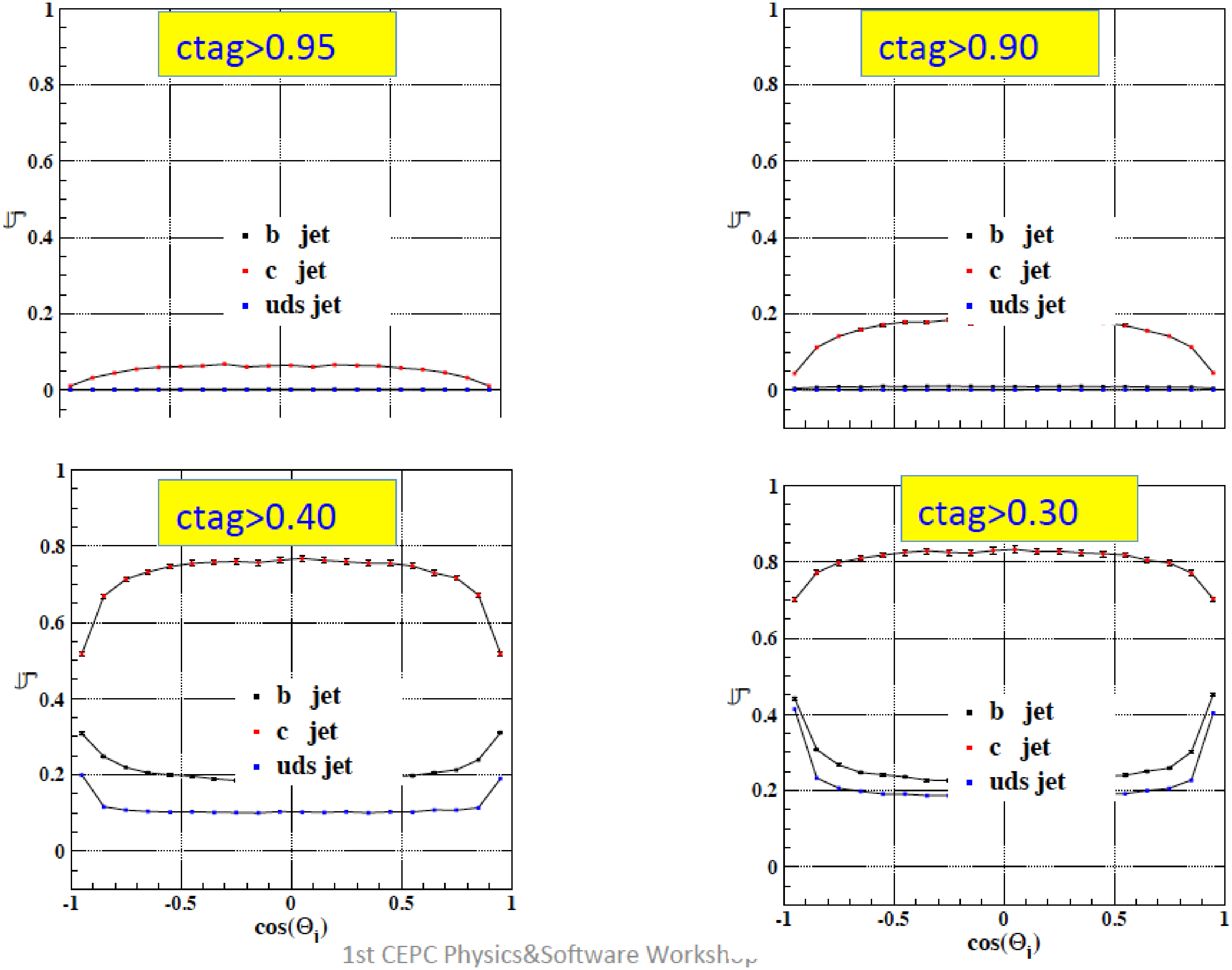


TMVA based method from ILC Study:
http://indico.ihep.ac.cn/event/5592/contribution/16/material/slides/0.pdf

btag performance



ctag performance



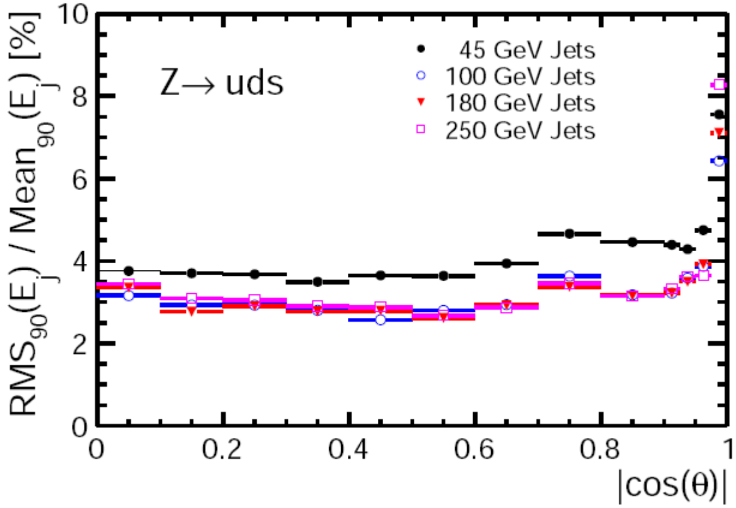
2016-03-27 1st CEPC Physics&Software Worksh

Analysis tool

FSClasser Vs self-wrote analysis code...

Pandora, Dante Gabriel Rossetti, 1869

Pandora is great



100 GeV Jets

$$\alpha = 0.315\left(\frac{B}{4}\right)^{-0.19}\left(\frac{R}{1.68}\right)^{-0.49}(1 + 6.3e^{-\frac{N}{8.0}})$$

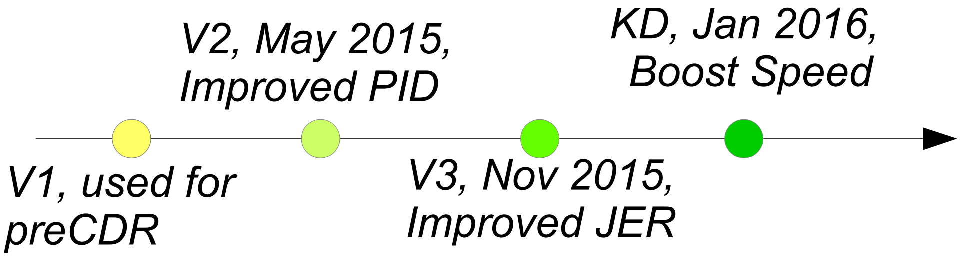
180 GeV Jets

$$\alpha = 0.42\left(\frac{B}{4}\right)^{-0.31}\left(\frac{R}{1.78}\right)^{-0.61}(1 + 21.6e^{-\frac{N}{7.1}})$$

2007, M. Thomson, Optimising GLDC for PFA

http://www.hep.ph.ic.ac.uk/calice/others/070530lcws07/thomson1.pdf

Vendangeuse, William Adolphe Bouguereau, 1875



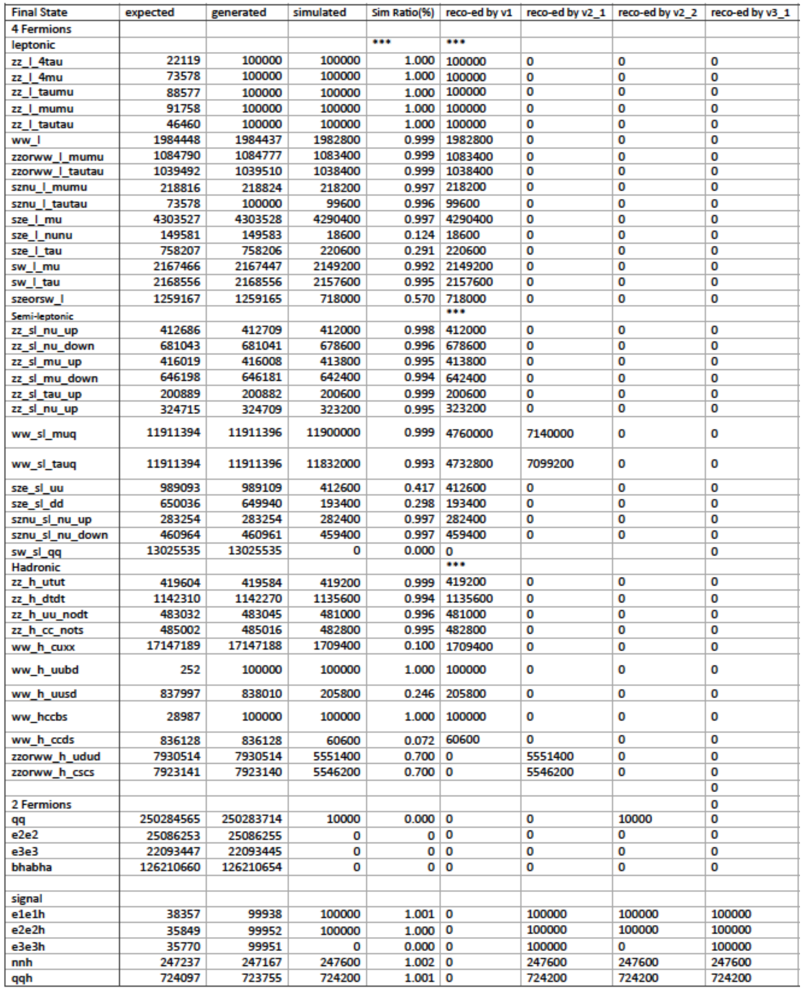
V1, used for preCDR

V2, May 2015, Improved PID

V3, Nov 2015, Improved JER

KD, Jan 2016, Boost Speed

Table 1

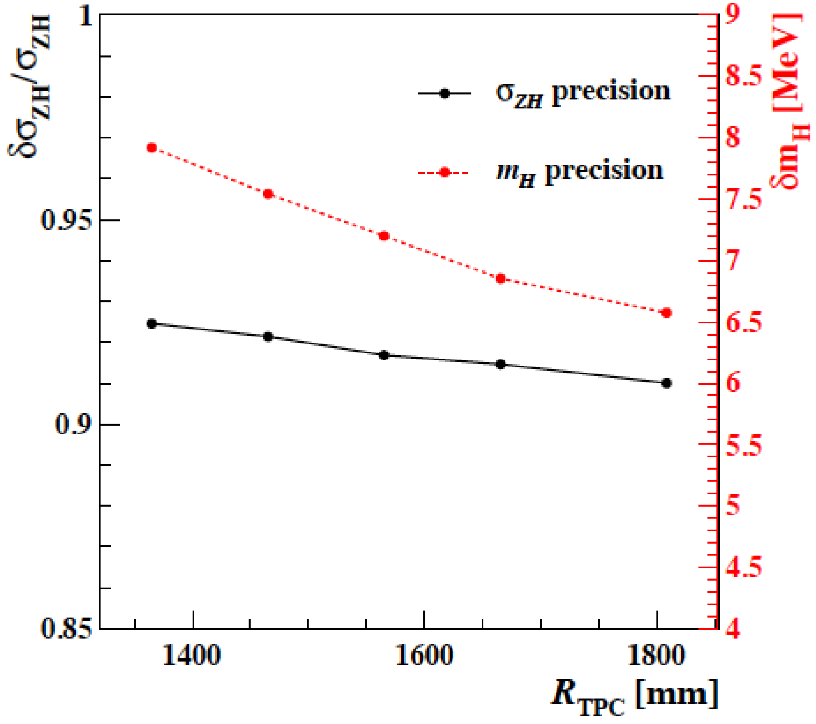


Final State	expected	generated	simulated	Sim Ratio(%)	reco-ed by v1	reco-ed by v2_1	reco-ed by v2_2	reco-ed by v3_1
4 Fermions								
leptonic				***	***			
zz_l_4tau	22119	100000	100000	1.000	100000	0	0	0
zz_l_4mu	73578	100000	100000	1.000	100000	0	0	0
zz_l_taumu	88577	100000	100000	1.000	100000	0	0	0
zz_l_mumu	91758	100000	100000	1.000	100000	0	0	0
zz_l_tautau	46460	100000	100000	1.000	100000	0	0	0
ww_l	1984448	1984437	1982800	0.999	1982800	0	0	0
zzorww_l_mumu	1084790	1084777	1083400	0.999	1083400	0	0	0
zzorww_l_tautau	1039492	1039510	1038400	0.999	1038400	0	0	0
sznu_l_mumu	218816	218824	218200	0.997	218200	0	0	0
sznu_l_tautau	73578	100000	99600	0.996	99600	0	0	0
sze_l_mu	4303527	4303528	4290400	0.997	4290400	0	0	0
sze_l_nunu	149581	149583	18600	0.124	18600	0	0	0
sze_l_tau	758207	758206	220600	0.291	220600	0	0	0
sw_l_mu	2167466	2167447	2149200	0.992	2149200	0	0	0
sw_l_tau	2168556	2168556	2157600	0.995	2157600	0	0	0
szeorsw_l	1259167	1259165	718000	0.570	718000	0	0	0
Semi-leptonic					***			
zz_sl_nu_up	412686	412709	412000	0.998	412000	0	0	0
zz_sl_nu_down	681043	681041	678600	0.996	678600	0	0	0
zz_sl_mu_up	416019	416008	413800	0.995	413800	0	0	0
zz_sl_mu_down	646198	646181	642400	0.994	642400	0	0	0
zz_sl_tau_up	200889	200882	200600	0.999	200600	0	0	0
zz_sl_nu_up	324715	324709	323200	0.995	323200	0	0	0
ww_sl_muq	11911394	11911396	11900000	0.999	4760000	7140000	0	0
ww_sl_tauq	11911394	11911396	11832000	0.993	4732800	7099200	0	0
sze_sl_uu	989093	989109	412600	0.417	412600	0	0	0
sze_sl_dd	650036	649940	193400	0.298	193400	0	0	0
sznu_sl_nu_up	283254	283254	282400	0.997	282400	0	0	0
sznu_sl_nu_down	460964	460961	459400	0.997	459400	0	0	0
sw_sl_qq	13025535	13025535	0	0.000	0			0
Hadronic					***			
zz_h_utut	419604	419584	419200	0.999	419200	0	0	0
zz_h_dtdt	1142310	1142270	1135600	0.994	1135600	0	0	0
zz_h_uu_nodt	483032	483045	481000	0.996	481000	0	0	0
zz_h_cc_nots	485002	485016	482800	0.995	482800	0	0	0
ww_h_cuxx	17147189	17147188	1709400	0.100	1709400	0	0	0
ww_h_uubd	252	100000	100000	1.000	100000	0	0	0
ww_h_uusd	837997	838010	205800	0.246	205800	0	0	0
ww_hccbs	28987	100000	100000	1.000	100000	0	0	0
ww_h_ccds	836128	836128	60600	0.072	60600	0	0	0
zzorww_h_udud	7930514	7930514	5551400	0.700	0	5551400	0	0
zzorww_h_cscs	7923141	7923140	5546200	0.700	0	5546200	0	0
								0
2 Fermions								0
qq	250284565	250283714	10000	0.000	0	0	10000	0
e2e2	25086253	25086255	0	0	0	0	0	0
e3e3	22093447	22093445	0	0	0	0	0	0
bhabha	126210660	126210654	0	0	0	0	0	0
signal								
e1e1h	38357	99938	100000	1.001	0	100000	100000	100000
e2e2h	35849	99952	100000	1.000	0	100000	100000	100000
e3e3h	35770	99951	0	0.000	0	100000	0	100000
nnh	247237	247167	247600	1.002	0	247600	247600	247600
qqh	724097	723755	724200	1.001	0	724200	724200	724200

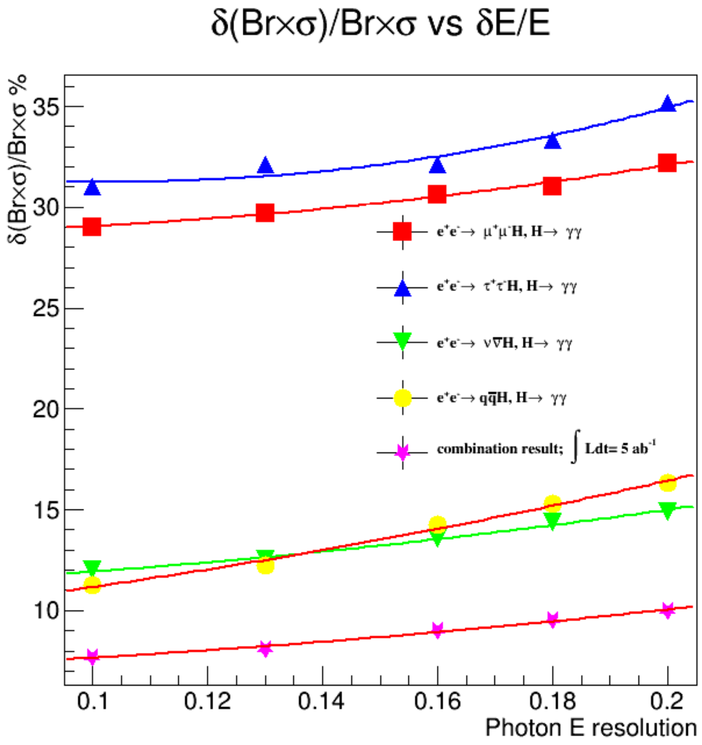
Short of resource...


Optimisation

TPC Radius & ECAL resolution



δ(Br×σ)/Br×σ vs δE/E



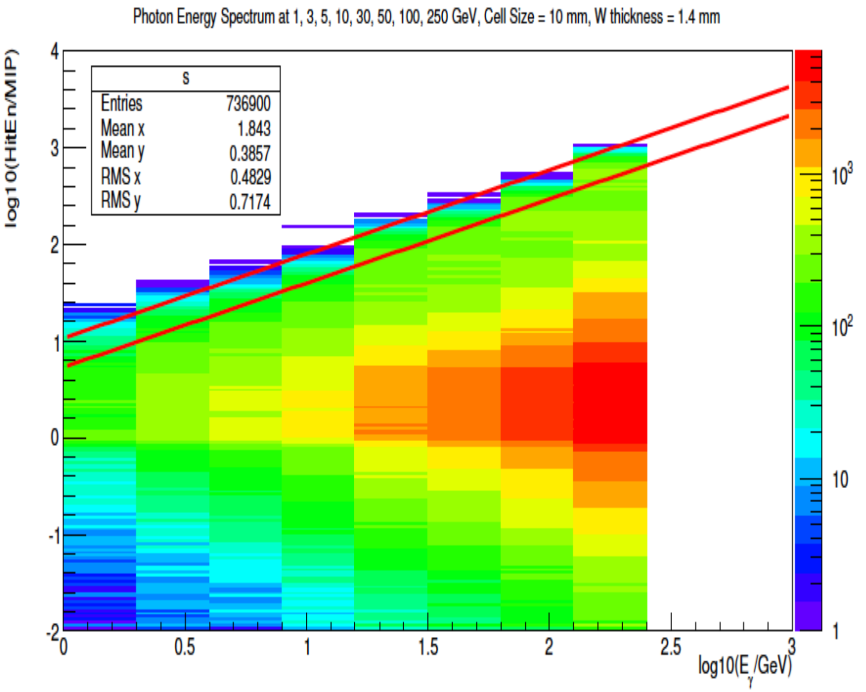
$$\delta m_H = 36.286 \times (1+0.092\times e^{-1.820\cdot R_{\rm TPC}})\mathrm{MeV}.$$

$$\frac{\delta\sigma_{ZH}}{\sigma_{ZH}} = 0.485 \times (1+e^{-0.094\cdot R_{\rm TPC}})$$

H->di photon branching ratio measurement

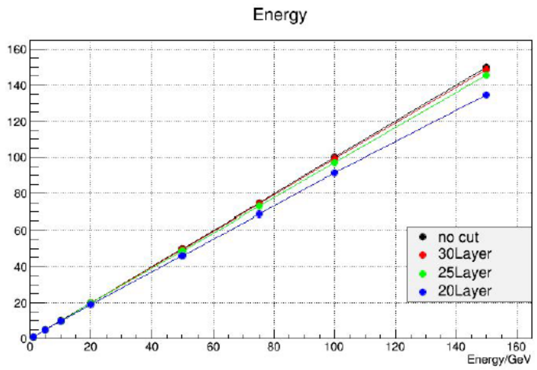
Calorimeter: Saturation & Leakage

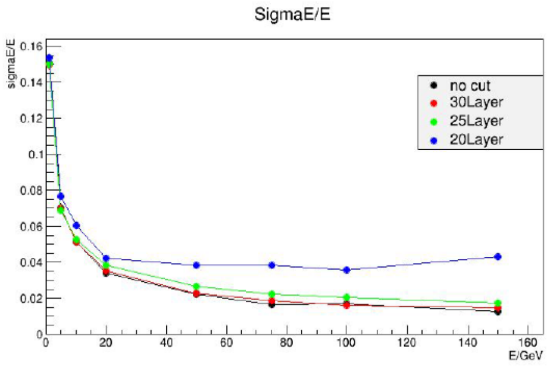
Calo optimization effort: UCAS(陈石)， IHEP(成栋，赵航，树正，学正， etc)

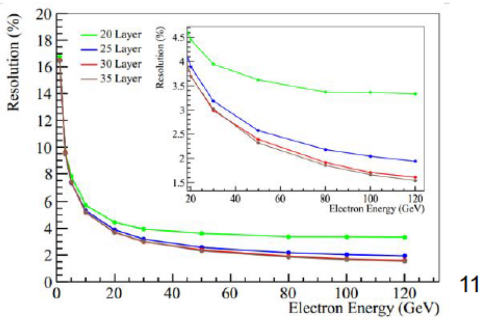


单层结构：2mm塑闪+2mmPCB+3mmW

入射粒子为gamma







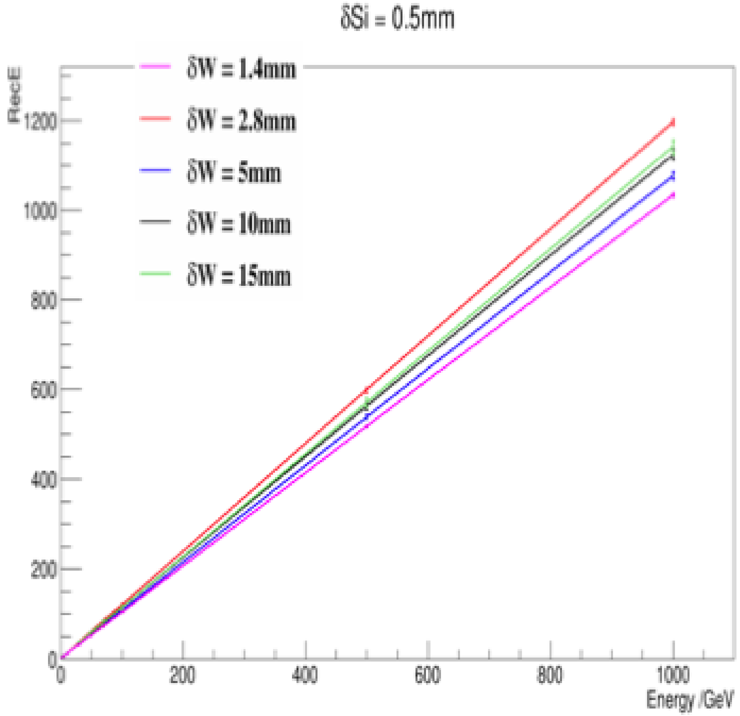
L 1sigma value = 0.87x - 0.24yy + 0.97y - 0.43z + 0.82

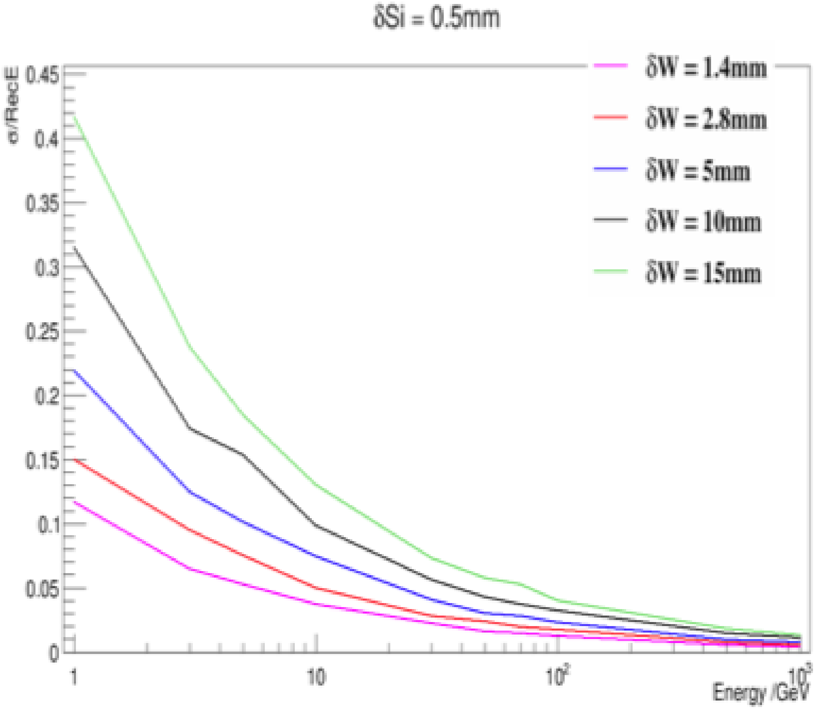
x = log10(energy) y = log10(cell size) z = log10(angle)

Eg, Saturate at 175 GeV photon, 20mm cell size: 2500 MIP

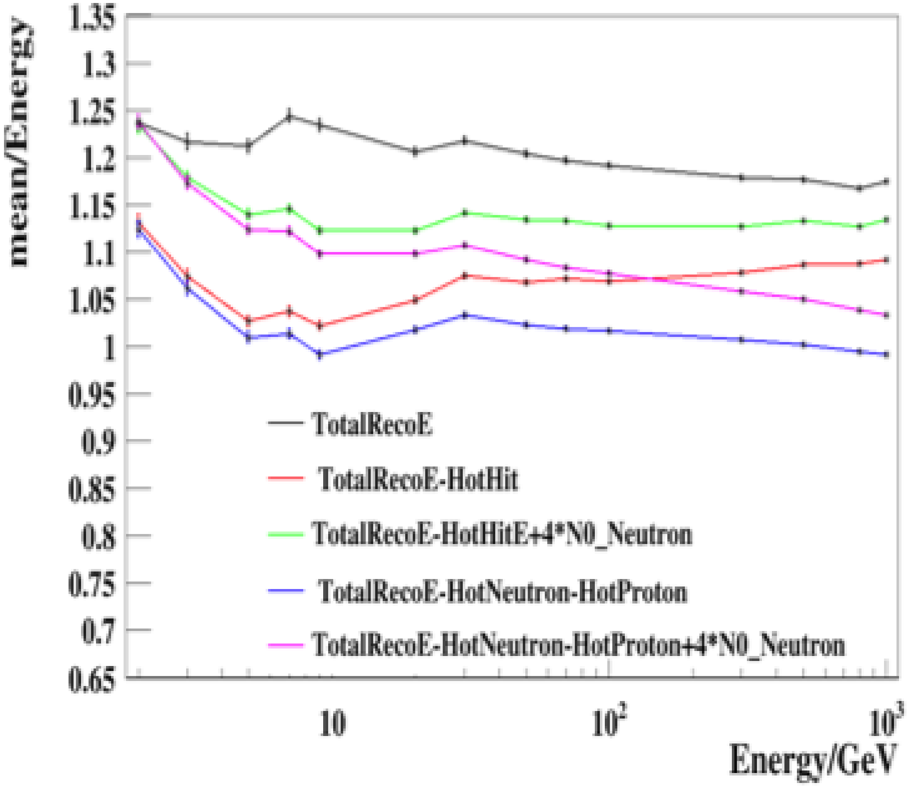
Calorimeter: Linearity & Resolution

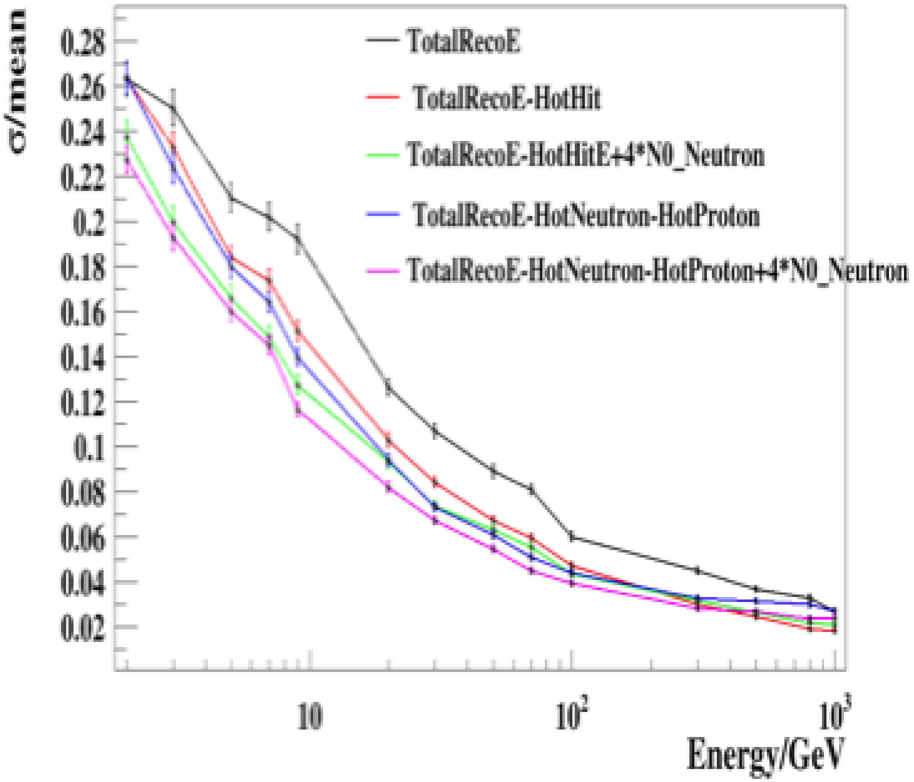
- 敏感层Si = 0.3， 0.5， 0.75 （单位：mm）
- 吸收层W = 1.4， 2.8， 5， 10， 15 （单位：mm）





Calorimeter: Energy Estimation of Hadronic object

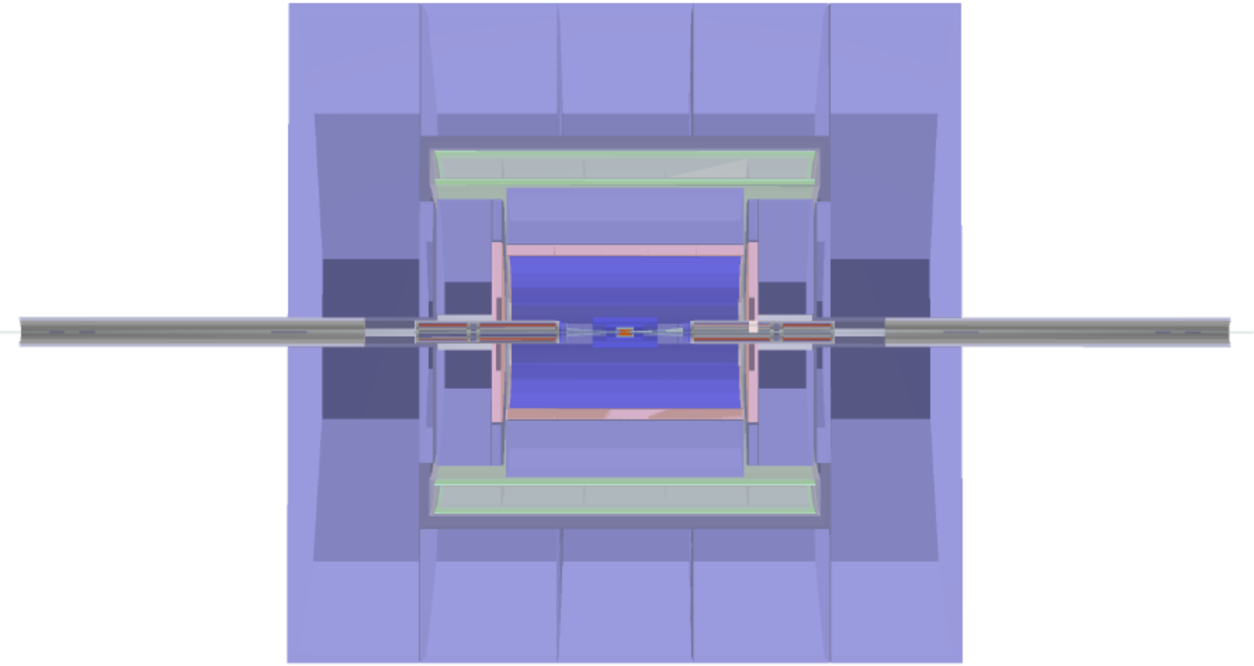


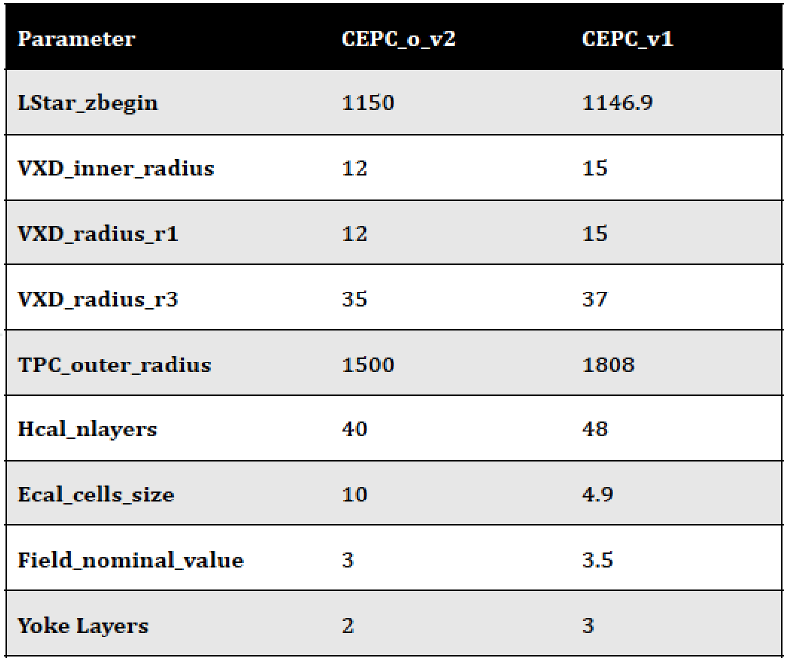


Default Energy Cut: 0.15 MIP, time < 10000 ns

Using Time, Energy & Nature information, Hadron energy measurement can be Dramatically improved... (up to 40%)

CEPC_o_v2





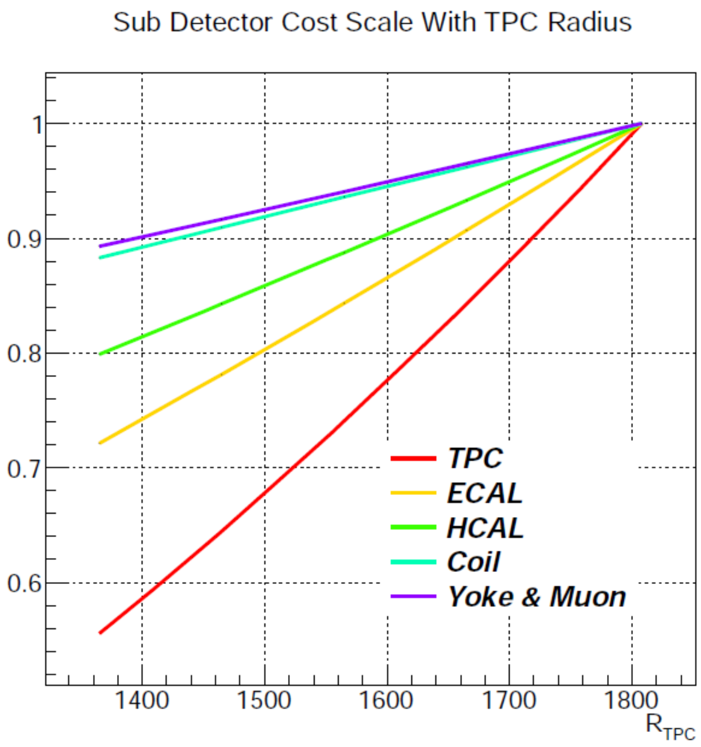
Parameter	**CEPC_o_v2**	**CEPC_v1**
LStar_zbegin	1150	1146.9
VXD_inner_radius	12	15
VXD_radius_r1	12	15
VXD_radius_r3	35	37
TPC_outer_radius	1500	1808
Hcal_nlayers	40	48
Ecal_cells_size	10	4.9
Field_nominal_value	3	3.5
Yoke Layers	2	3

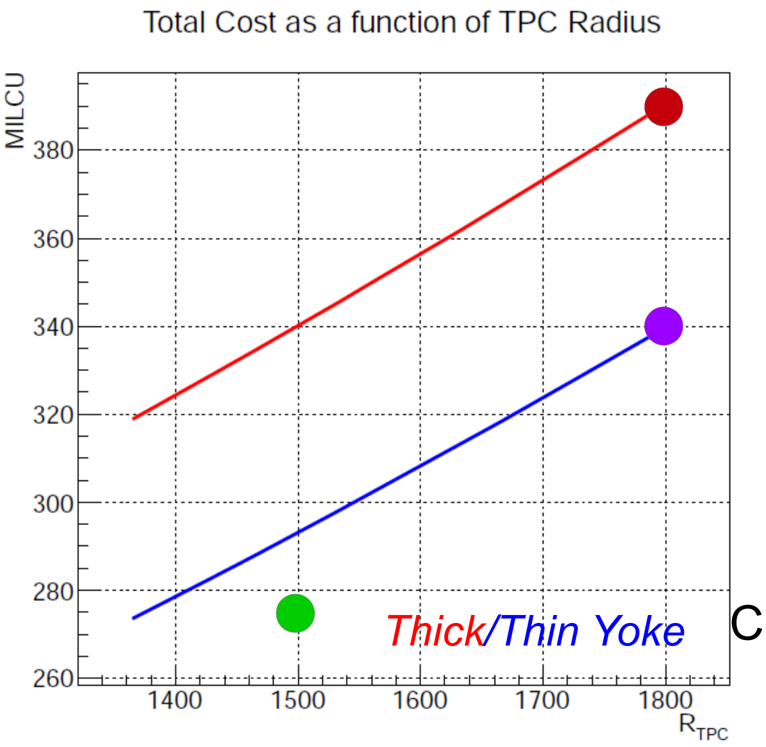
Parameter put by hand, motivated by:
- Saving the cost.
- Closer VTX inner layer, better flavor tagging?

Reconstruction: be aware of TPC boundary & B-Field Strength

Implemented by Xuyin (NanKai U)

Cost estimation: extrapolate from ILD

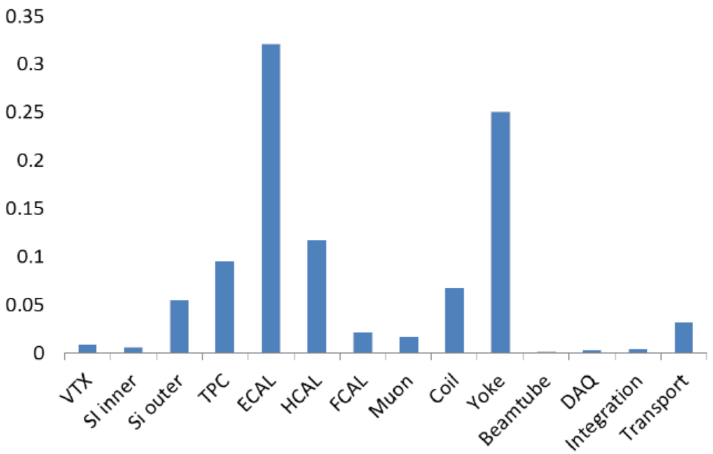




ILD

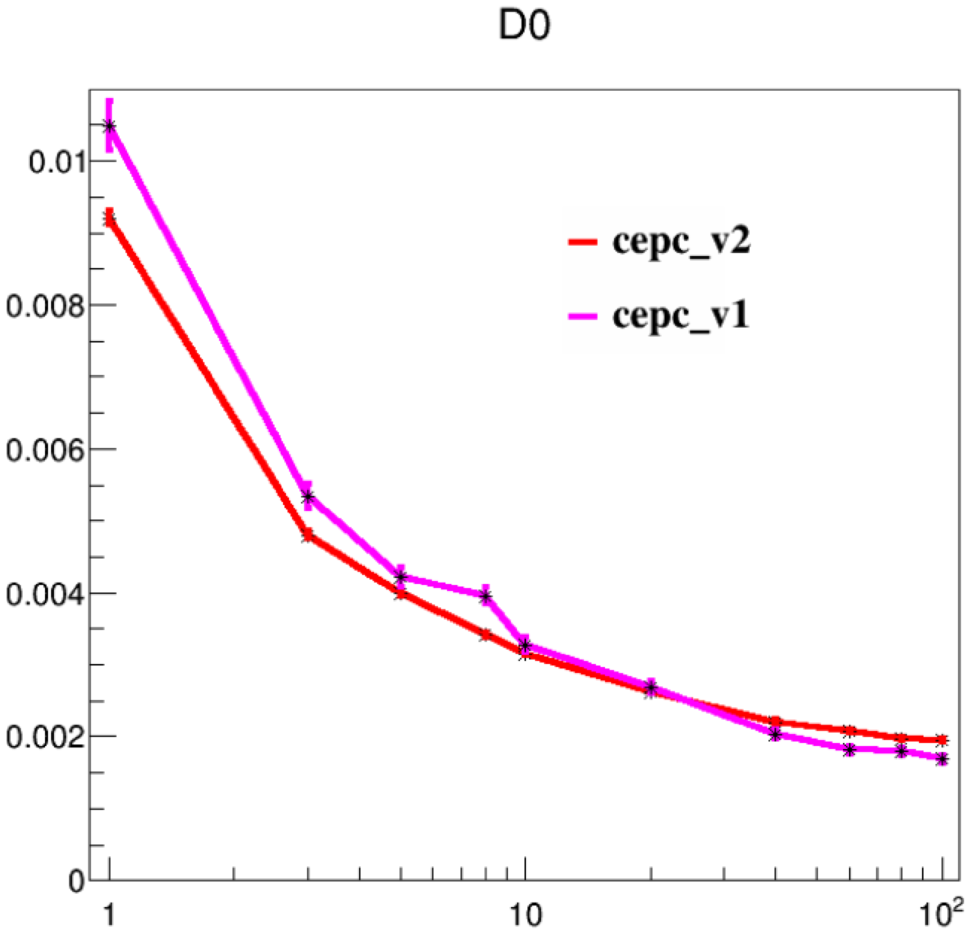
CEPC_v1

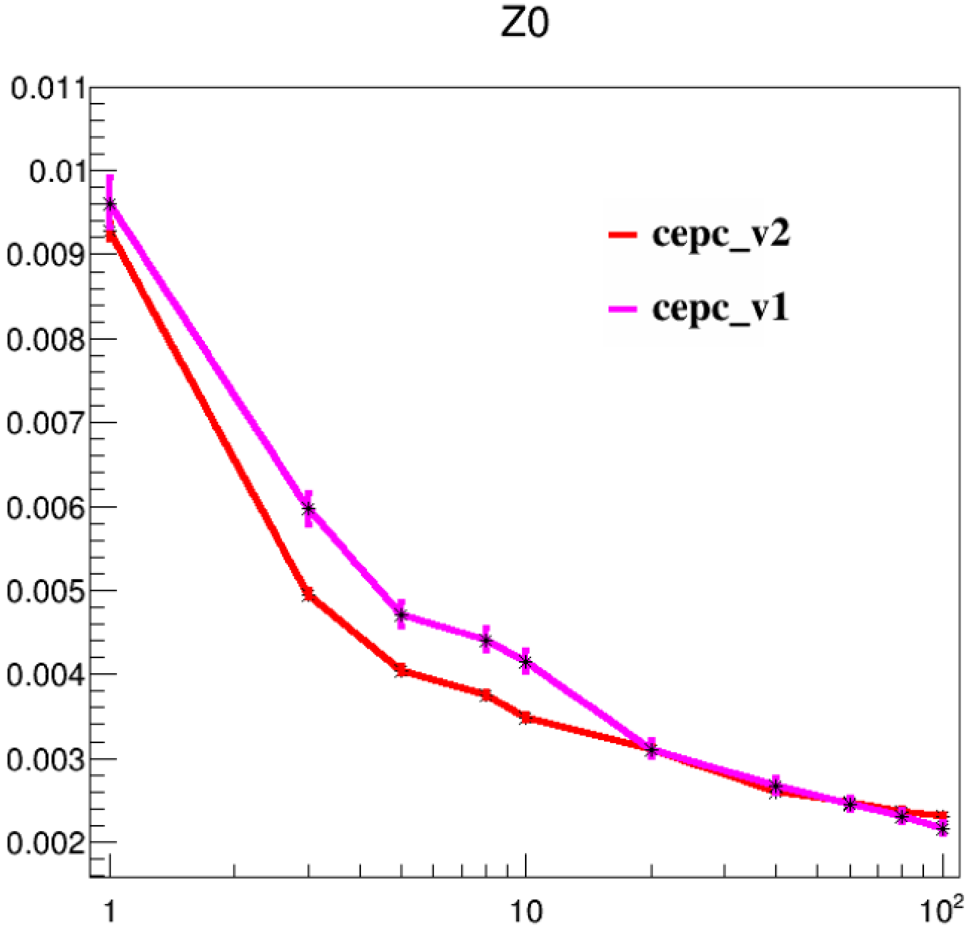
CEPC_o_v2



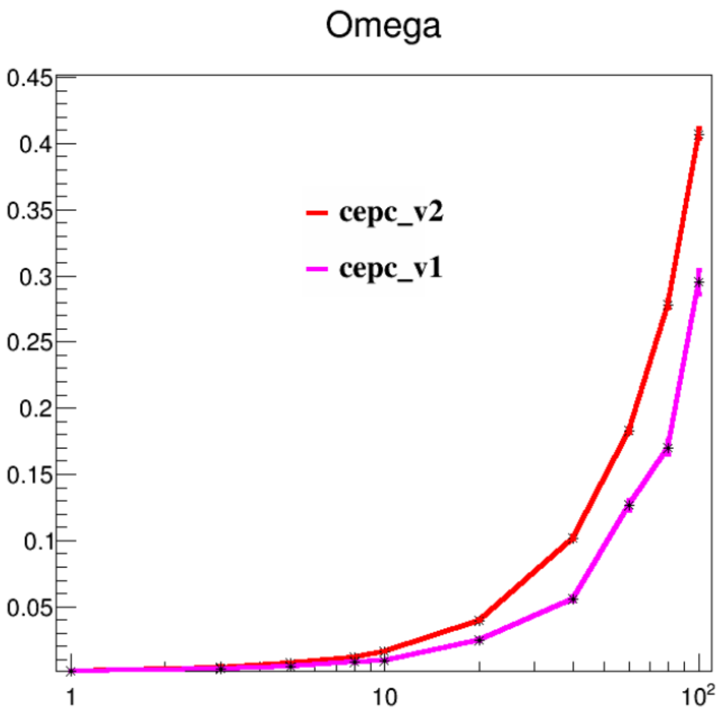
ILD -> CEPC_v1: reduced by 13%
CEPC_v1 -> CEPC_v2: reduced by further 25%

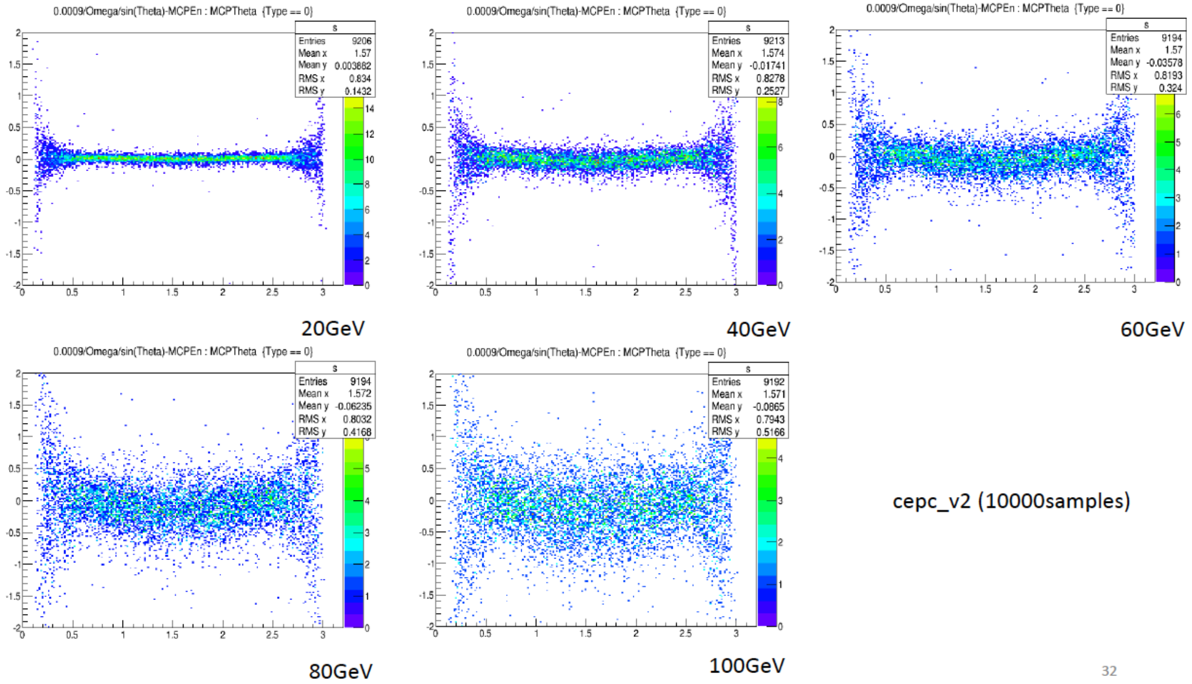
Performances: Tracks





Performances: Tracks





20GeV

40GeV

60GeV

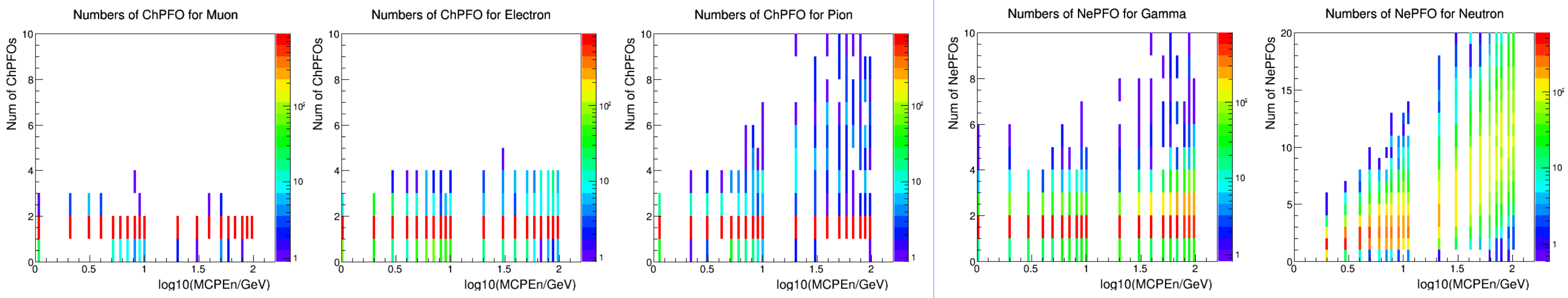
80GeV

100GeV

cepc_v2 (10000samples)

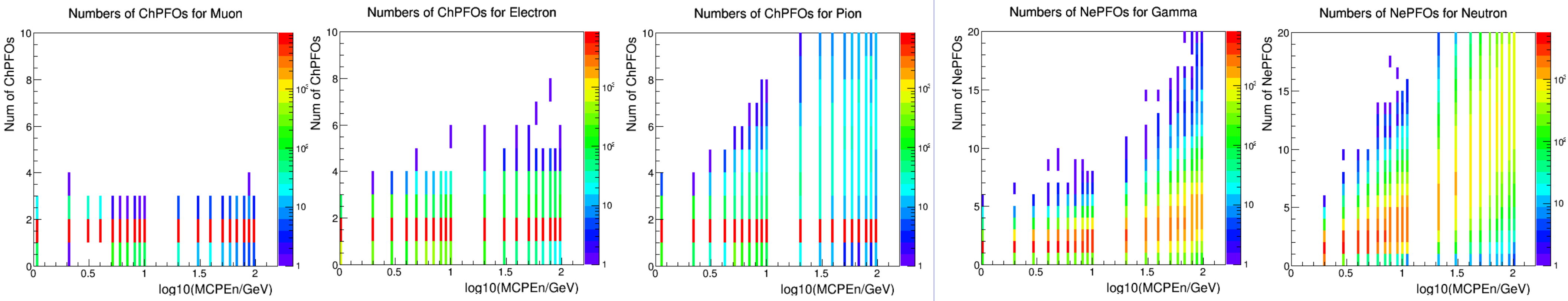
Banana Shape emerged at high energy... ~o(10^{-3}) level

Single particle: clustering



CEPC_v1

CEPC_o_v2

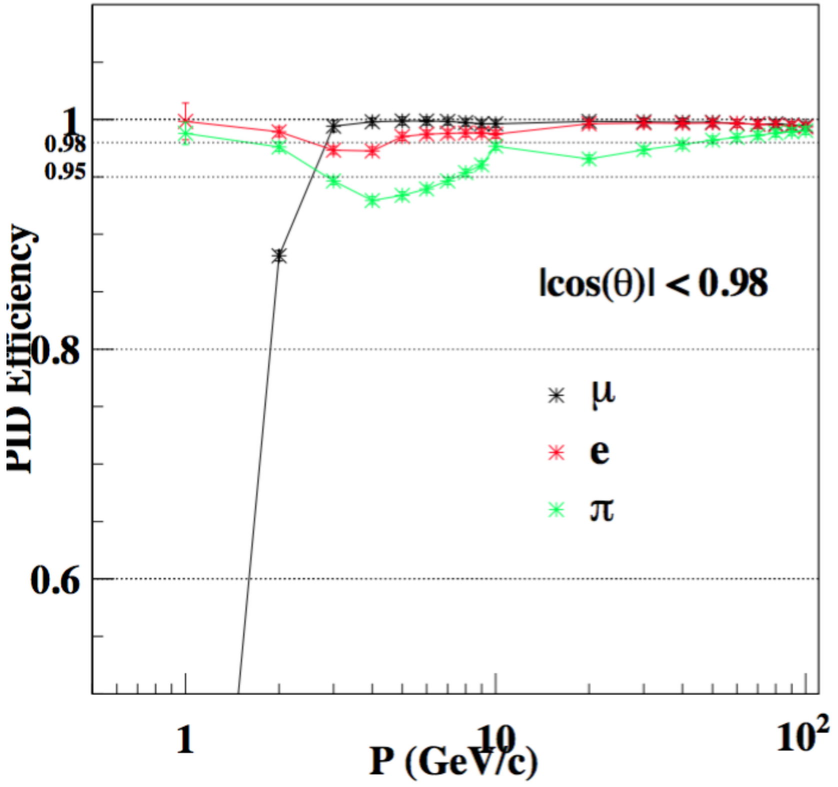


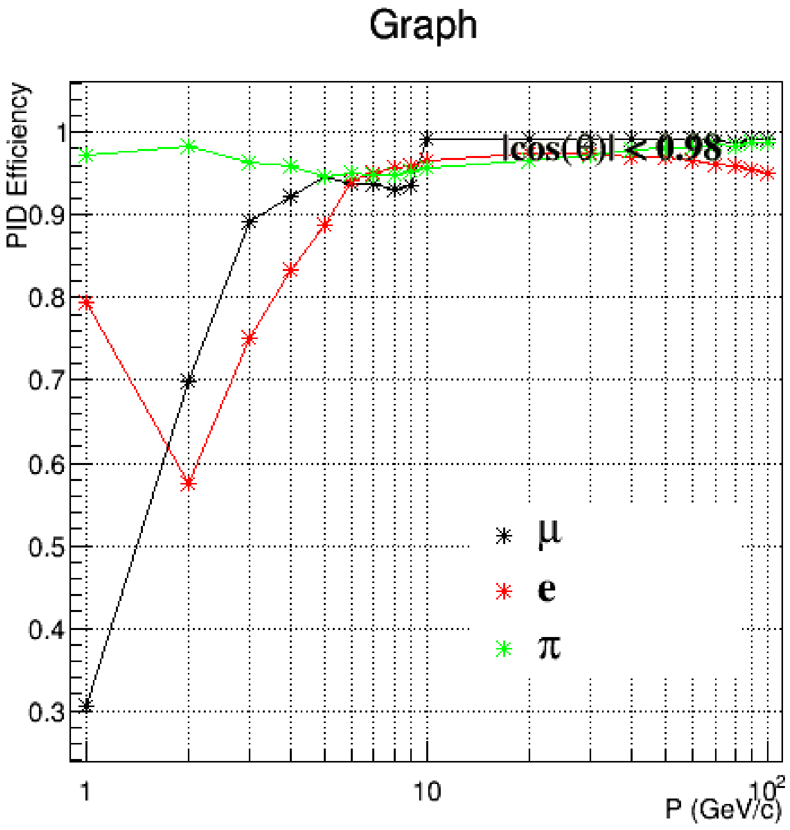
Charged

Neutral

Splitting is more severer in new geometry: need adjustment

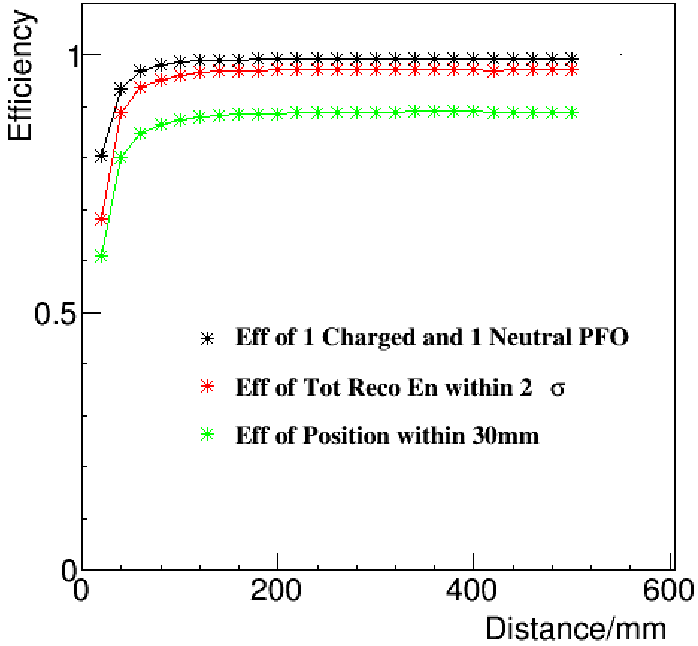
Performances: Lepton ID

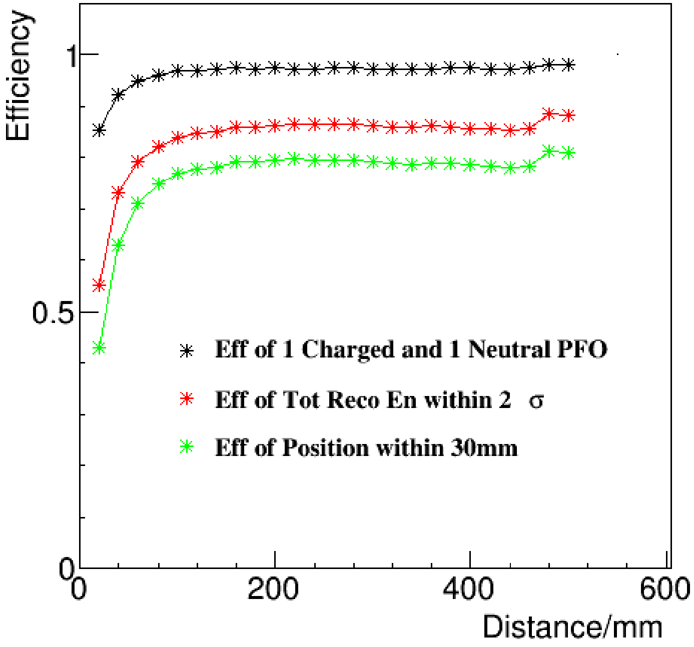


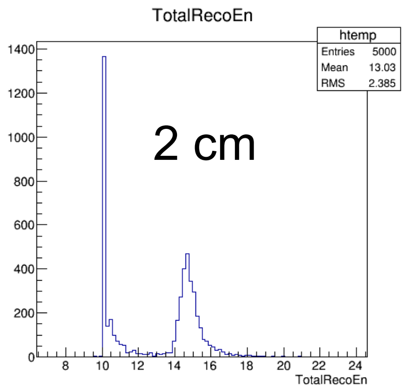


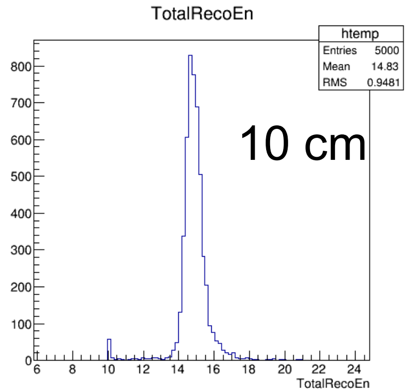
a bit crazy – mainly due to the parameter shift due to the Calo change...

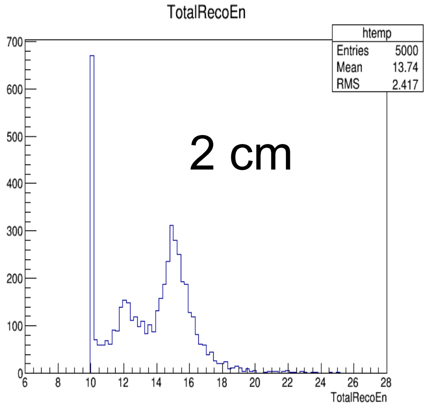
Performances: Overlap particles

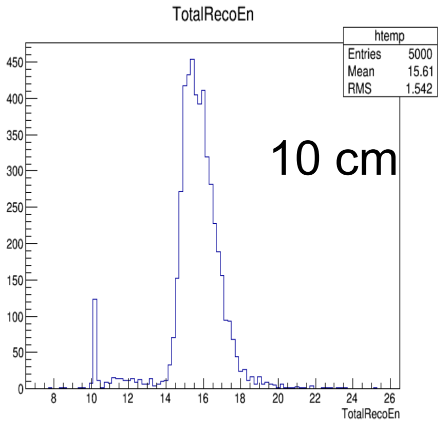






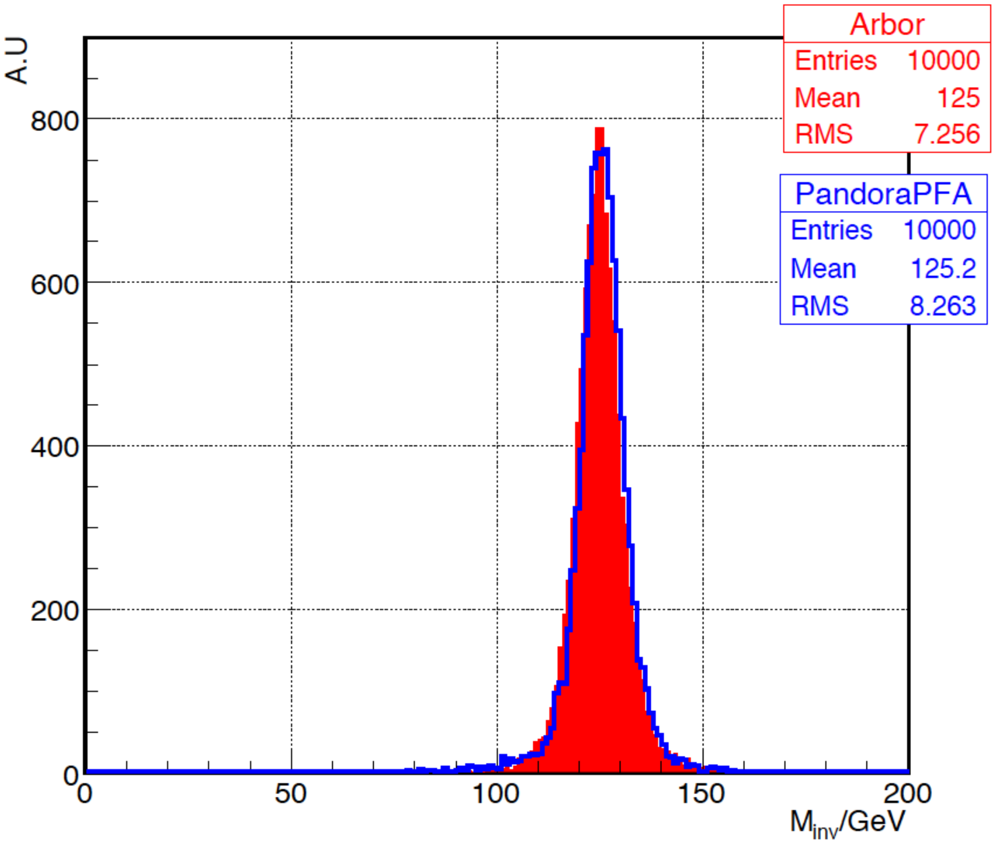


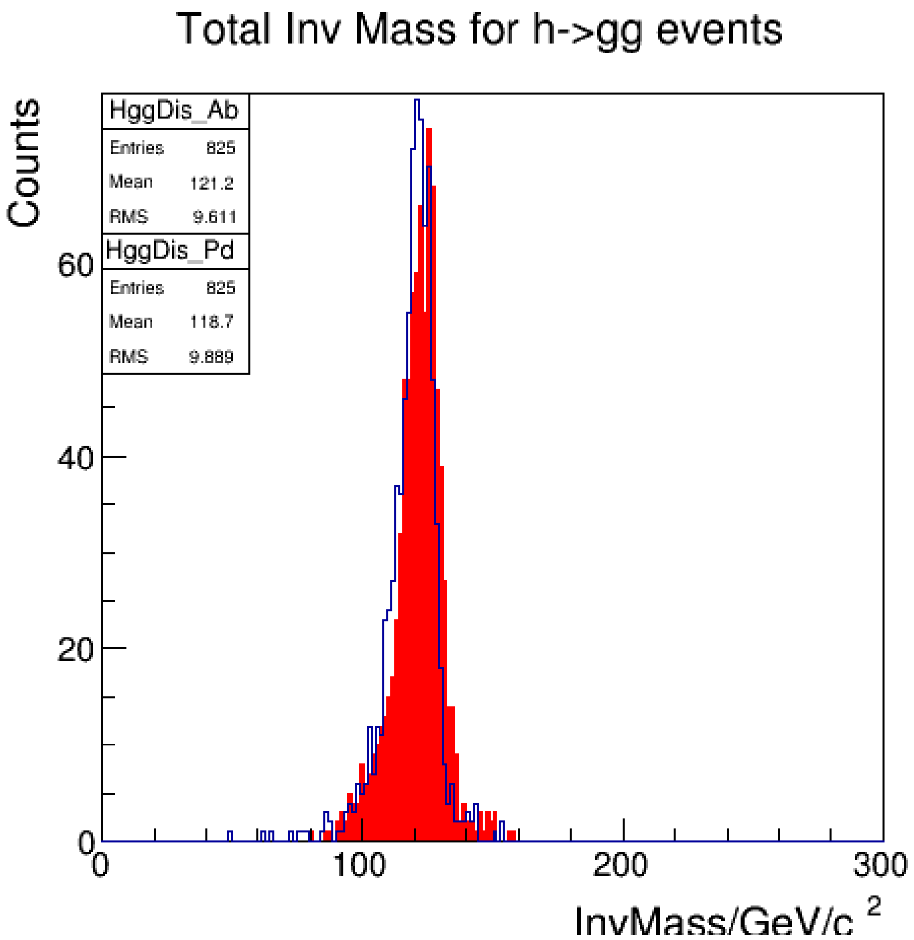




Anomaly in Matching observed

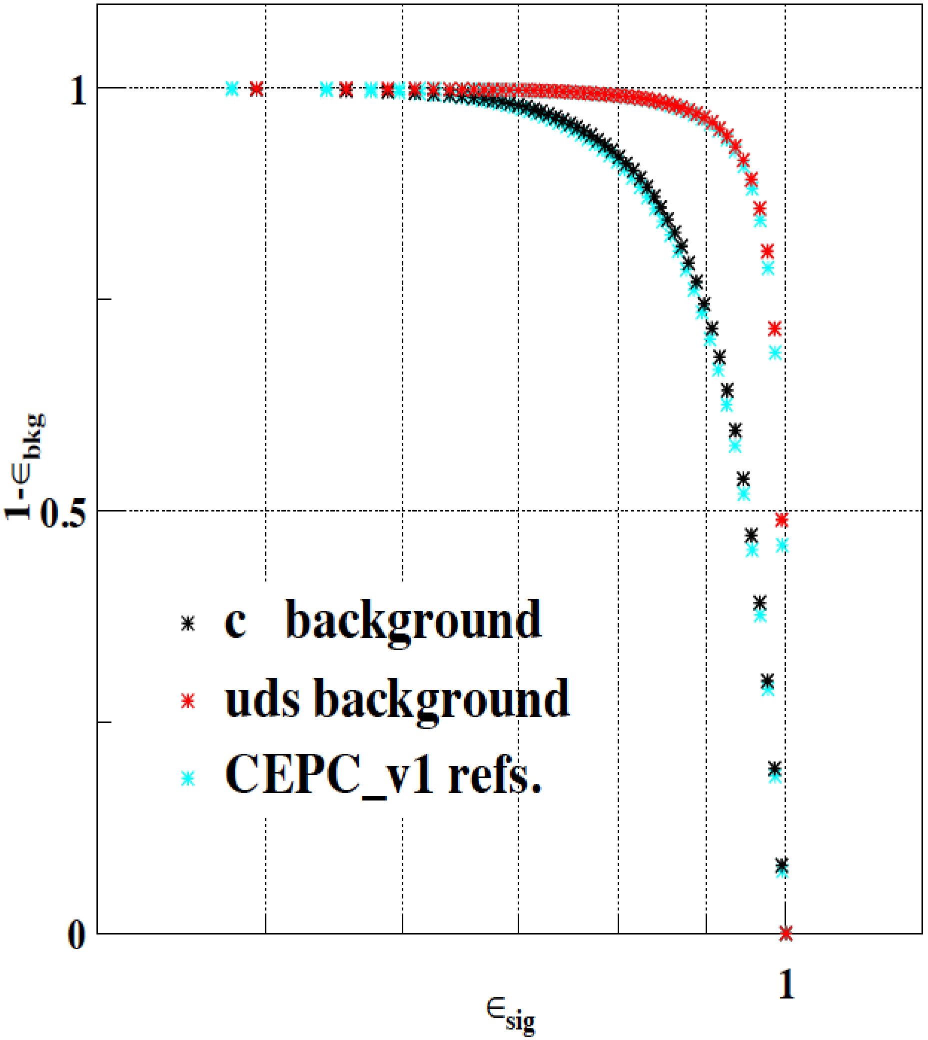
Performances: Jets

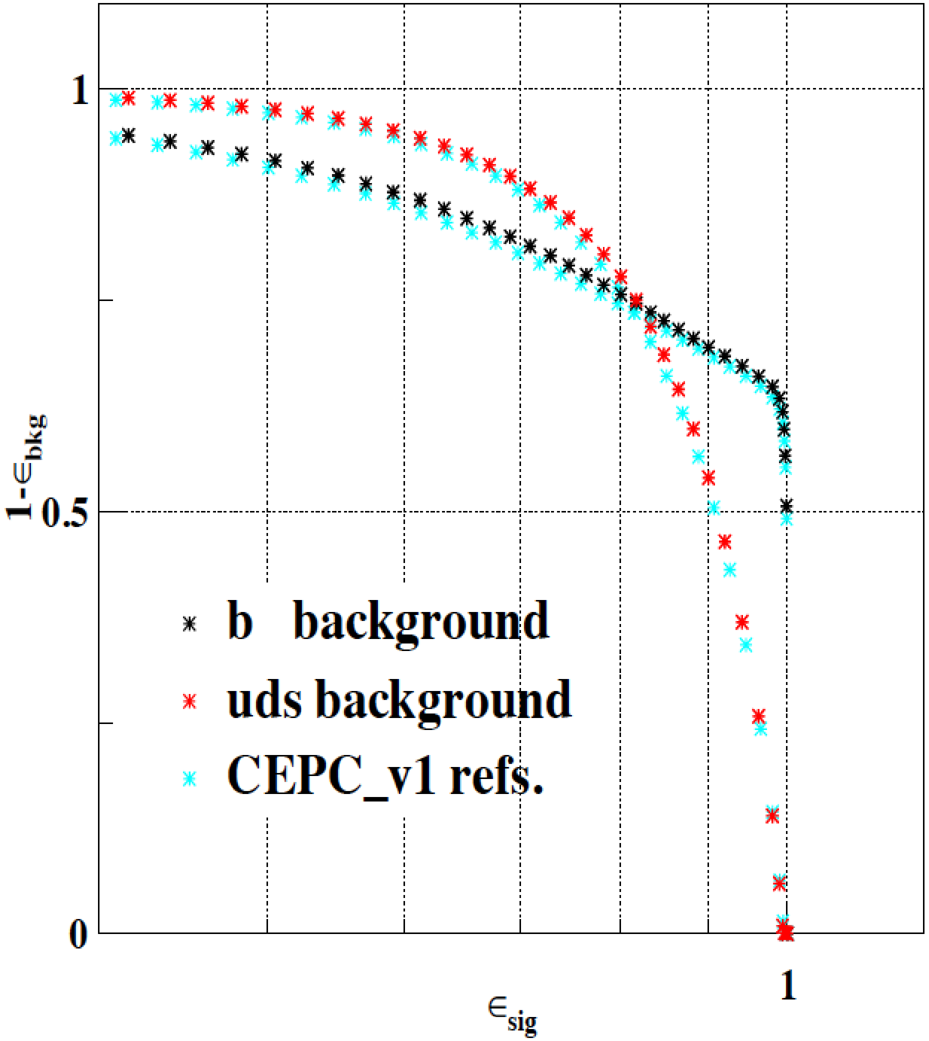




- *Without optimization... similar tendency observed at both Pandora & Arbor*

Performances: Flavor Tagging





CEPC_v1 → CEPC_o_v2

W.R.T CEPC_v1, Reduce:

Total cost ~ 25%;
ECAL power/FEE: 75%;
HCAL thickness/channels ~ 20%;
B-Field to 17% (3.5 → 3);
VTX inner radius: 25%;

Qualitatively: everything goes into the Expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for **OPTIMIZATION,** especially the PFA

Performance	adapted
Tracking: D0, Z0	20% ↑ @ E < 20 GeV (VTX); 5% ↓ @ E > 20 GeV (B-Field effect);
Lambda, Phi	worse
Omega	worse
PFA: Clustering	Slightly worse
Matching	worse
Separation	~10% ↓
PID	3-5% ↓ @ E > 10 GeV; 10% ↓ @ E < 10 GeV;
JER	20% ↓
Flavor Tagging	Improved up to 5%↑

CEPC_v1 → CEPC_o_v2

W.R.T CEPC_v1, Reduce:

Total cost ~ 25%;
ECAL power/FEE: 75%;
HCAL thickness/channels ~ 20%;
B-Field to 17% (3.5 → 3);
VTX inner radius: 25%;

Qualitatively: everything goes into the expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for **OPTIMIZATION,** especially the PFA

Performance	adapted	optimized*
Tracking: D0, Z0	10% ↑ @ E < 20 GeV (VTX); 5% ↓ @ E > 20 GeV (B-Field);	
Theta, Phi	worse	-
Omega	worse	-
PFA:Clustering	Slightly worse	same
Matching	~10% ↓	~5% ↓
Separation	~10% ↓	~2% ↓
PID	3-5% ↓ @ E > 10 GeV; 10% ↓ @ E < 10 GeV;	~1%↓
JER	20% ↓	~10%↓
Flavor Tagging	Improved up to 5%↑	?

**My personal expectation...*

CEPC_v1 → CEPC_o_v2

W.R.T CEPC_v1, Reduce:

Total cost ~ 25%;
ECAL power/FEE: 75%;
HCAL thickness/channels ~ 20%;
B-Field to 17% (3.5 → 3);
VTX inner radius: 25%;

Qualitatively: everything goes into the expected direction

Quantitatively: ???

Reconstruction: Adapted, lots of effects needed for **OPTIMIZATION,** especially the PFA

Performance	adapted	optimized*	Manpower/people*month
Tracking: D0, Z0	20% ↑ @ E < 20 GeV (VTX); 5% ↓ @ E > 20 GeV (B-Field);		4
Theta, Phi	worse	-	
Omega	worse	-	
PFA:Clustering	Slightly worse	same	-
Matching	~10% ↓	~5% ↓	6
Separation	~10% ↓	~2% ↓	2
PID	3-5% ↓ @ E > 10 GeV; 10% ↓ @ E < 10 GeV;	~1%↓	4
JER	20% ↓	~10%↓	4
Flavor Tagging	Improved up to 5%↑	?	3

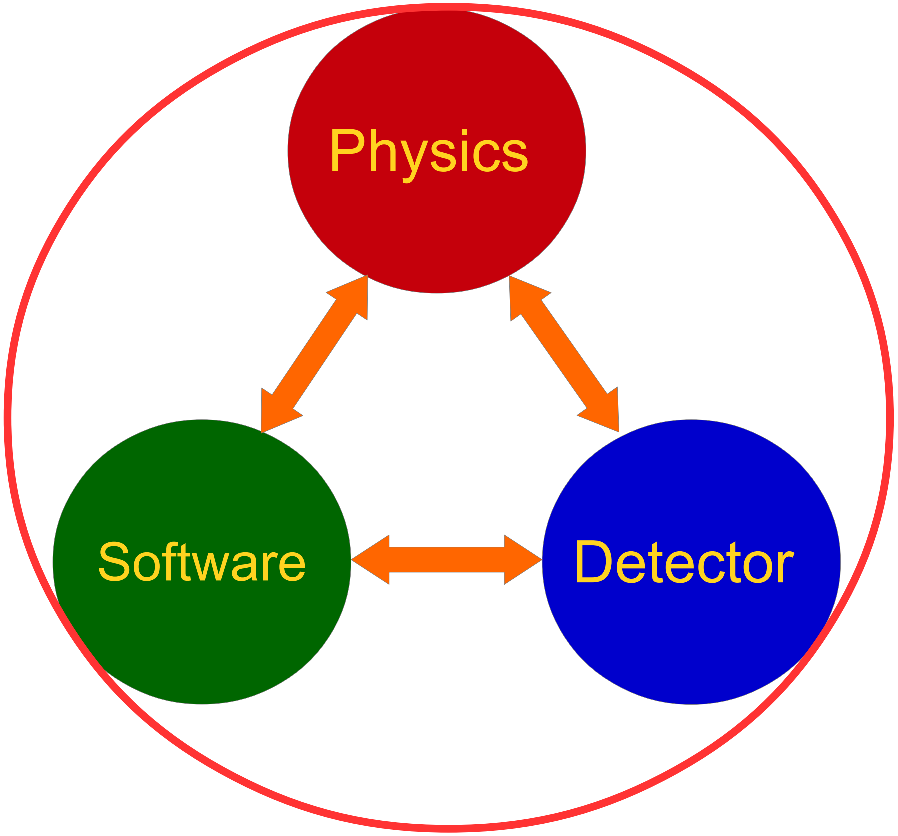
*Given the current status, 23 people*month is needed to fully optimize the performance.*
For the next geometry, the needed manpower will be half

Activities & Cooperations

- Physics, optimization & PFA working group...
- Supported by IHEP innovation funding...
  - Short term visit;
  - Training;
  - Physics Software workshop(s);
  - Computing resource;

- Cooperations:
  - Chicago University & Wuhan University: Photon reconstruction & H->di photon analysis
  - SLAC: Irradiation studies, etc
  - LLR, France: PFA & Tau finder
  - Shandong University, ZhongShan University: Software framework
  - Nankai University: simulation toolkit
  - ...

Perspective at CDR



- Benchmark detector(s)
  - **Optimized PFA detector**
    - Scan key para (size; B-field)
    - Compare different technology: develop adequate digitizer, validate at prototype test...
    - Understand the MDI constrain
  - Other concepts/option?
- Software
  - Adopt & **Optimize** to Benchmark detector
  - Simulation: more realistic...
  - Framework & computing: prototype
  - Software team: regular release & validation (need at least 2 professional software experts...)
- Physics
  - Higgs, EW & BSM @ Benchmark Geometry
  - Scan Benchmark Physics analyses at different geometry
- Decent Documentation & Publication

5-year perspectives & resource demand

- Perspective:
  - Physics: understand the physics requirement to CEPC detector, and demonstrate the physics potential from Higgs, EW & BSM
  - Software: develop, maintain and optimize the full set of SCRAC tools, develop future software framework and computing tools
  - Detector Design **GLOBAL** optimization: taken into account the physics goal, constrains (collision environments, detector hardware technologies), detector performance and cost
- Resources:
  - Computing & Storage: 1 PB storage & 1k CPU for Higgs Run.
    - 1 M/evt; 1 evt/CPU*min;
    - Higgs Run: o(10^9) Physics events: 1M Higgs; Z pole: o(10^{10}) Physics events
  - Manpower: 12 FTE 、 9 PostDoc 、 30 Students
    - Analysis: 2 FTE + 2 PostDoc + 14 Students
    - Software:
      - **SCRAC**: 5 FTE + 3 PostDoc + 6 Students;
      - Framework & Computing: 3 FTE + 2 PostDoc + 4 Students
    - **Optimization**:
      - 2 FTE + 2 PostDoc + 8 Students

Personally I hope we also have outreach experts...

Summary

- Status of SCRAC: healthy & lots of progress
  - Simu: mastered existing tool, freely edit geometry
  - Tracking & Flavor Tagging: mastered the ILC tools
  - PFA reconstruction: leptons, photons, taus, Jets are reconstructed at high efficiency & accuracies...
- Optimization:
  - Explored at many different P.o.V
  - **The REAL game: SCRAC**@CEPC_o_v2, adaption is straight forward, optimization demands lots of manpower & expertise
- *Lots of Activities & Global cooperations*
- *Toward CDR: optimized SCRAC & benchmark geometr(ies)*
  - *One Benchmark geometry: PFA detector. Optimization is secured*
  - *Open to other concepts...*
- *In 5 years, personal vision:*
  - ***Global optimization**: to achieve the physics goal in a feasible & most efficient way – Joint efforts between Theory-Pheno study, Simulation, Detector hardware development and Accelerator Study...*
  - *Strong team: Software Frame & SCRAC, Analysis...*

Thank you!

- *Special thanks to Gang, Xuyin, Chengdong, Libo, Binsong, Dan, and the full analysis team – who not only produce the physics results, but also valid/polishes the reconstruction tools!*
- *Apologize for being not able to cover the discussion & progress made in Software framework & computing*

Back up

Software framework consideration

Use an existing one 🙂 vs Develop from beginning 😢

- Consideration of the choice for CEPC
  - Enough services and functionalities
  - Easy to use
  - Future supports
- Almost all widely used frameworks can satisfy our requirements
- Several potential candidates are investigated and compared

Framework candidates investigation

- Marlin: currently used by CEPC(with uncertain official support)
- Gaudi: very popular for collider physics experiments, most familiar to us, very comprehensive but a bit heavy
- ROOT: very flexible and powerful, but need more manpower for some service functionalities development
- ART: optimized for high intensity physics experiments and a little complex
- SNiper: lightweight and optimized for non-collider experiments

	Marlin	Gaudi	ROOT	ART	SNiPER
User Interface	XML	Python, TXT	Root script	FHiCL	Python
Adoption	ILC	Atlas, BES3, DYB	Phenix, Alice	Mu2e, NOvA, LArSoft, LBNF	JUNO, LHAASO

Computing considerations

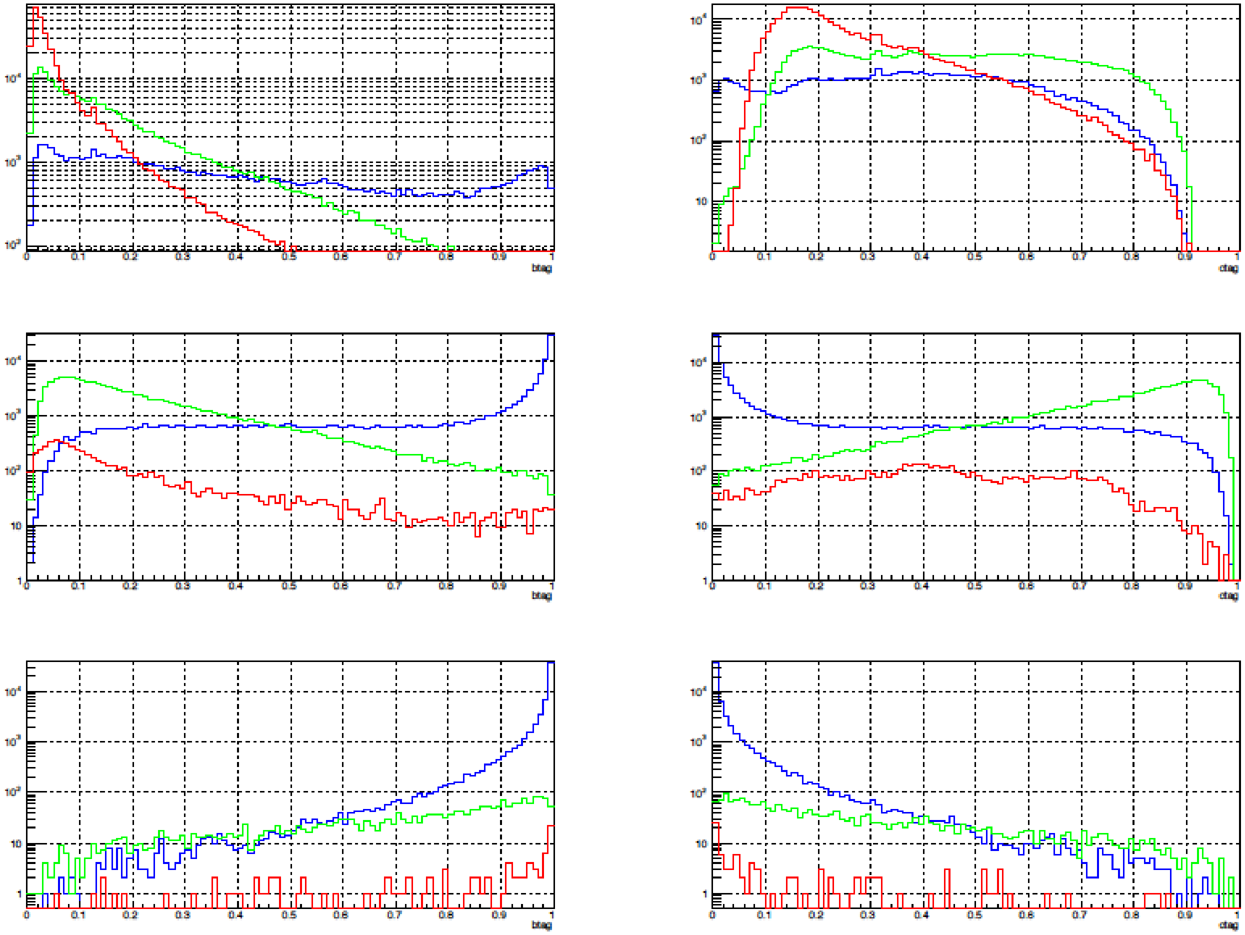
- It is still far to confirm the computing technology now used for 30 years more
- But we believe the technology is evolving step by step
- Now the main computing task is to study and follow the latest computing technology to prepare for the future, including
  - Cloud computing
  - Distributed computing
  - Multi-cores computing
  - High performance computing
  - Unified distributed data management and access
  - "Smart" network, high bandwidth future network
  - …….

Step by step: performance

Nvtx=0

Nvtx=1

Nvtx>1



bbar

ccbar

Interesting/crucial topics to explore

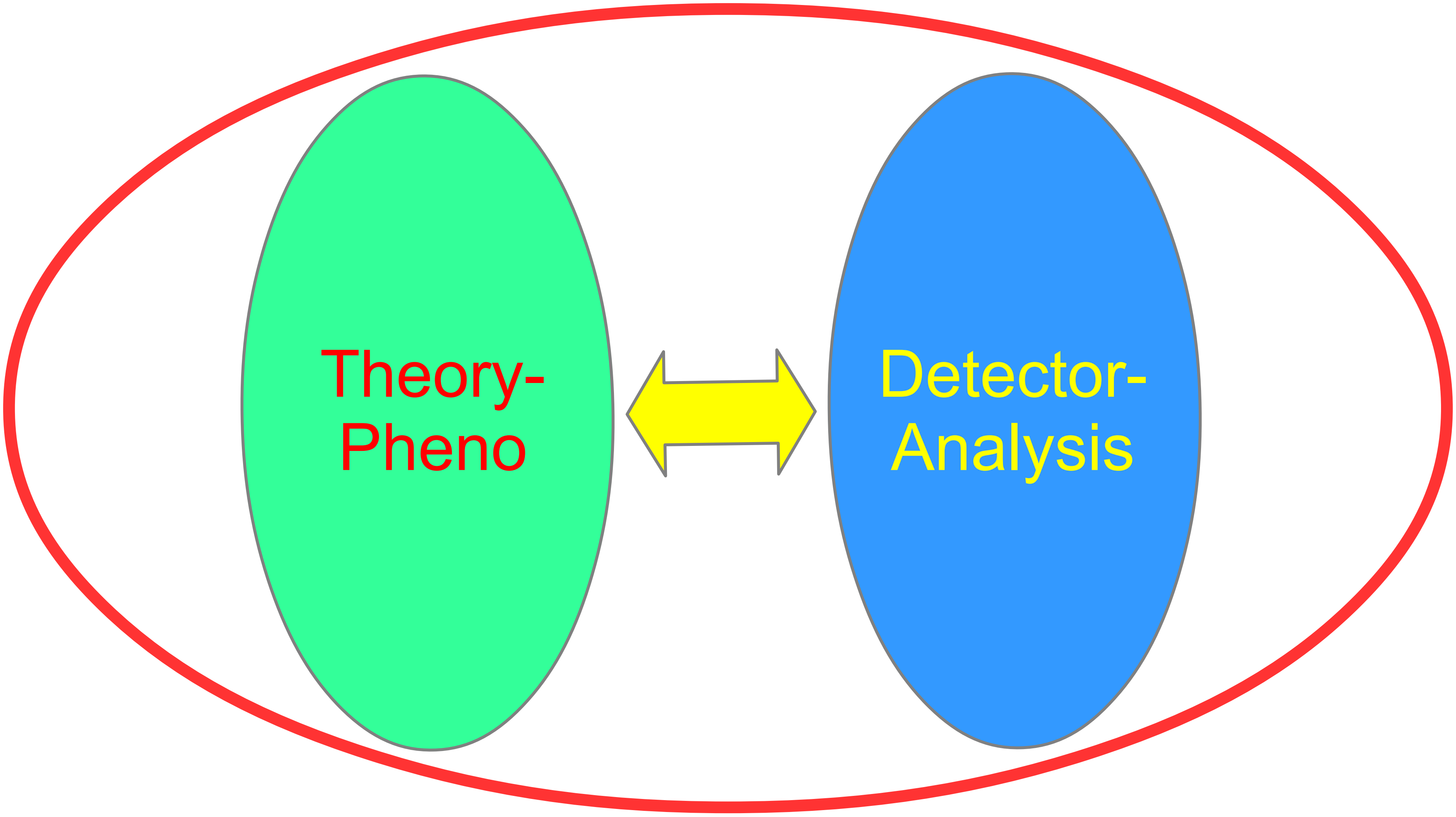
- Jet Clustering
  - Distinguish 2 jets, 4 jets and 6 jets events from each other;
  - Identify the boson (color singlet) origin of different jets;
- Analysis
  - sigma(ZH) determination from qqH recoil
  - Data driven method to determine the Higgs observables
  - Systematic estimation for Higgs/Z pole runs

Reconstruction

Algorithms	Objects
Tracking	Sub-detector Tracks
	Merged Tracks
PFA: Clustering	Calorimeter Clusters
Matching	Final State Particles
PID	Final State Particles with Type flag/information
PFA: Advanced Object finder	Isolated Lepton
	Converted Photon
	Electron with BS photon recovery
	Tau
	MET
Jet Clustering	Jets
Flavor Tagging	Jets with b, c, uds, (gluon), multiple flags...

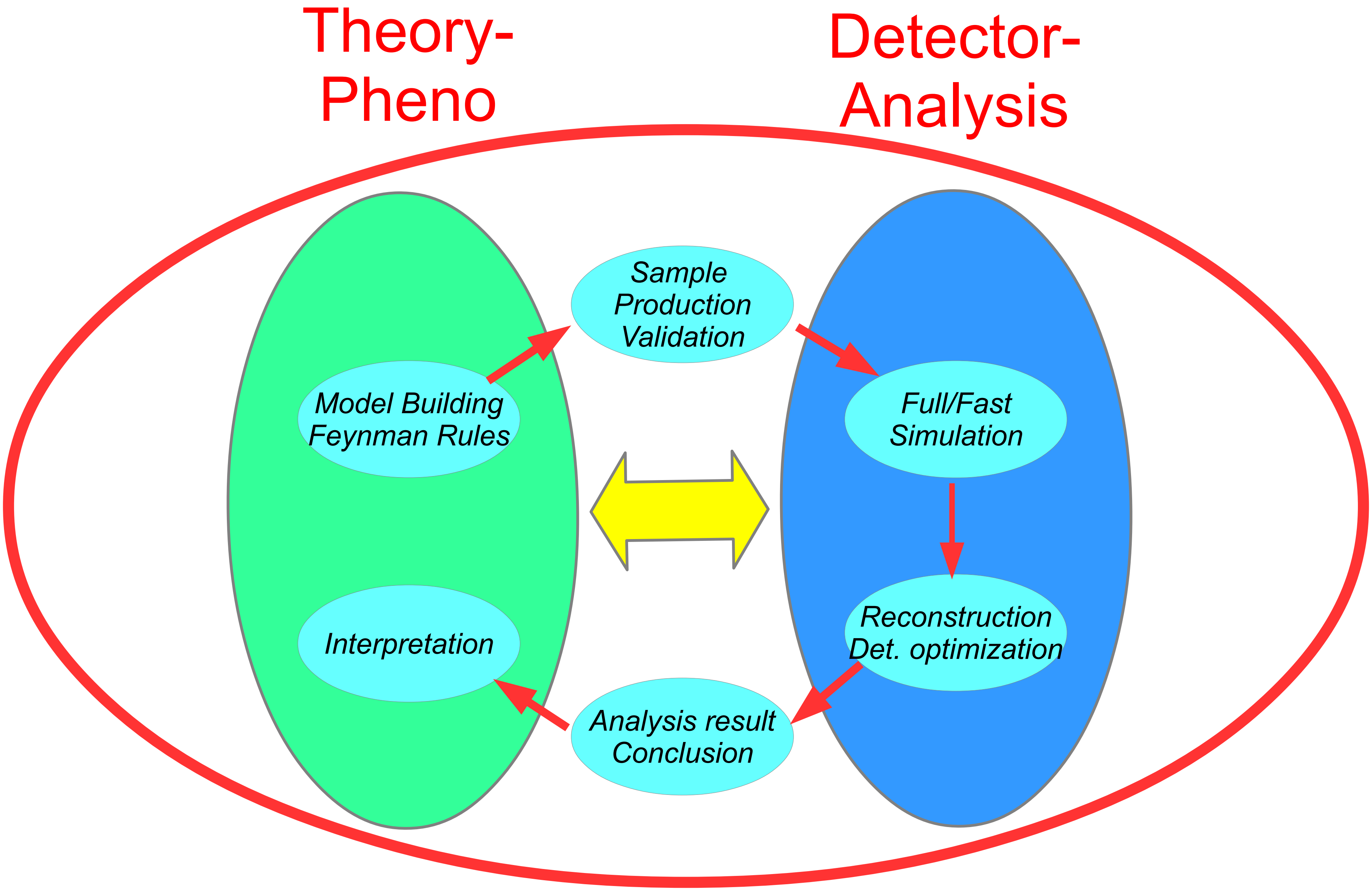
Detector optimization: general receipt

- Understand the motivation and carefully modify/edit the geometry accordingly
- Verify the detector geometry
    - Hit map
    - Object construction at sub-detector level: tracks , clusters & vertex
- **Adjust/optimize the reconstruction** & Understand the detector performance
    - Single particle level: reconstruction/id efficiency
    - Overlap particle level: separation performance, essential for PFA
    - Multi-particle object: Tau & Jets
    - *Tech. oriented, Time consuming & need strong expert (see Gang's talk)*
- Re-process the benchmark physics analysis



Theory-Pheno

Detector-Analysis



Theory-Pheno

Detector-Analysis

Model Building
Feynman Rules

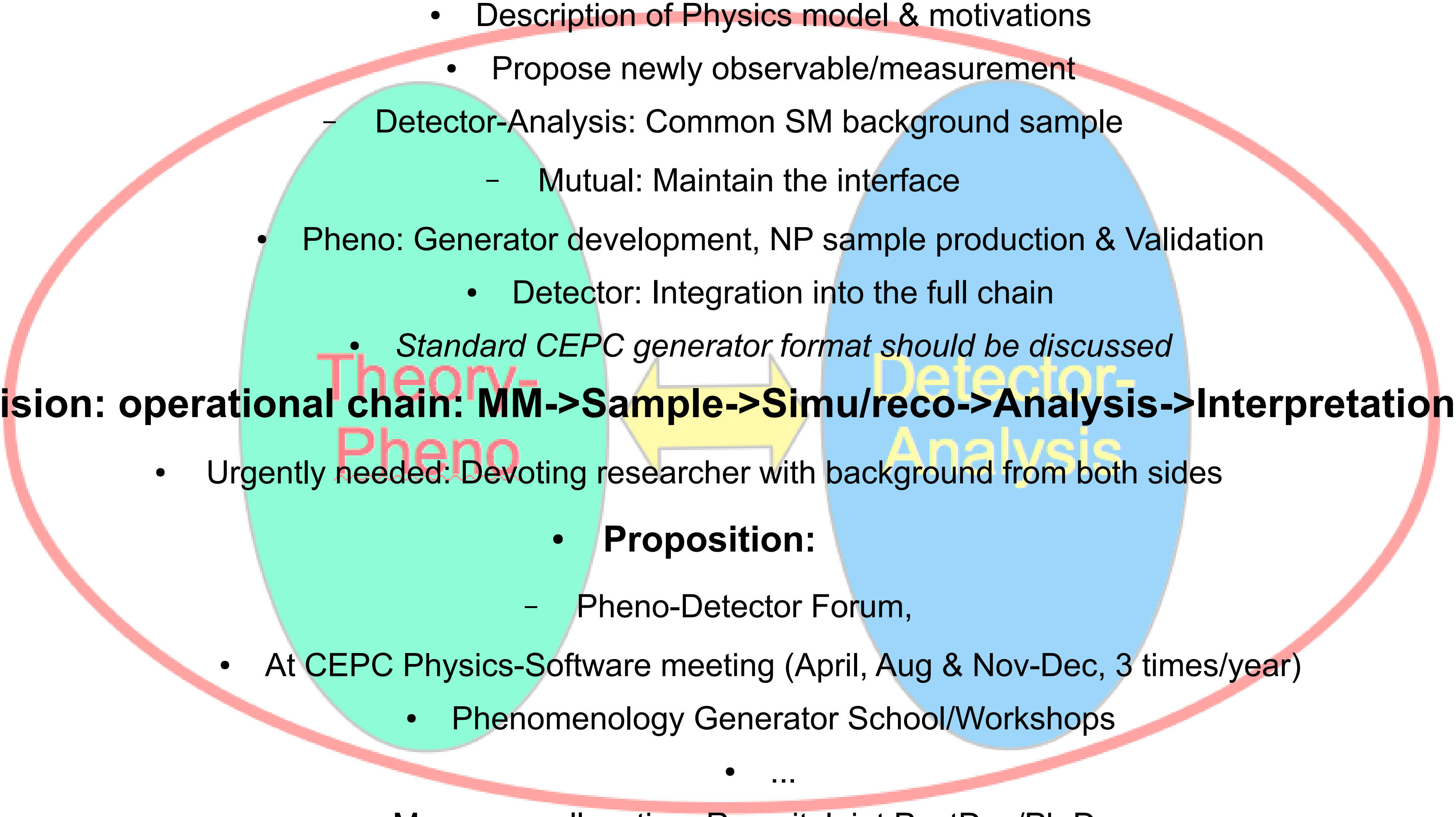
Sample
Production
Validation

Full/Fast
Simulation

Reconstruction
Det. optimization

Analysis result
Conclusion

Interpretation



- **Team work...**
  - Theory-Phenology: Model Building - Interpretation
    - Description of Physics model & motivations
    - Propose newly observable/measurement
  - Detector-Analysis: Common SM background sample
  - Mutual: Maintain the interface
    - Pheno: Generator development, NP sample production & Validation
    - Detector: Integration into the full chain
    - *Standard CEPC generator format should be discussed*
- **Vision: operational chain: MM->Sample->Simu/reco->Analysis->Interpretation**
  - Urgently needed: Devoting researcher with background from both sides
- **Proposition:**
  - Pheno-Detector Forum,
    - At CEPC Physics-Software meeting (April, Aug & Nov-Dec, 3 times/year)
    - Phenomenology Generator School/Workshops
    - ...
  - Manpower allocation: Recruit Joint PostDoc/Ph.D
  - Support relevant works: short term visit, travel, etc