STCF Offline Software Status

黄性涛, 艾小聪

代表STCF软件工作组

2026超级陶粲装置研讨会, 西安, 2026年7月3日

STCF physics requirements

Requirements of offline software at STCF

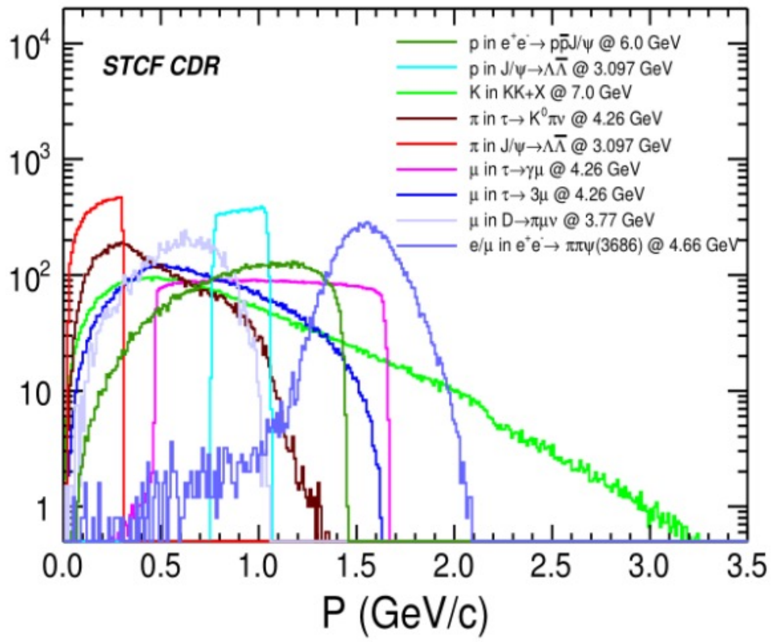
- Higher event rate, background hits ratio (~2/3 hits are from backgrounds), CPU consumption at STCF than BESIII. we need to
  - develop **highly accurate and efficient** detector simulation and digitization algorithms to meet the demands of high luminosity, high background for **different detector options**
  - **reconstruct** the tracks and photon **with good efficiency and resolution,** and **identify** them **at high accuracy**, with **good speed**

Process	Physics Interest	Optimized Subdetector	Requirements
$\tau \to K_s \pi \nu_\tau$, $J/\psi \to \Lambda\bar{\Lambda}$, $D_{(s)}$ tag	CPV in the τ sector, CPV in the hyperon sector, Charm physics	ITK+MDC	acceptance: 93% of 4π; trk. effi.: $> 99\%$ at $p_T > 0.3$ GeV/c; $> 90\%$ at $p_T = 0.1$ GeV/c $\sigma_p/p = 0.5\%$, $\sigma_{\gamma\phi} = 130\ \mu$m at 1 GeV/c
$e^+e^- \to KK + X$, $D_{(s)}$ decays	Fragmentation function, CKM matrix, LQCD etc.	PID	π/K and K/π misidentification rate $< 2\%$ PID efficiency of hadrons $> 97\%$ at $p < 2$ GeV/c
$\tau \to \mu\mu\mu$, $\tau \to \gamma\mu$, $D_s \to \mu\nu$	cLFV decay of τ, CKM matrix, LQCD etc.	PID+MUD	μ/π suppression power over 30 at $p < 2$ GeV/c, μ efficiency over 95% at $p = 1$ GeV/c
$\tau \to \gamma\mu$, $\psi(3686) \to \gamma\eta(2S)$	cLFV decay of τ, Charmonium transition	EMC	$\sigma_E/E \approx 2.5\%$ at $E = 1$ GeV $\sigma_{\text{pos}} \approx 5$ mm at $E = 1$ GeV
$e^+e^- \to n\bar{n}$, $D_0 \to K_L \pi^+\pi^-$	Nucleon structure Unity of CKM triangle	EMC+MUD	$\sigma_T = \frac{300}{\sqrt{p^3(\text{GeV}^3)}}$ ps

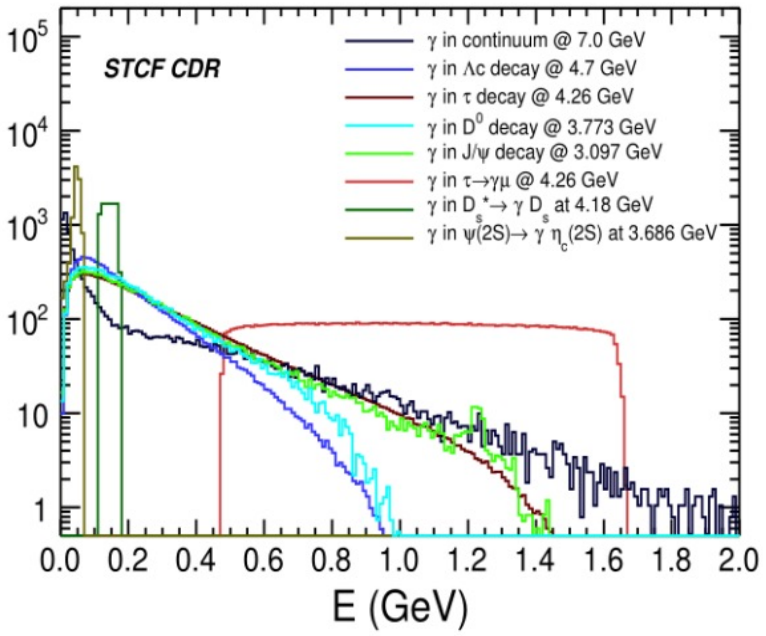
Particles at STCF

- Charged particles
  - **e, μ, K, π, proton** (most have p < 2 GeV, lots have p < 400 MeV)
- Neutral particles
  - **γ** (energy coverage: 25 MeV - 3.5 GeV) and **K_L, neutron** (up to 1.6 GeV)

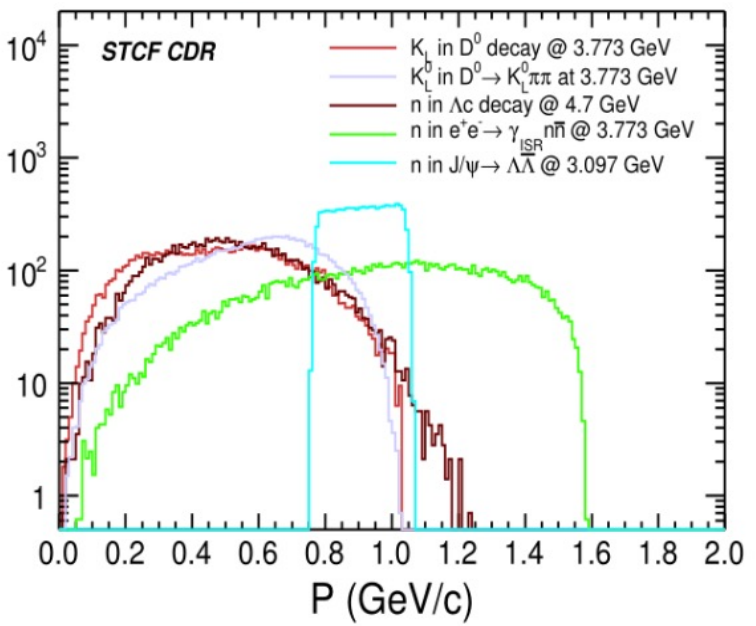
Charged particle momentum



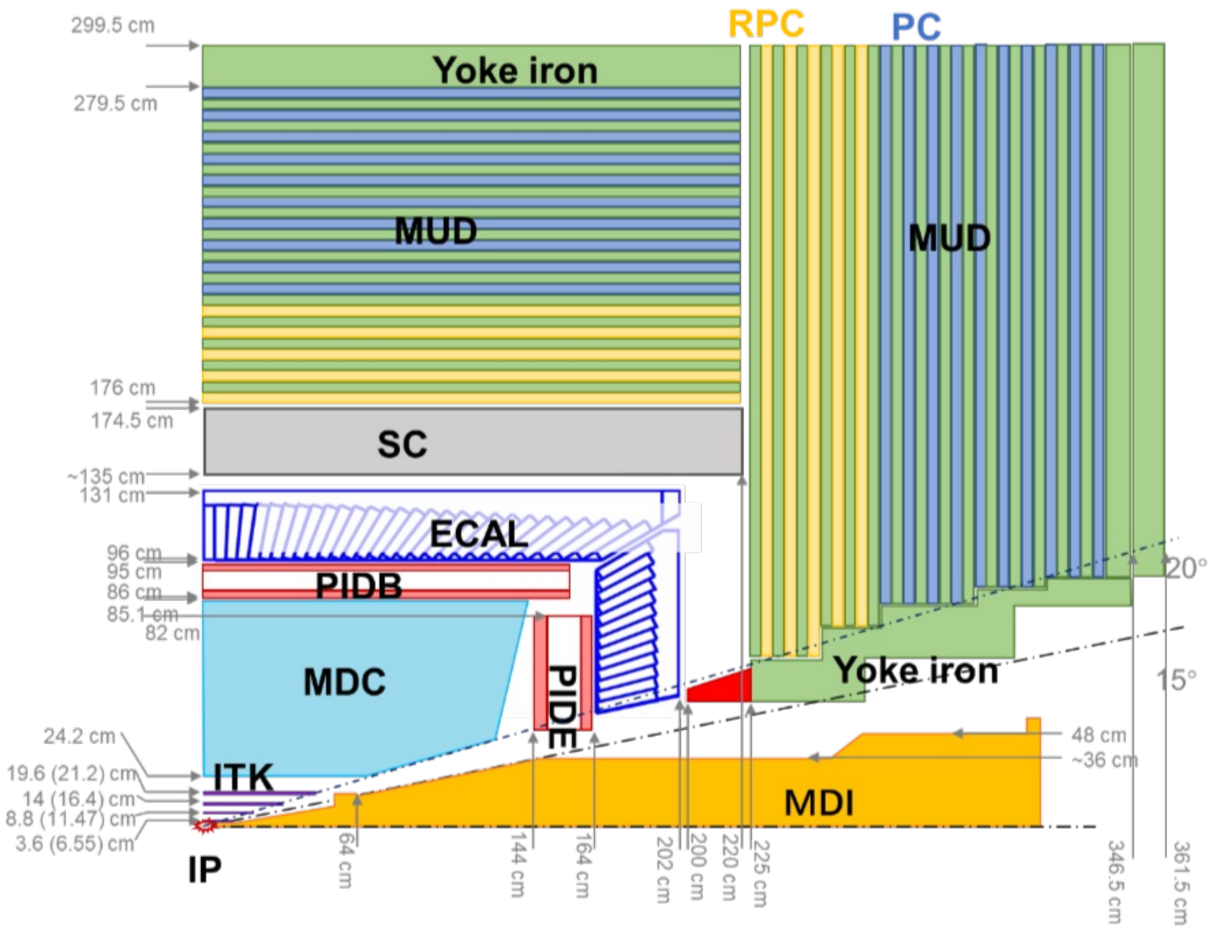
Photon energy



K_L, neutron momentum



Detector performance requirements



ITK

- Material < 0.01 X0, σ_{xy} < 100 um

MDC

- σ_{xy} < 130 um, σ_p/p < 0.5% at 1 GeV/c
- dE/dx resolution < 6%

RICH→BTOF & RICH/DTOF/ ASHIPH

- π/K PID efficiency > 97% up to 2 GeV/c
  @mis-ID rate 2%

EMC

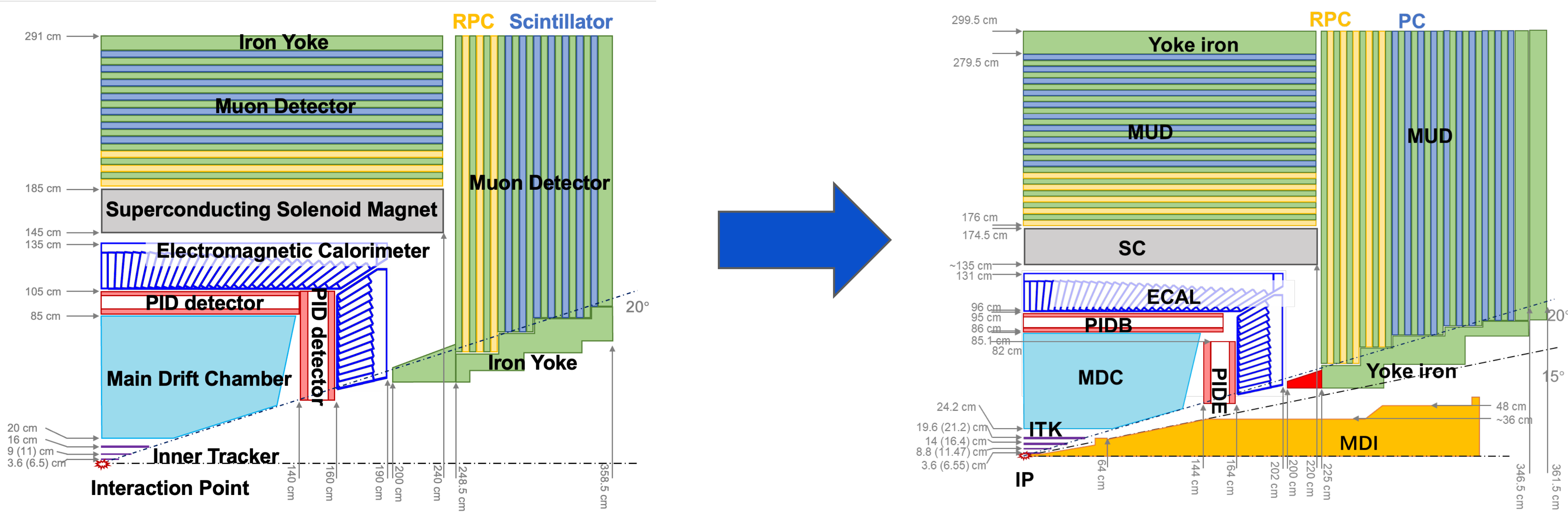
- σ_E < 2.5%, σ_{pos} < 5 mm, σ_t < 300 ps @ 1 GeV

MUD

- μ PID efficiency > 95% with $\pi\rightarrow\mu$ mis-ID
- rate < 3.3% @ p = 1 GeV/c

More about the latest detector design in [Qian Liu's talk]

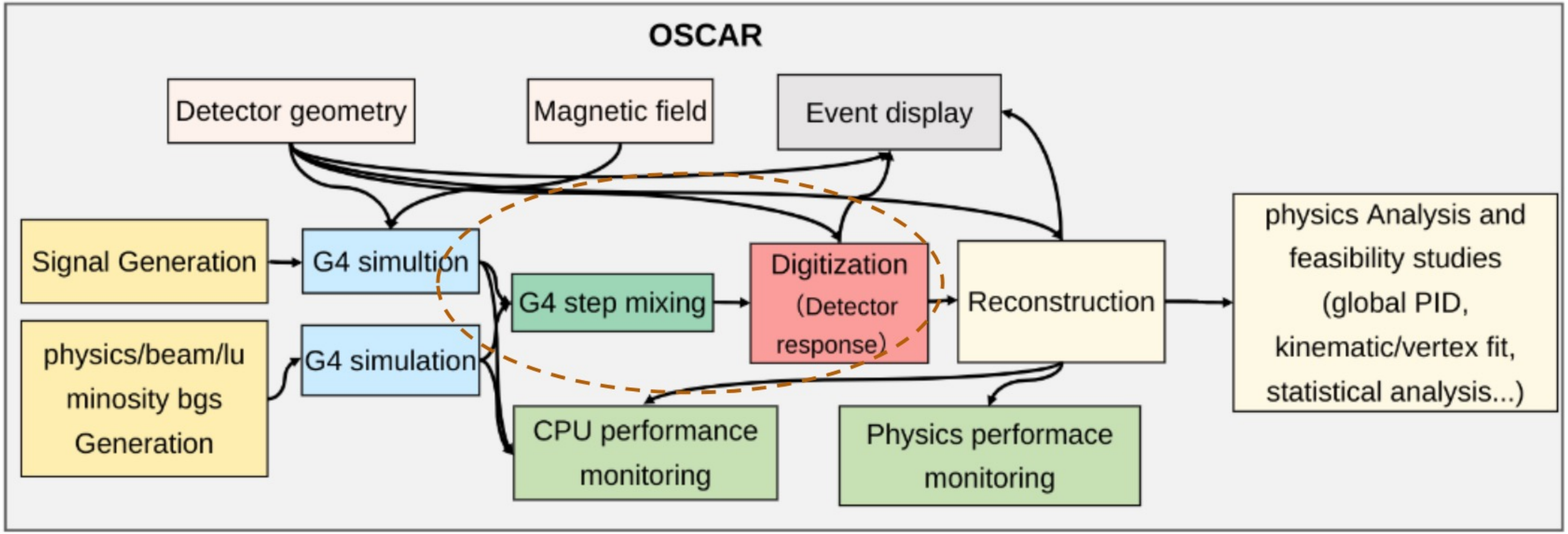
Detector geometry update



- An optimized geometry was developed to improve performance at high background levels.
  - **Tracker:** ITK 3 → 4 layers; MDC 48 → 42 layers (remove innermost & rearrange rest) .
  - **PID & ECAL:** barrel longer & narrower (PID: RICH → BTOF; ECAL: +2 crystal rings)
    endcaps narrower & outward (PID: RICH/DTOF/ASHIPH; ECAL: rearranged)
  - **MUD:** 3 RPC + 7 PS → 5 RPC + 10 PS.

STCF offline software (OSCAR)

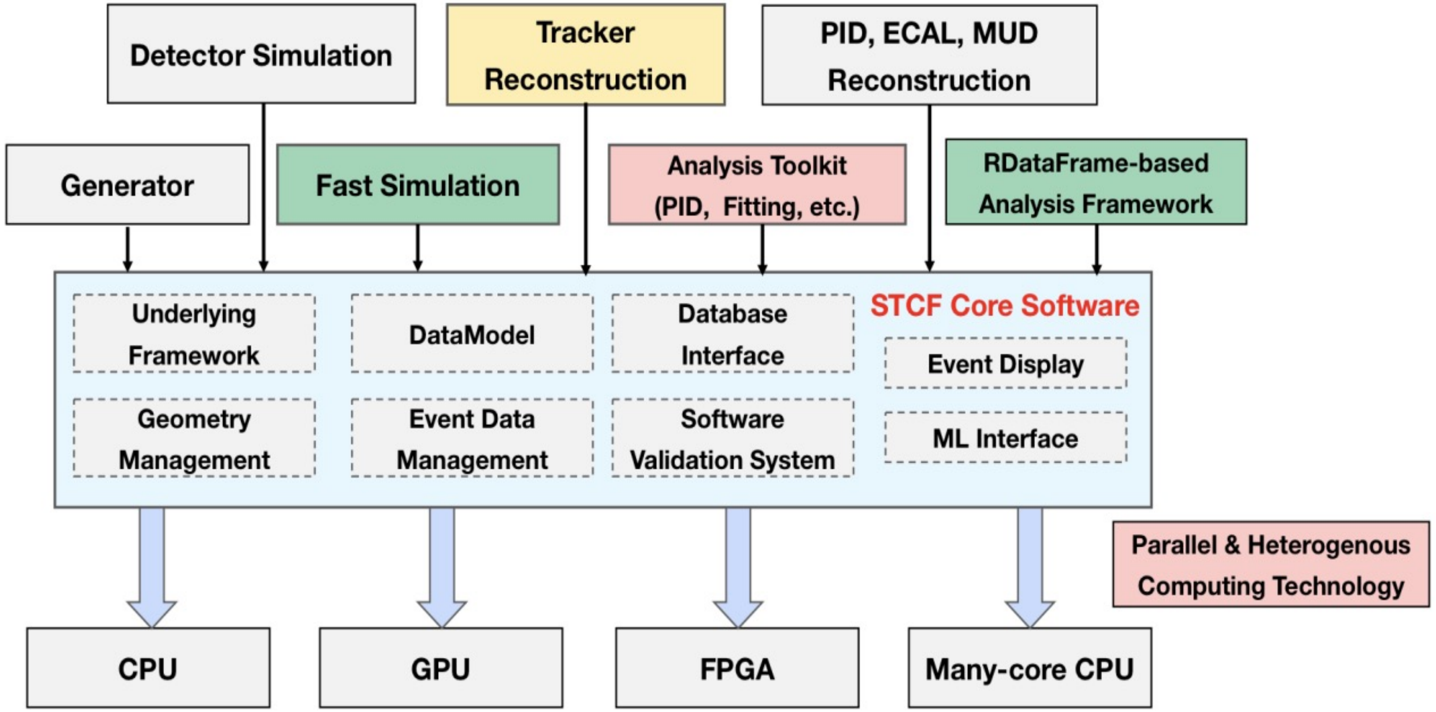
- OSCAR is developed to provide common functionalities for **detector simulation**, **digitization** , **calibration**, **reconstruction** and **physics analysis** at STCF
  - A full chain of simulation + digitization + reconstruction + analysis has been established
  - ML techniques exploited in simulation, reconstruction and analysis



STCF Core Software

OSCAR core software features

- The OSCAR core software is developed based on the SNiPER framework and the (partially) the **Key4hep** Stack, with lots of state-of-the-art techniques
  - Event Data Model based on **podio**
  - Geometry based on **DD4hep**
  - Tracking with **ACTS** (as an option)
  - Analysis framework with **RDataFrame**
  - ML inference with **ONNX**
  - Parellel computing with **TBB**
- Many new features are being develope
  - Fast simulation based on **GAN**
  - Software deployment with **Spack**
  - Support for **heterogenous** resources
  - Modern CI powered by **LLM**

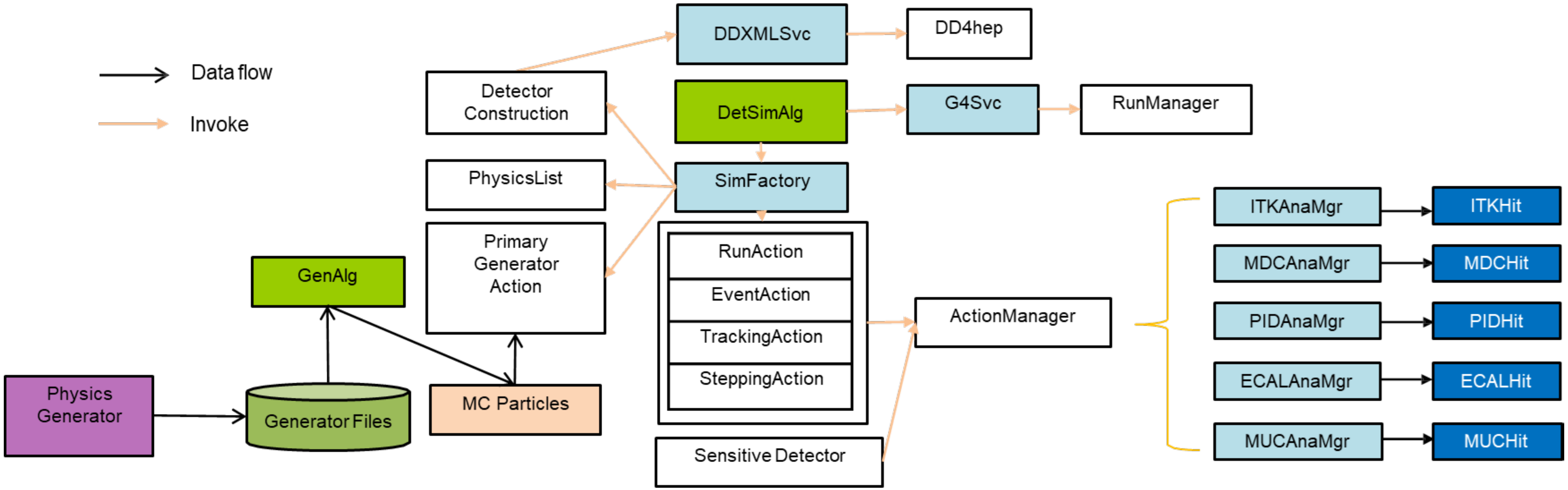


Comput.Softw.Big Sci. 9 (2025) 1, 3
Mod.Phys.Lett.A 39 (2024) 40, 2440006
JINST 18 (2023) 03, P03004
J.Phys.Conf.Ser. 2438 (2023) 1, 012054

Simulation, digitization, background mixing

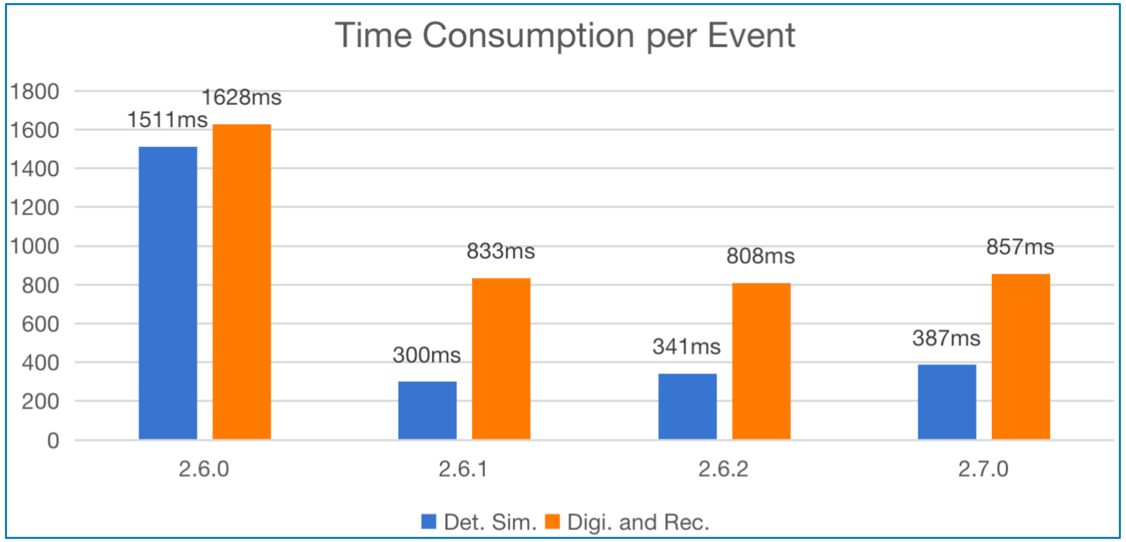
Full simulation

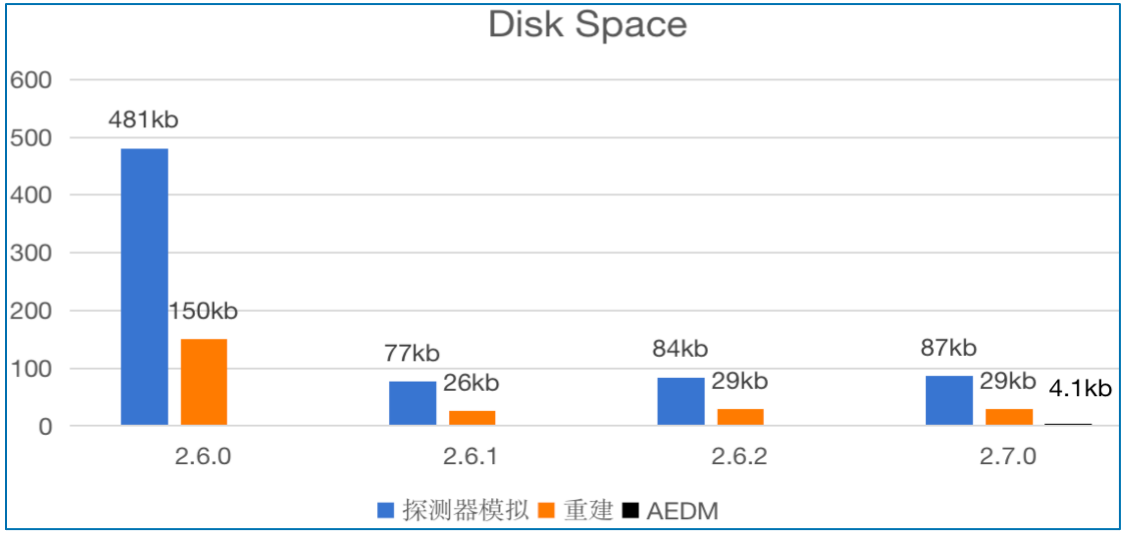
- Built **full simulation chain** from generated events to Geat4 hit information
- Providing **flexible configuration** in different steps
  - **Generator** for different physics topics i.e. Babayaga, KKMC, Phokhara, DIAG36, BBBrem …
  - **Geometry** for different detector design options
  - **User actions** for recording MC truth information, G4Step level

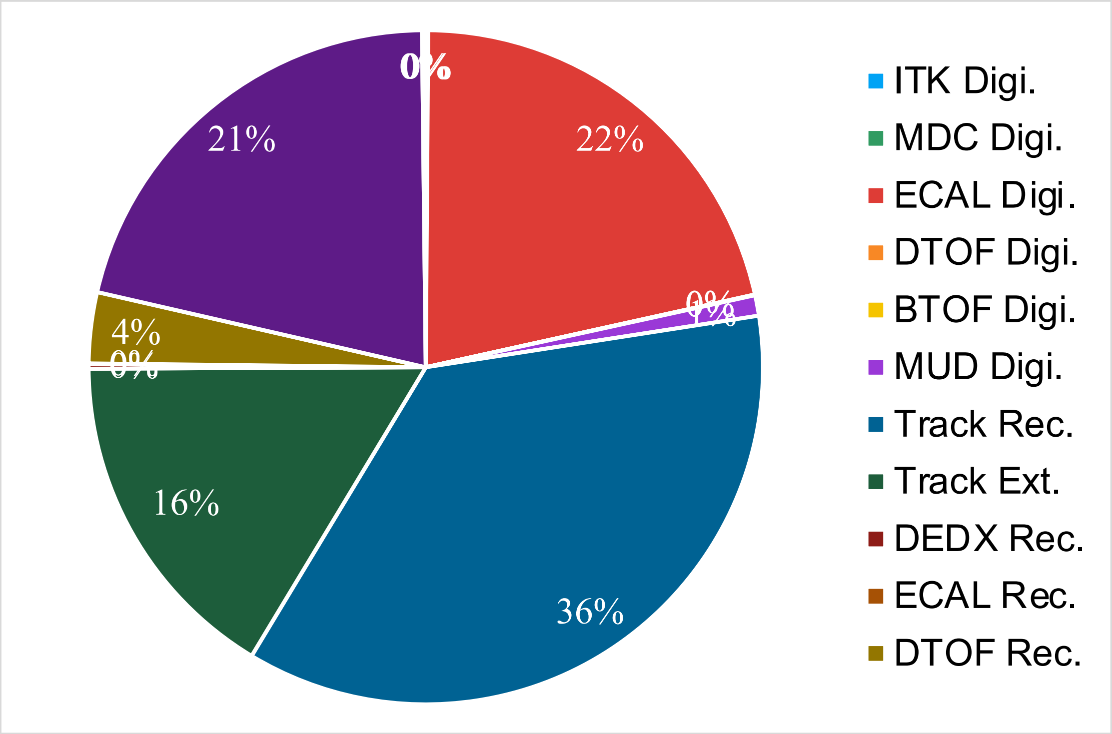


CPU performance of the full chain

- Much improved CPU consumption and disk space (comparable to BESIII now)



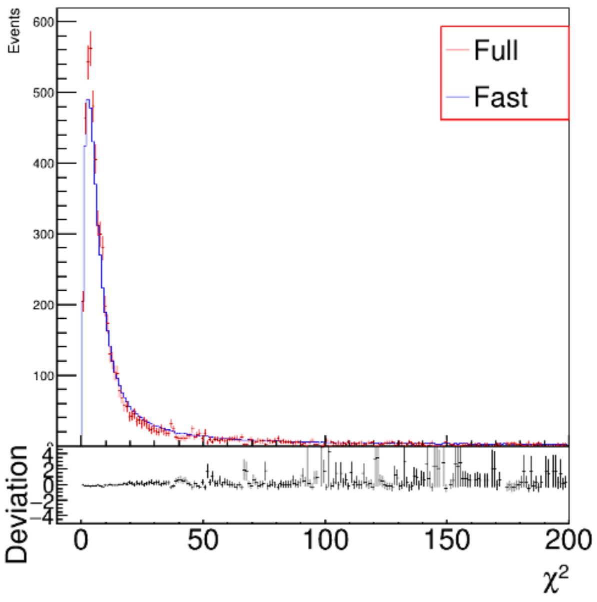


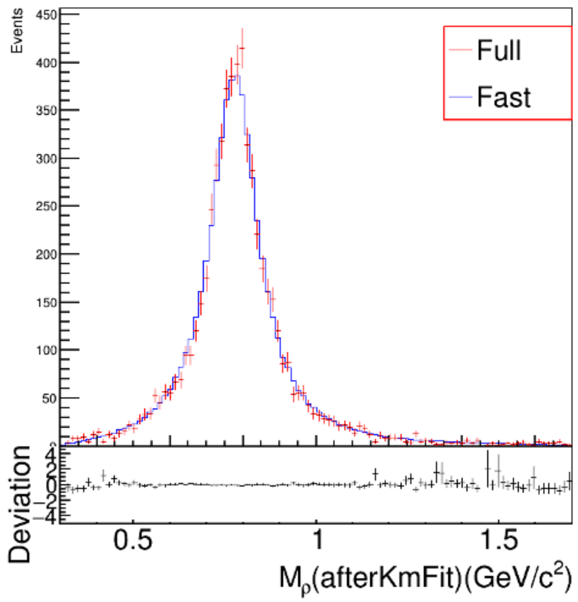


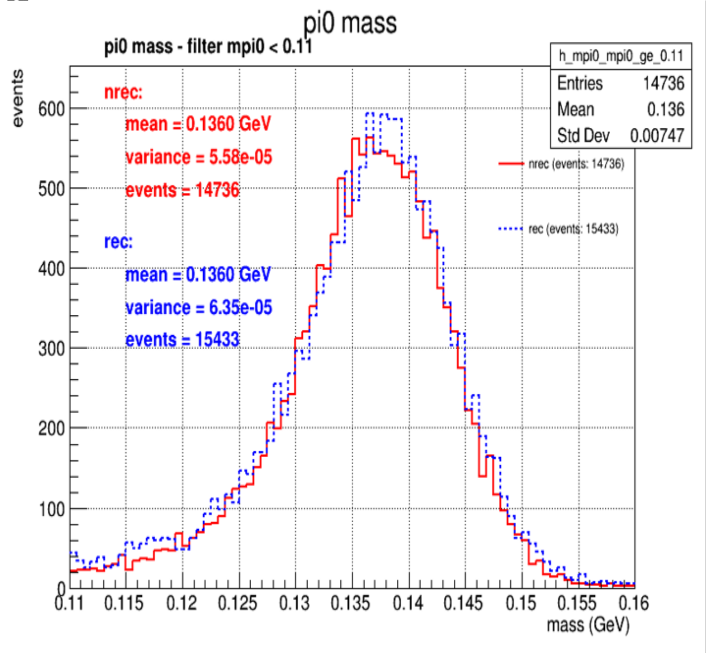
Digitization and reconstruction in OSCAR-2.7.0

Fast simulation

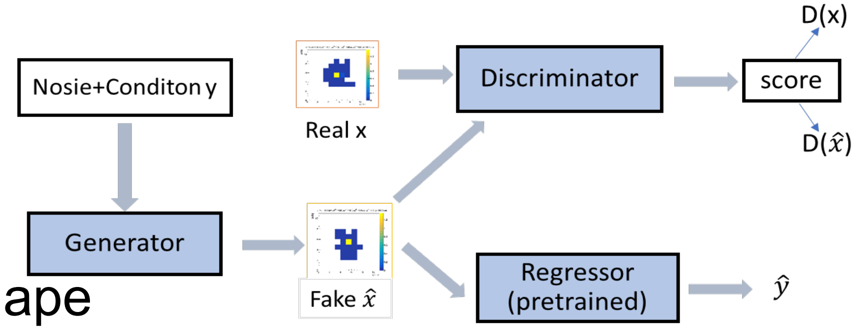
- Parameterizations of STCF **detector responses (scalable efficiency, resolution…)** to
- "fake" responses from full simulation + digitization + reconstruction
- Performance validated against full simulation using multiple channels: $\rho\pi^0$, $\pi\pi J/\psi$, $\Lambda\bar{\Lambda}$…
- Speed up and storage reduction:
  - ~**2 ms** per event, ~**1.8 kb** storage for a 4 prong J/ψ decay event
- Feasible detector response scaling realization for detector optimization
- Muti-dimentional sampling under study, consistency improved





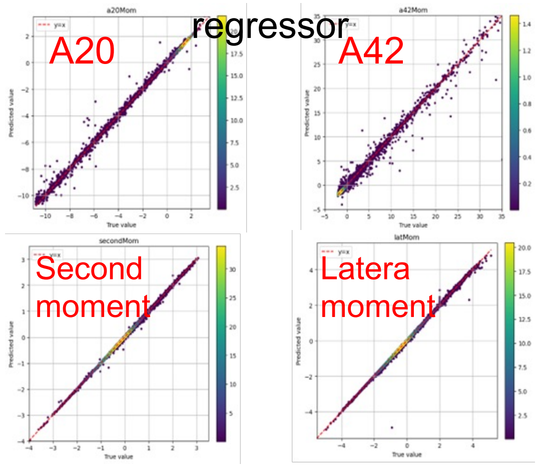


GAN EMC Fast simulation

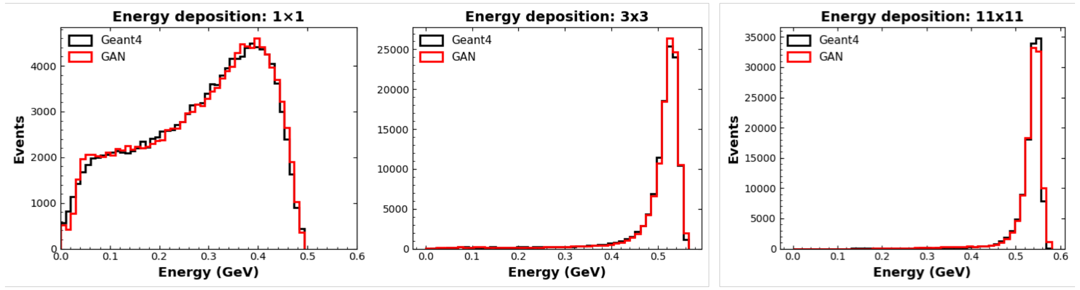


- Model architecture and training strategy
  - **Regressor**： **a pretrained network**, to supervise shower-shape observables, including A20, A42, second moment, and lateral moment.
  - **Input condition**: particle momentum, inject direction, momentum direction, and z-position.
  - **Training Strategy:** The 0–2 GeV range was divided into four intervals, with an independent model trained for each interval.
- **Generation speed**: ～0.15 ms per particle on CPU, achieving **over 100 times speedup**. (Intel Xeon Gold 6338@2.00GHz)

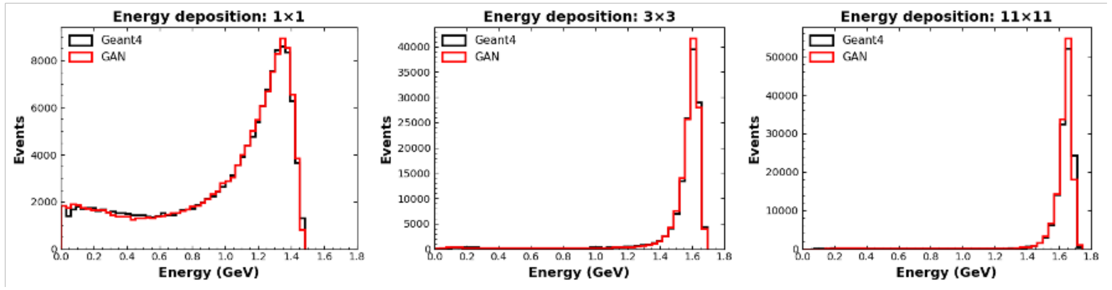
Performance of regressor



0.6GeV test for electrons:

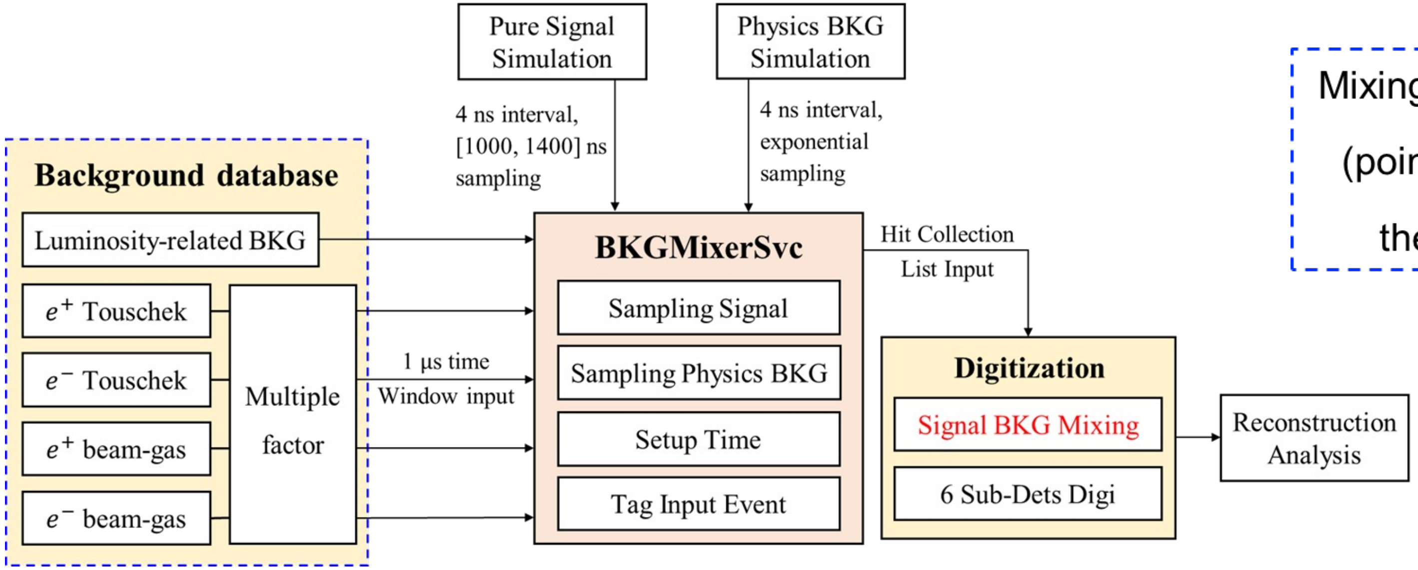


1.8GeV test for electrons:



Background mixing and digitization

- Signal particles mixing with background particles at the **Geant4 step level**
  - To vastly improve the speed, optimization is ongoing to pre-digitize the background hit and do the **mixing partially at digi. hit level**
- Geant4 step info digitized to emulate the **detector measurement** (i.e. input for reconstruction)
  - Considering **electronic response, noise and other effects**
  - All sub-detector uses **sampling method,** including **MDC dE/dx** and **hit position**



Mixing in Geant4 Step Level

(points or hits in OSCAR)

then to do digitization

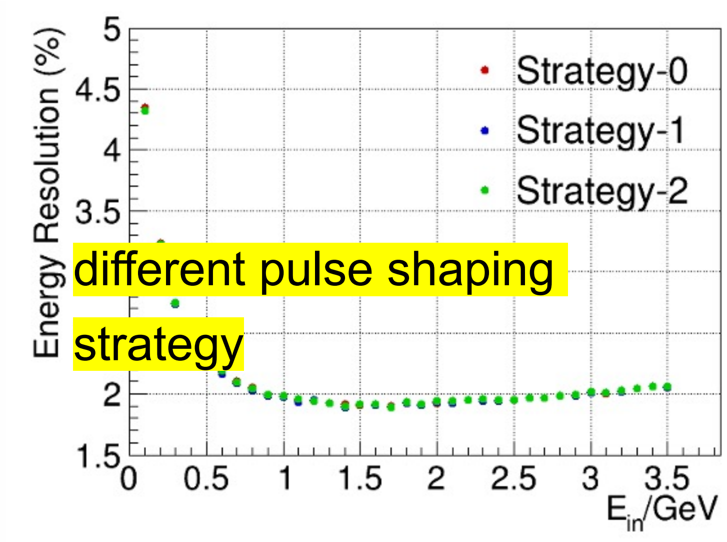
Flexible Configuration for digitization

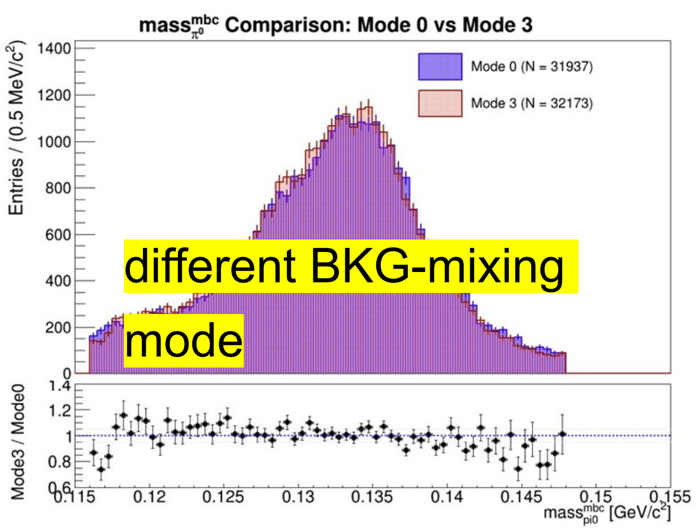
- Flexible configuration to **balance accuracy and efficiency** of digitization and BKG mixing.
- Some optimization setups to save the **consumption of CPU time amd RAM**.

ITKW	Mode: smearing, sampling or Garfield simulation	
ITKM	Mode： sampling or simulate chage propogation	Sampling Method: sampling from diffrent files, depend on particle type or deposition energy
MDC	Mode: smearing, sampling or Garfield simulation	
BTOF/RICH	Mode: single or double gaussion sampling for MCP-PMT TTS	
RICH	Flexible sampling parameters	
ECAL	Shaping strategy: wave response sampling or convolution	Fitting Method: single templete fitting or pipeline method
MUD	Mode： smearing or detector response sampling	
BKG mixing	Mixing in hit or in digi level, diffrent mode for digi-mixing	

ECAL digitization optimization

- ECAL BKG-mixing and digitization **dominates the overall CPU time** among all sub-detectors.
- Extensive efforts have been undertaken to **minimize** the processing time
  - Fitting only over threshold **(~30 s/event**)
  - Sampling → convolution in pulse shaping **(~10 s/event**)
  - Mixing in didi level **(~300 ms/event**)



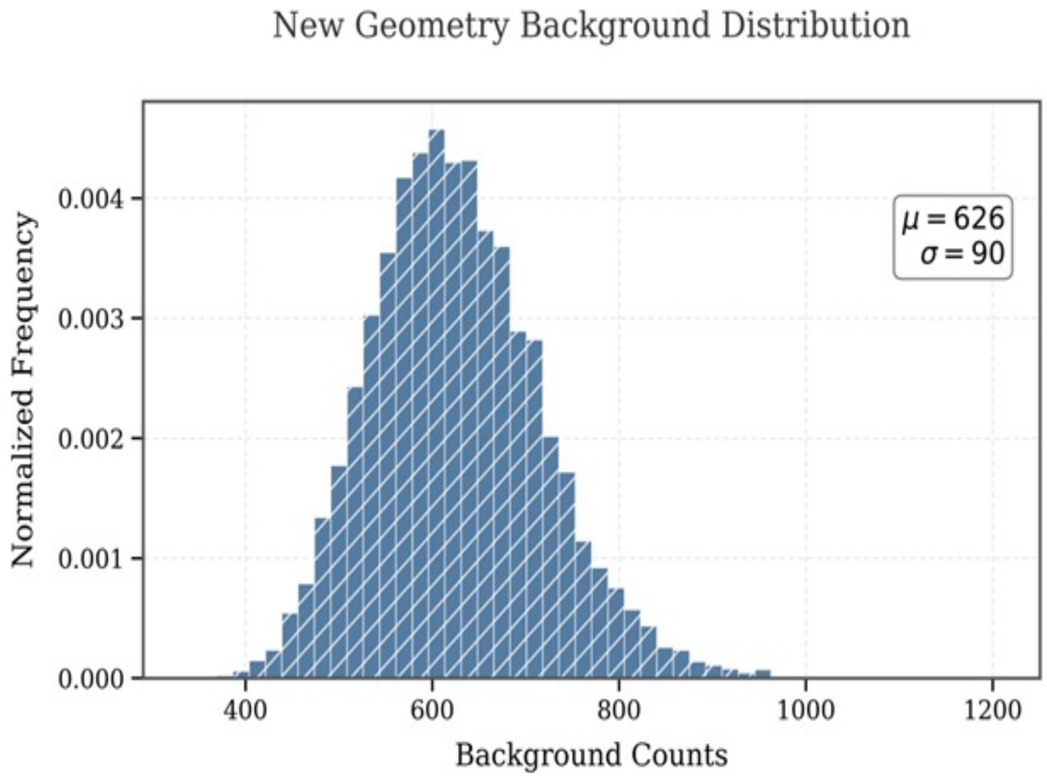


More details in Wang Bo and Zhou Zheyan'talks

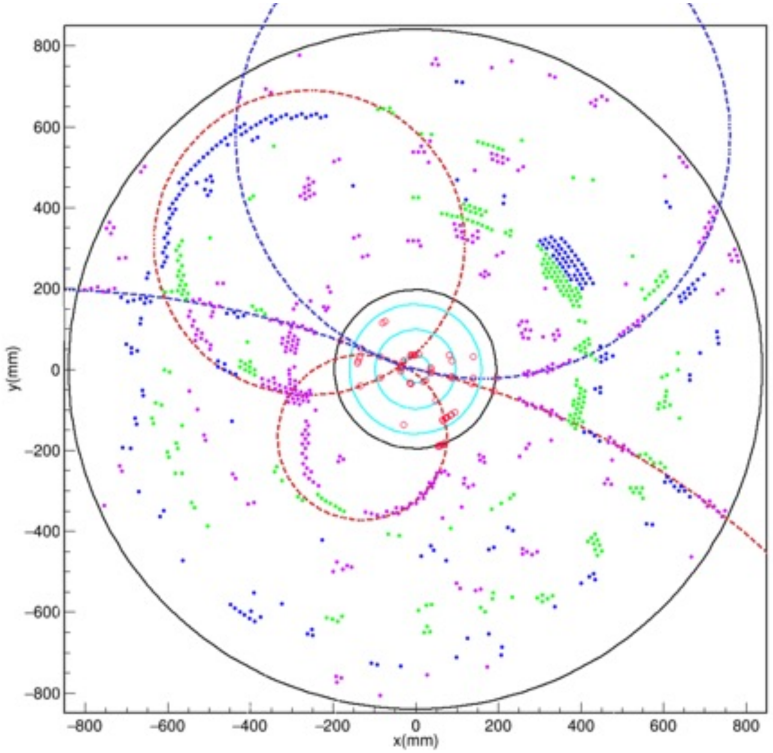
Tracking performance

STCF tracking environment

- Both prompt and **displaced** tracks exist with presence of **noise hits (~630/event)** from beam backgrounds
- Most tracks have p_T below 500 MeV



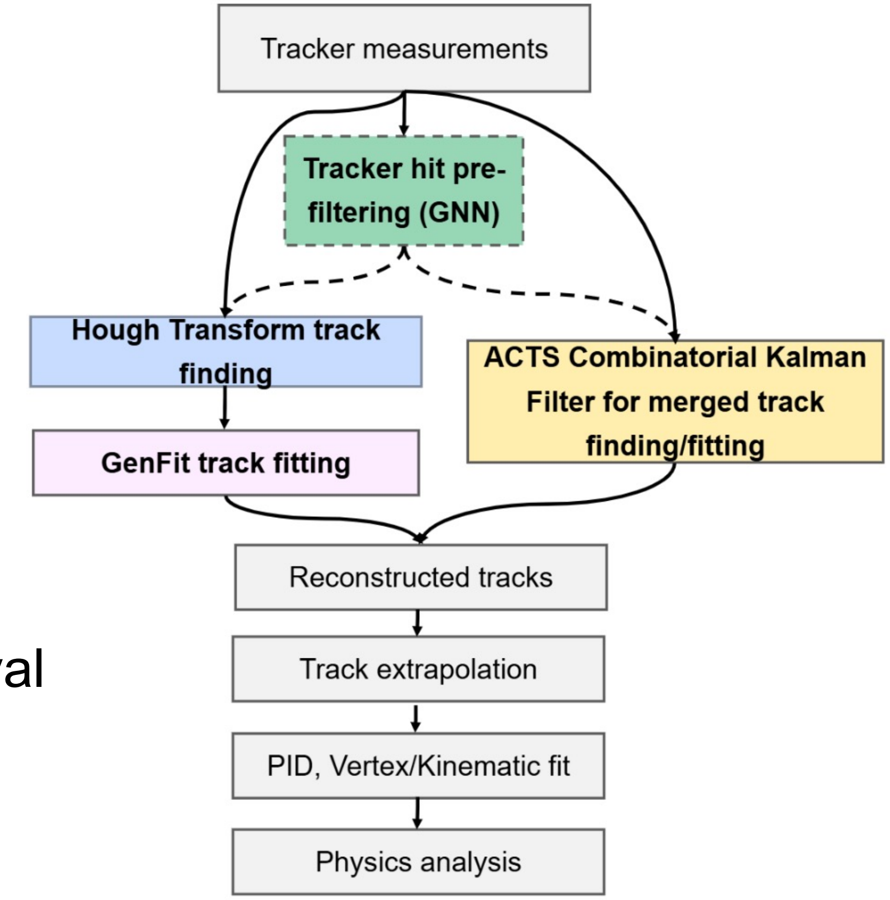
$J/\psi \to \Xi^-\bar{\Xi}^+ \to \Lambda\pi^-\bar{\Lambda}\pi^+ \to p\pi^-\pi^-\bar{p}\pi^+\pi^+$



Tracking strategies

Different strategies are investigated to improve the overall performance with new tracker geometry:

- **Hough transform** + **GenFit** shows decent performance for both prompt and long-lived particles
- **ACTS** is integrated to improve tracking efficiency of low p_T particles and reduce CPU overheads
- **GNN** shows promising performance in terms of noise removal



More details in talks:

→ [STCF tracker layout and tracking performance](#) (Jin Zhang)

→ [STCF MDC中基于图神经网络寻找径迹衰变拐点](#) (Xinghui Yang)

Hao Zhou et al., *NIMA 1075, 170357 (2025)*

Hao Li et al., *NUCL SCI TECH 36, 171 (2025)*

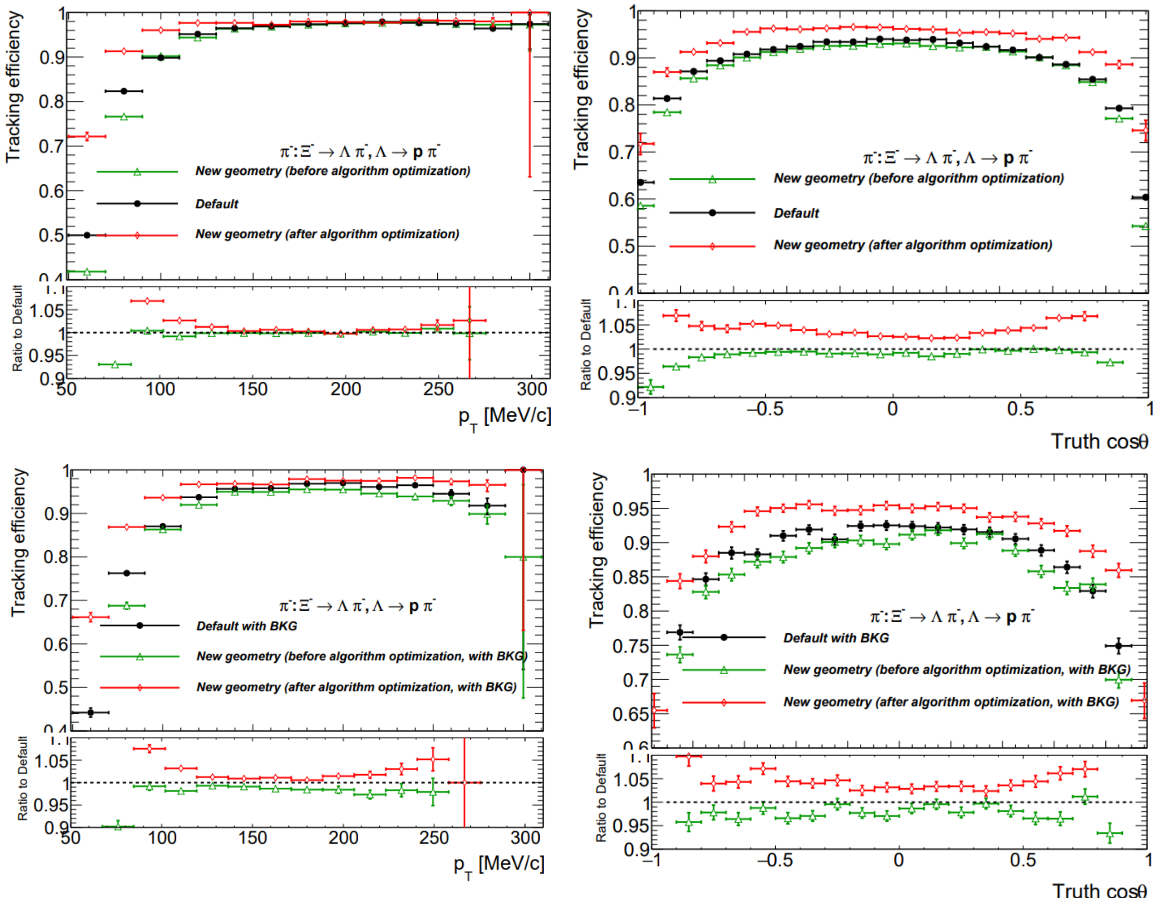
Xiaoqian Jia et al., *JINST 21, P03009 (2026)*

Zhenna Lu et al., ChinaXiv: chinaxiv-202604.00144 (received, NUCL SCI TECH)

Improved tracking efficiency after Hough optimization

- Optimize the algorithm with the new tracker geometry
  - Reconstruction efficiency for **secondary-vertex and low-momentum tracks improved significantly**

$$J/\psi \to \Xi^-\bar{\Xi}^+ \to \Lambda\pi^-\bar{\Lambda}\pi^+ \to p\pi^-\pi^-\bar{p}\pi^+\pi^+$$



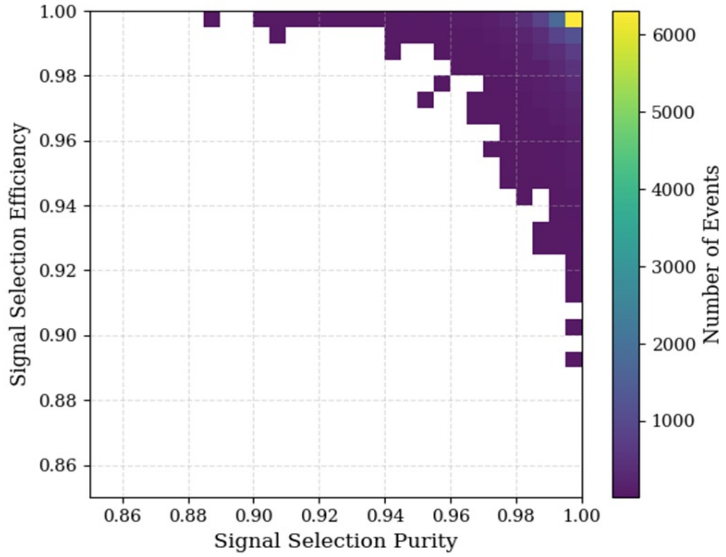
w/o background

w/ background

Performance gain with GNN and ACTS

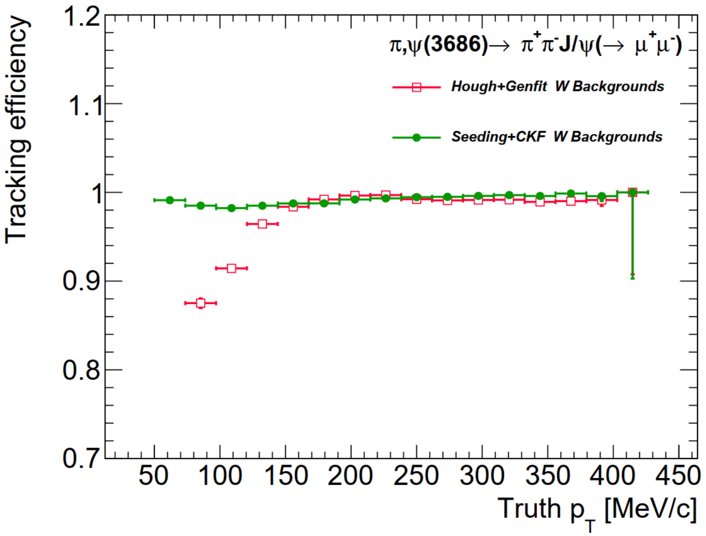
- GNN noise filter provide an exelent performance in reducing background
- ACTS can improve the efficiency at low p_T, but more fake tracks need to be resolved

$\psi(3686) \rightarrow \pi+\pi\text{-}J/\psi,\ J/\psi \rightarrow \mu+\mu\text{-}$

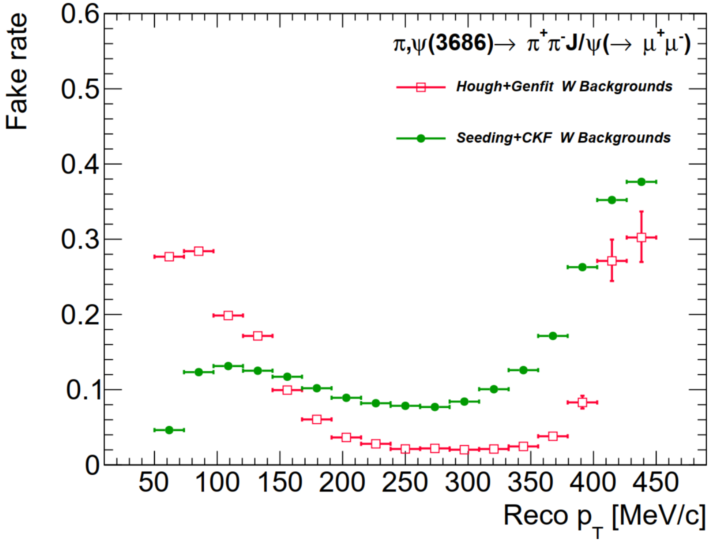


- 击中分类准确率：**99.4%**
- 信号挑选效率：**98.7%**
- 噪声移除率：**99.6%**

ACTS (seeding+CKF) tracking efficiency vs. pT
(π^- in $\psi' \rightarrow \pi^+\pi^-J/\psi \rightarrow \pi^+\pi^-\mu^+\mu^-$)

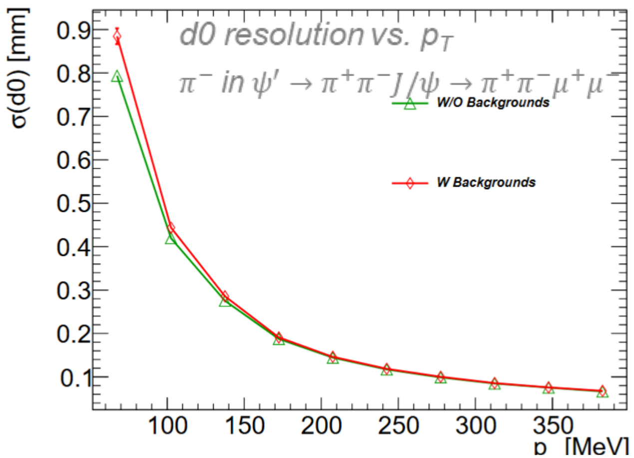


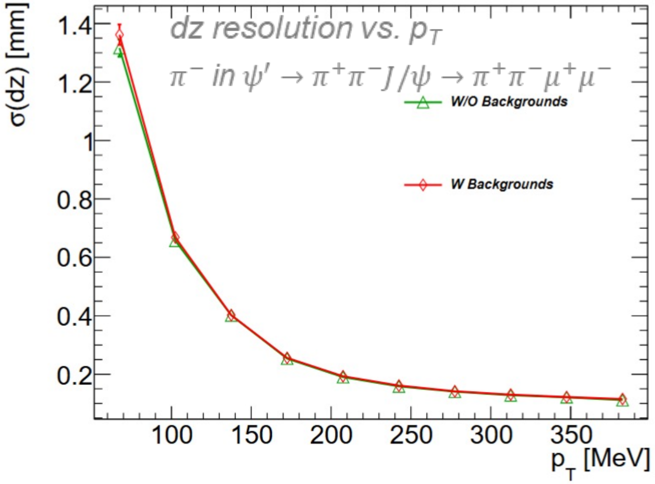
ACTS (seeding+CKF) fake rate vs. pT
(π^- in $\psi' \rightarrow \pi^+\pi^-J/\psi \rightarrow \pi^+\pi^-\mu^+\mu^-$)

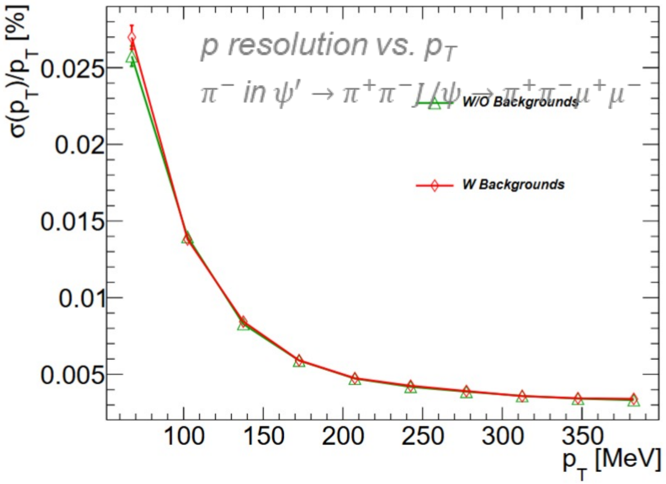


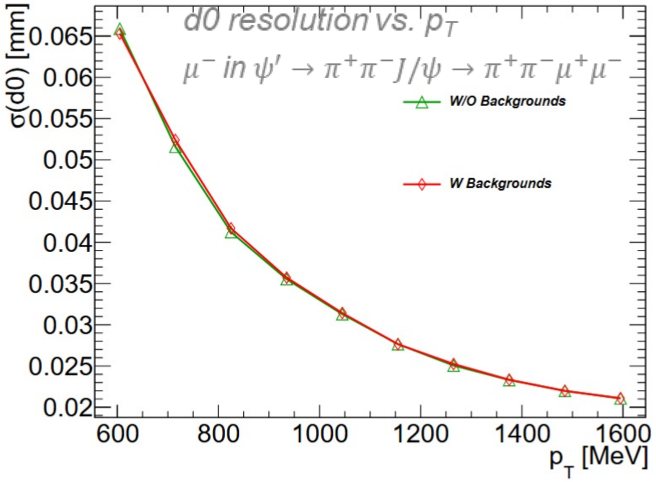
Decent tracking resolution with GenFit

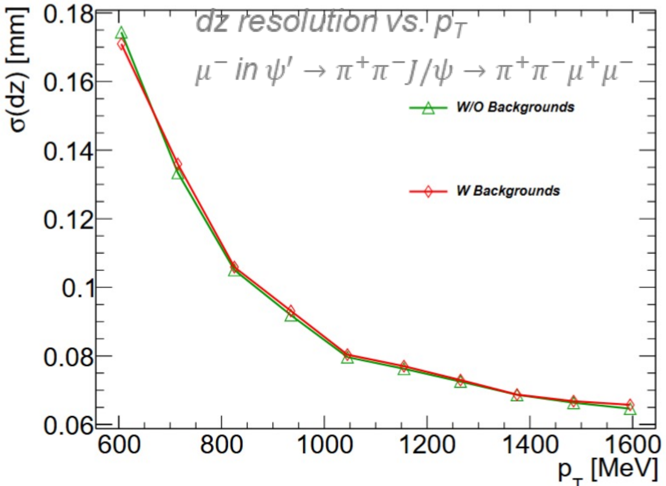
- d0 (z0) has resolution down to **32 (85) um** with p_T > **1 GeV**
- p_T relative resolution is ~**0.5%** at p_T = **1 GeV**

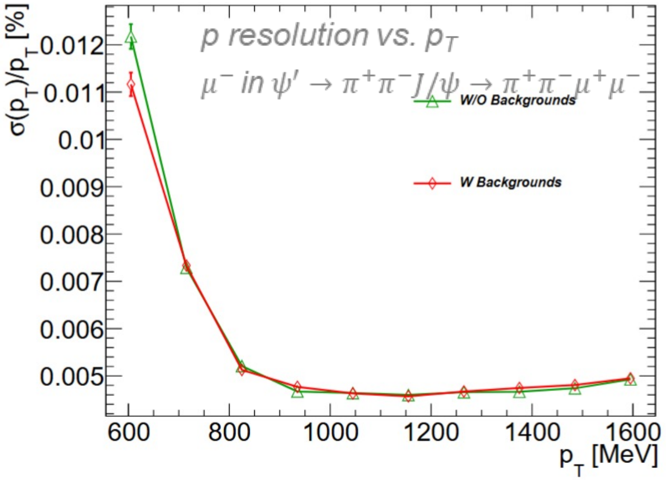










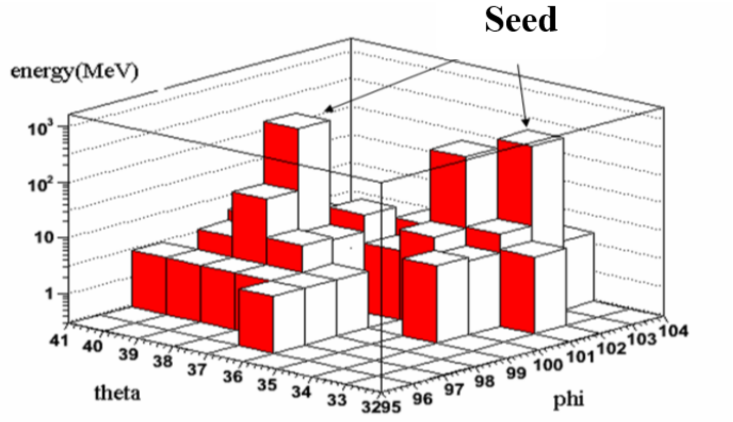


Photon performance

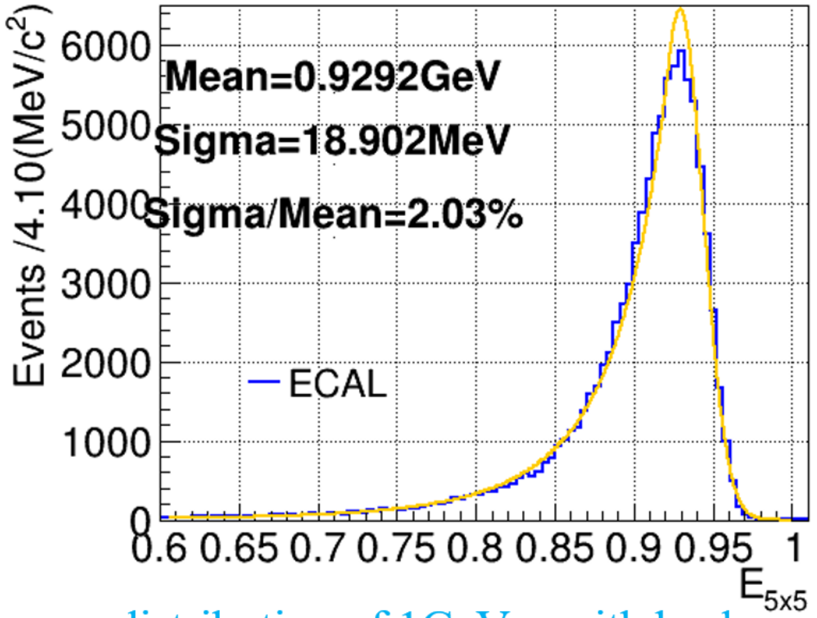
EMC reconstruction

ECALRecAlg

1. **Cluster search**
2. **Seed search**
3. **Clusters split into Showers**
4. **Calculation and correction of shower energy and position**



π^0 Cluster (two photons)



Energy distribution of 1GeV γ with background

- Fitted by Crystal Ball function
- Energy resolution defined by

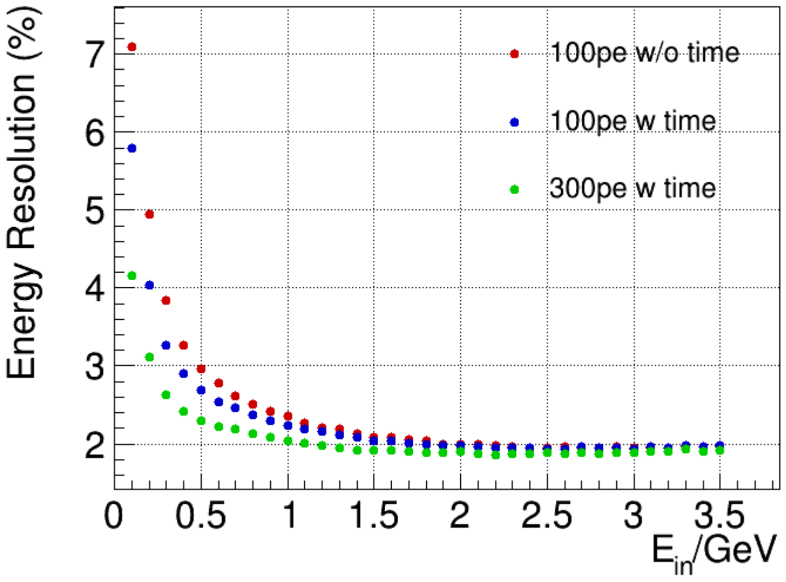
$$\sigma_E = \frac{FWHM}{2.355}$$

More details in Bo Wang's talk

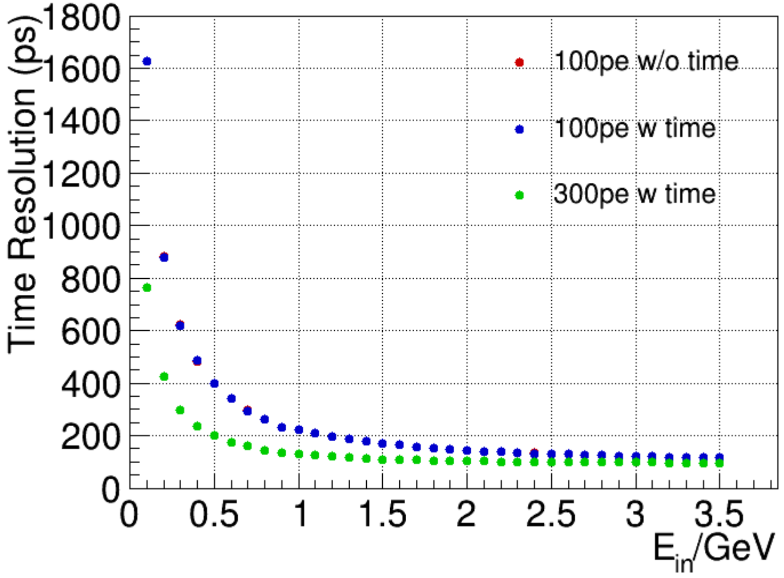
Photon reconstruction performance

- Using realistic digitization parameters from beam test results
- Single **photon energy**, **time and position resolution** with backgrounds can **meet physics requirements**

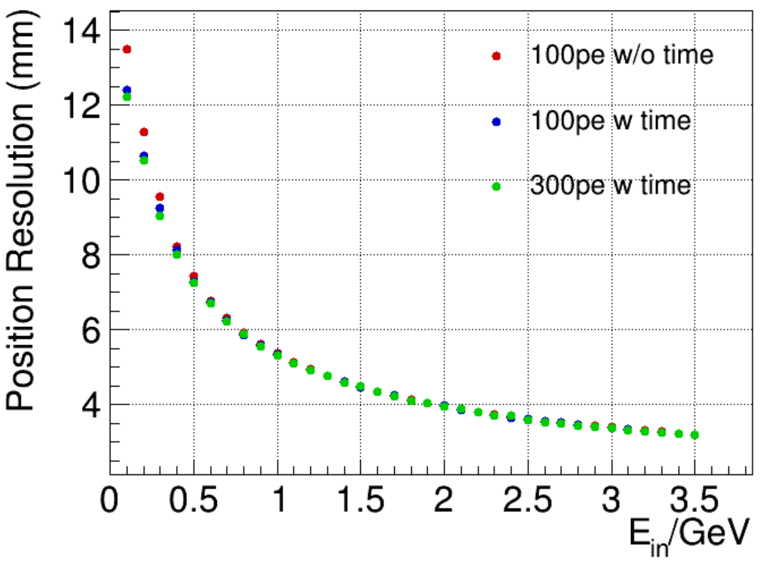
Energy resolution



Time resolution



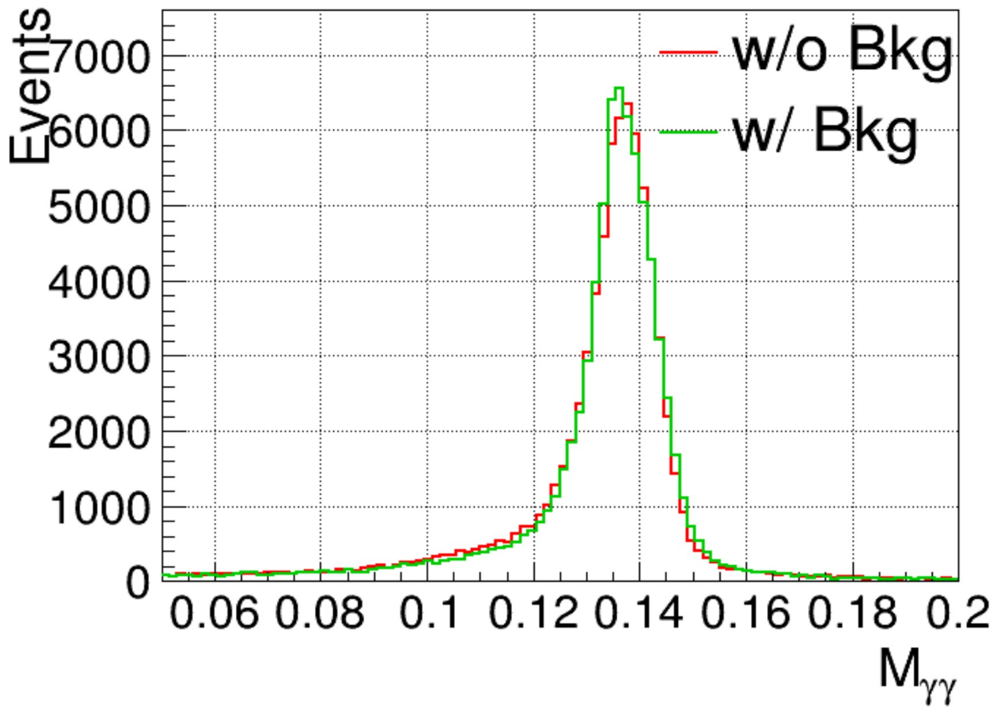
Position resolution



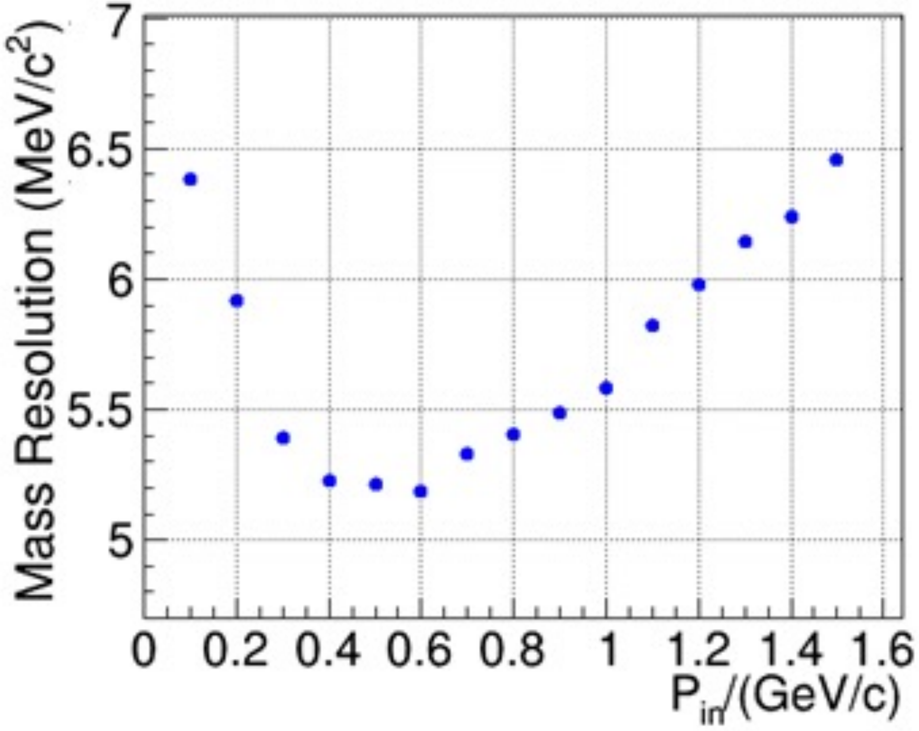
π^0 reconstruction performance

- Developed **4D EMC clustering algorithm** to improve photon resolution against beam backgrounds
  - No degradation of π^0 resolution with beam backgrounds

π^0 mass resolution in $J/\psi \to \rho\pi$



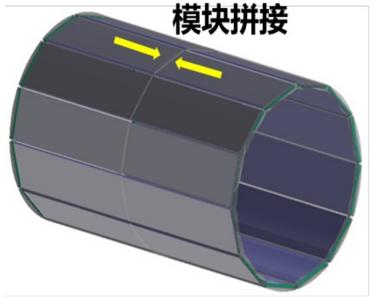
π^0 mass resolution vs π^0 momentum

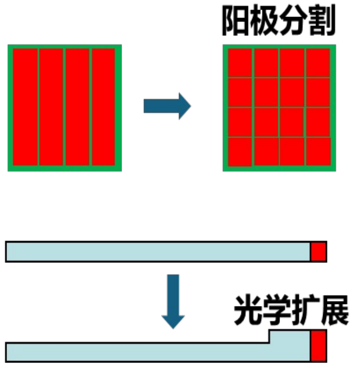


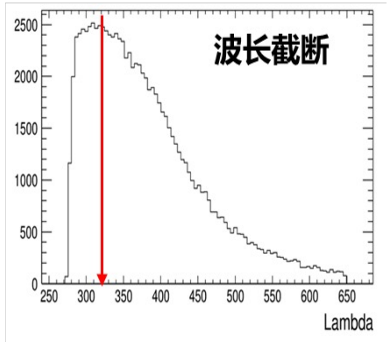
PID performance

π/K PID performance (BTOF)

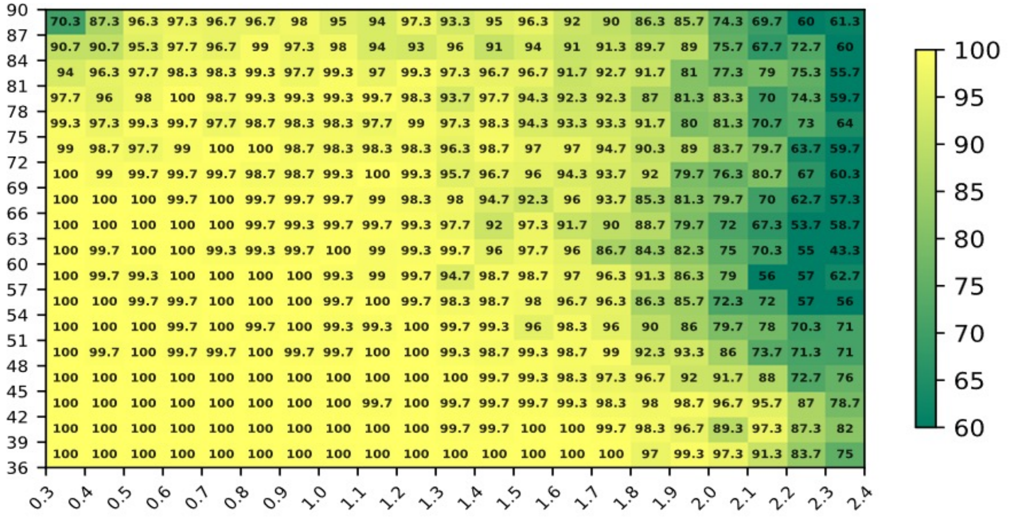
- Geometry optimized, improving performance at p > 2.5 GeV/c
- Three **BTOF** reconstruction algorithems (**Timing/Imaging/CNN method**), achieving >97% π eff @ K mis-ID=2%



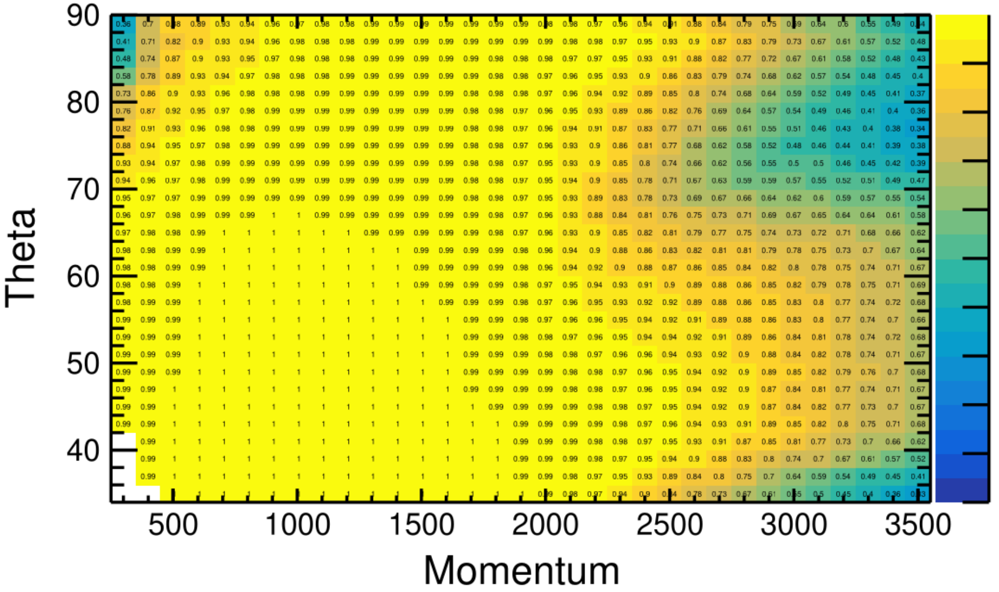




π/K PID efficiency (CNN method)



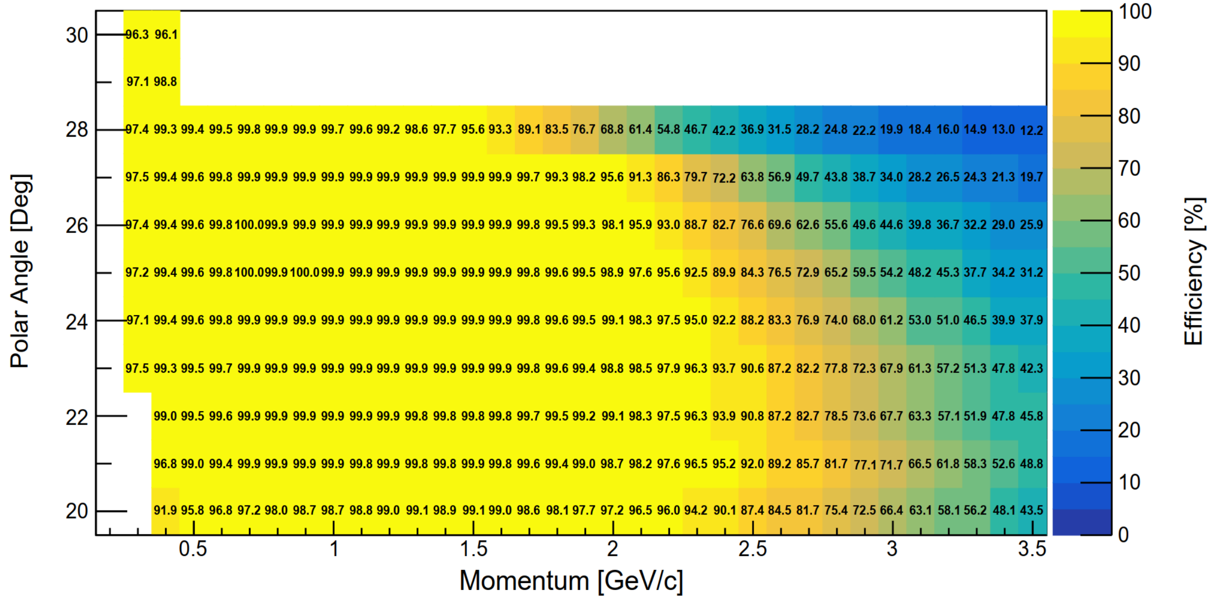
π/K PID efficiency (imaging method, optimized geometry)



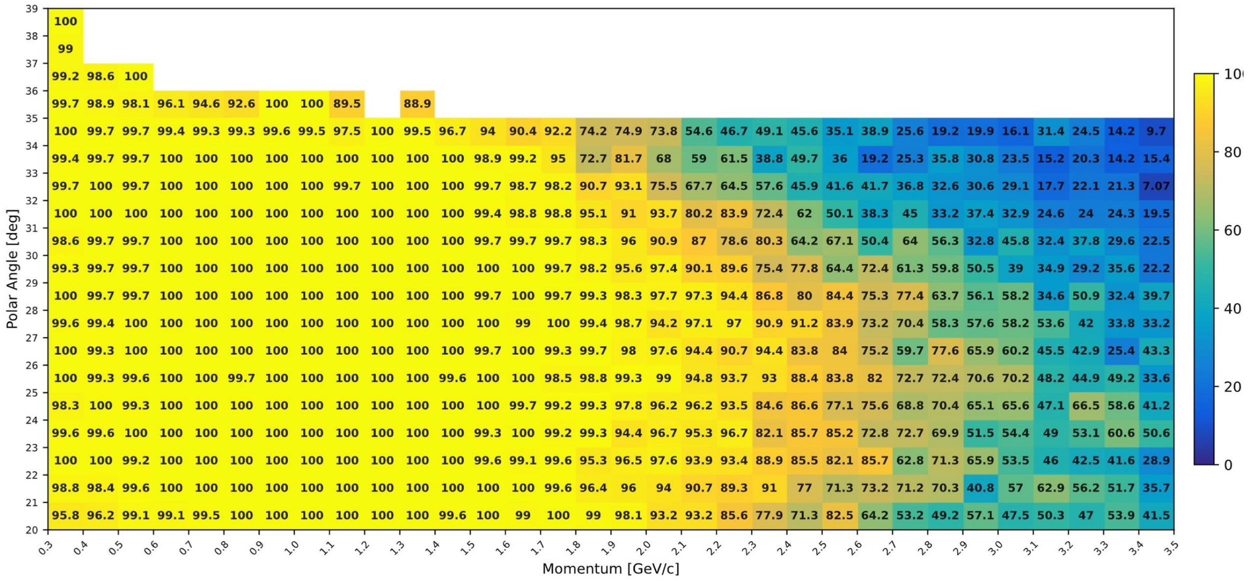
π/K PID performance (DTOF)

- Four **DTOF** reconstruction algorithms (**Timing/Imaging/CNN/Transformer method**), achieving >97% π eff @ K mis-ID=2%
  - At **high momentum**, Transformer outperforms CNN

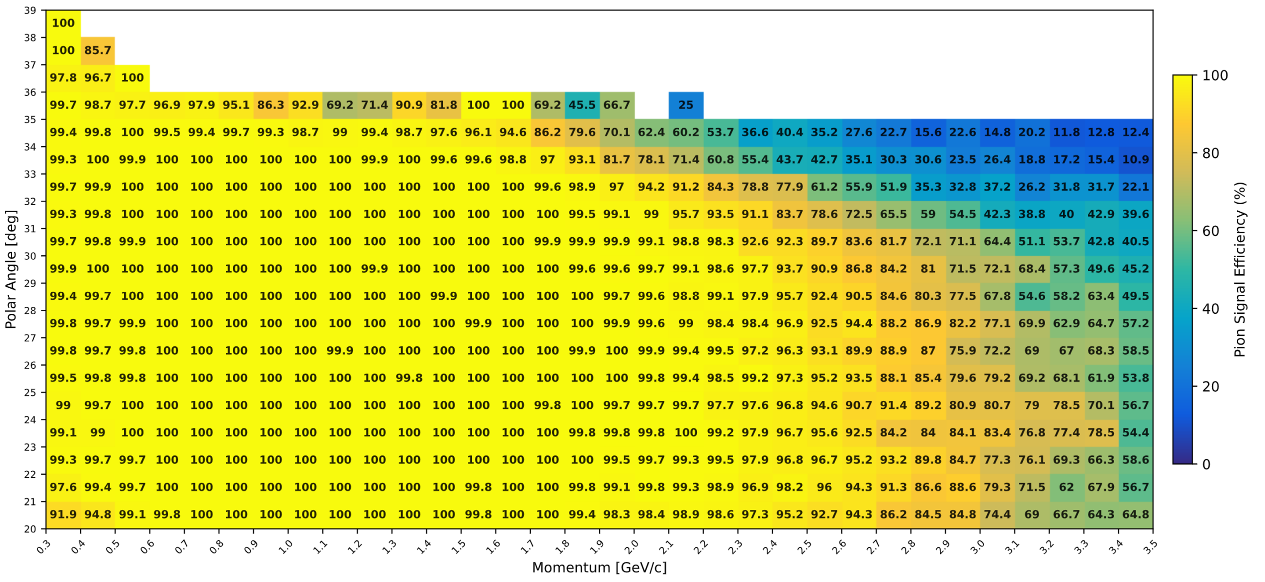
π/K PID efficiency
(imaging and timing hybrid, new geometry)



π/K PID efficiency (CNN method)

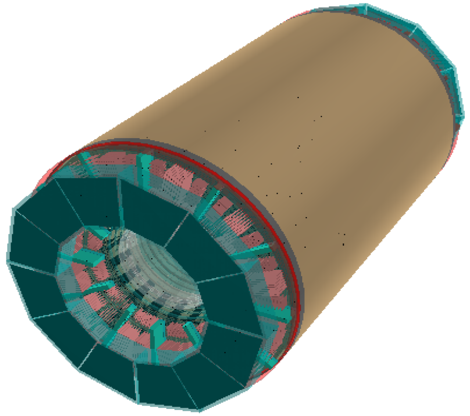


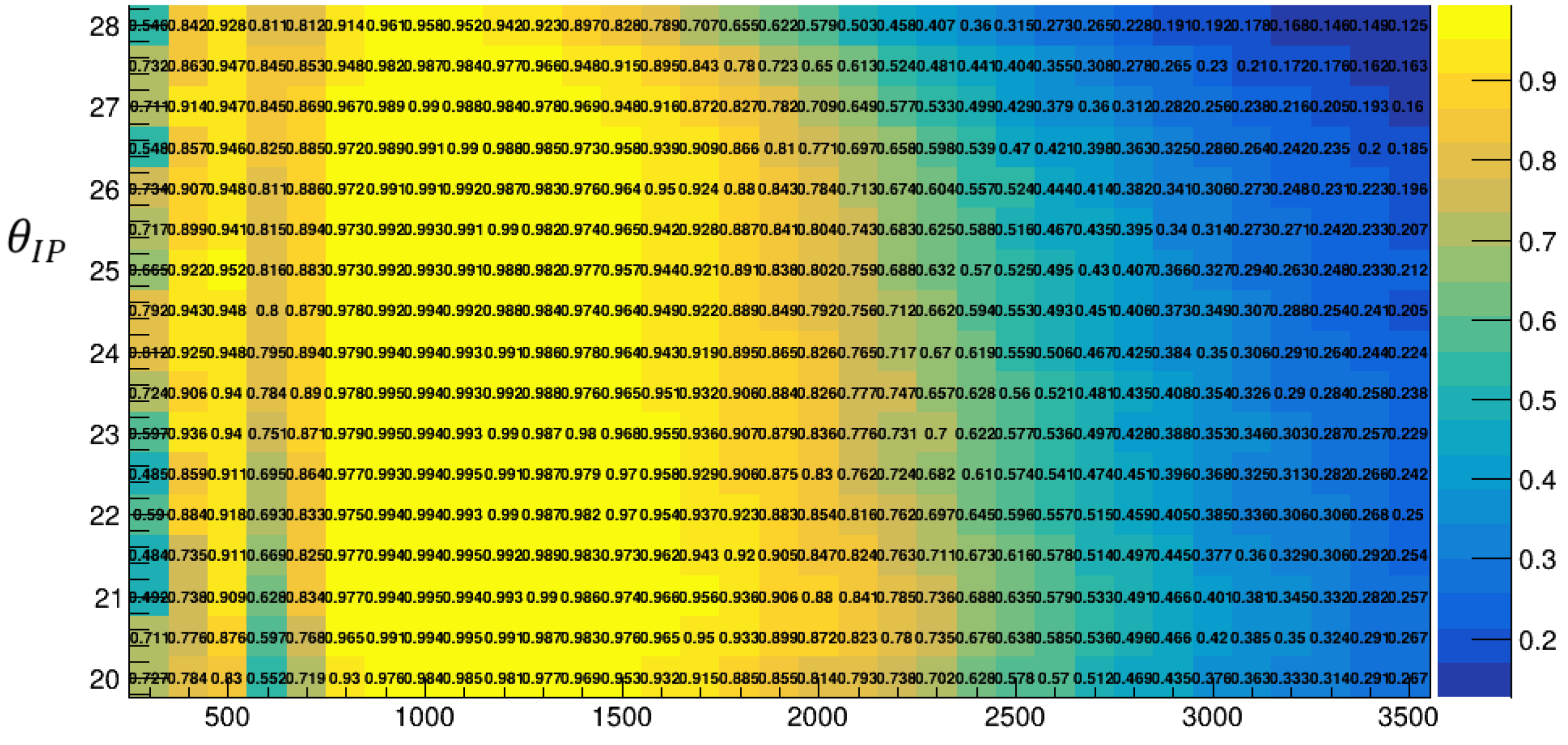
π/K PID efficiency (Transformer method)



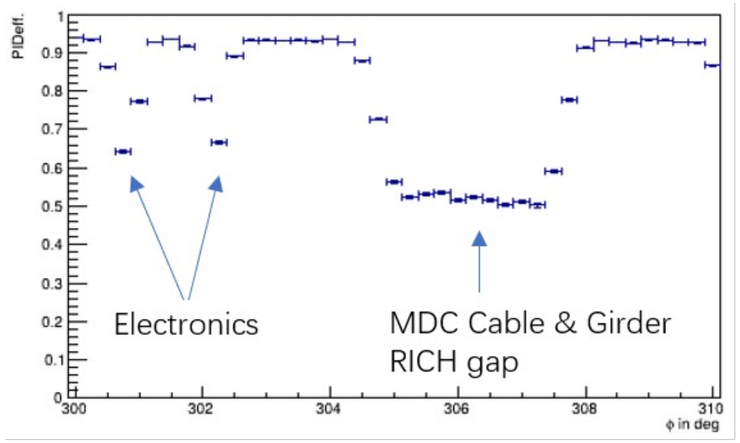
π/K PID performance (endcap RICH)

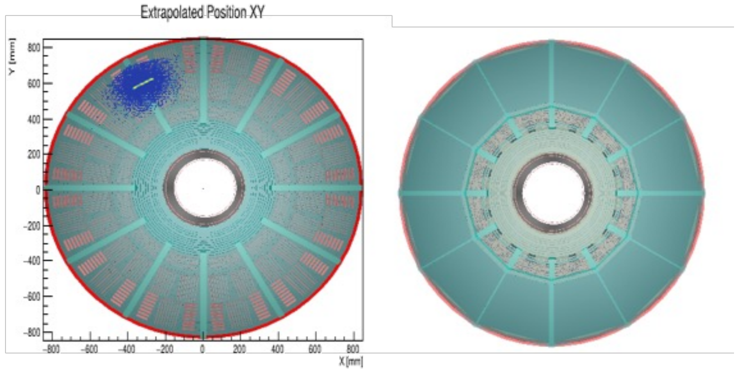
- Geometry: 4 sectors, 12 detetor plane
- PID efficiency: ~80% for pion @ 2 GeV/c
- Sensitive to the material budget of the MDC end caps
  - imporveent schemes are under inverstigation





π $PIDeff.$ @ $MissID$ $from$ $Kaon = 2\%$



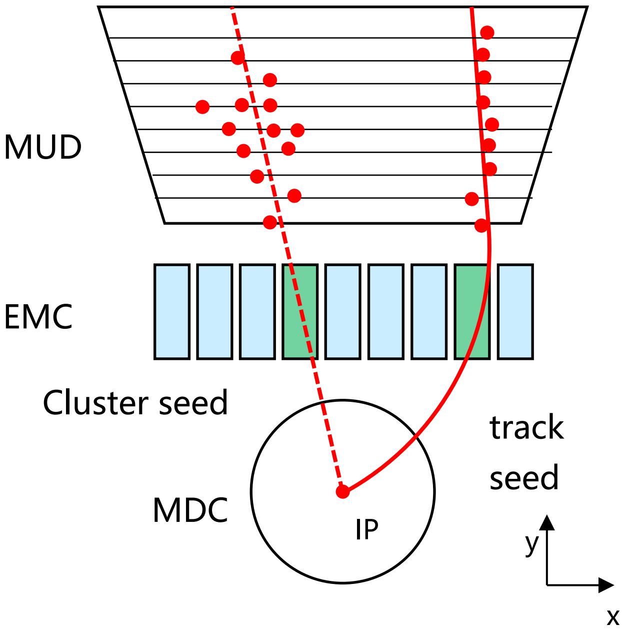


$\pi@\ P_{IP} = 2GeV/c, \theta_{IP} = 25°$

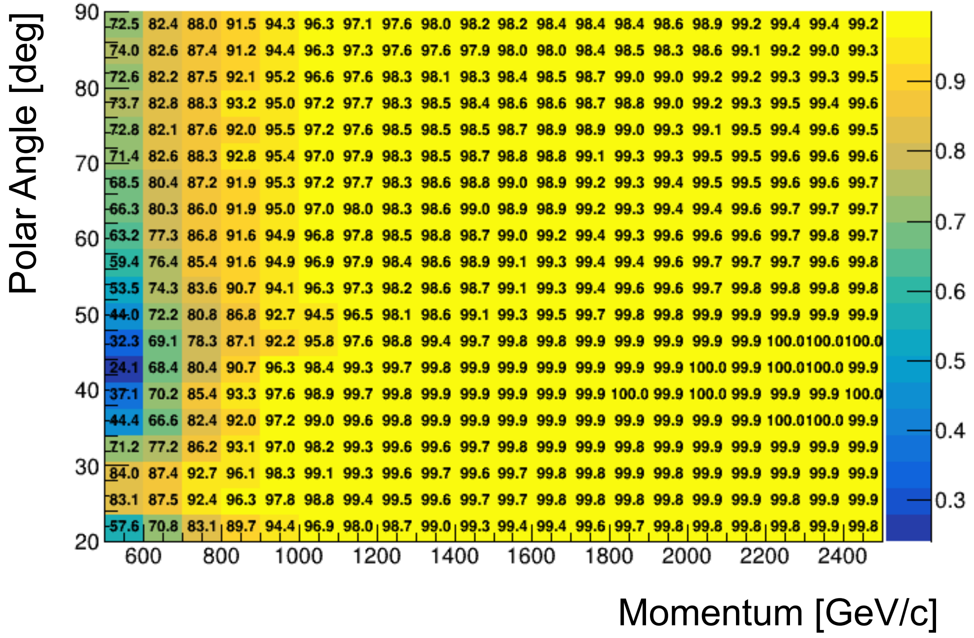
More details in Qingyuan Huang's talk

μ/π PID performance (MUD)

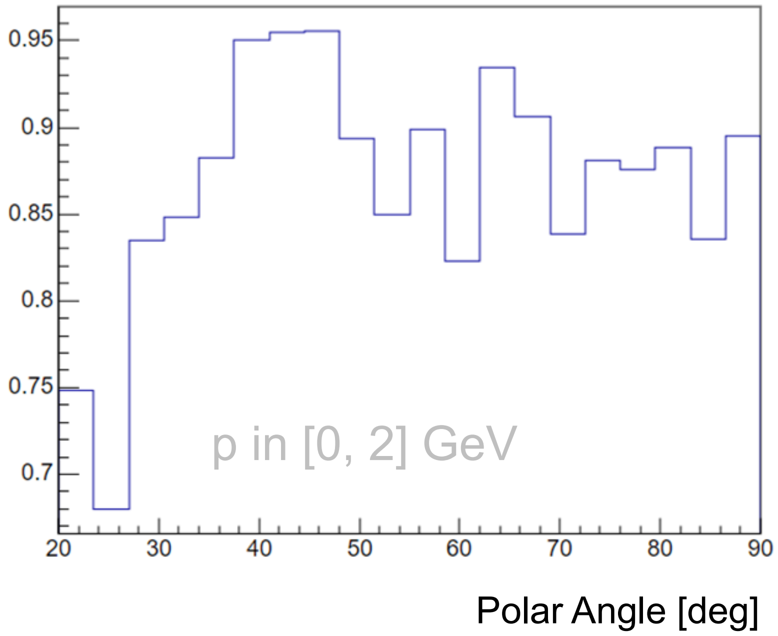
- **BDT-based algorithm** is developed to obtain μ/π and $(n, K_L)/\gamma$ identfication efficiency using features from tracker, EMC and MUD reco, with latest optimized MUD geometry
  - μ efficiency is above **95%** @ π suppression = 30 with momentum above **1 GeV**
  - (n, K_L) efficiency is **85%~95%** @ γ suppression = 30 in most polar angle (under optimization)



μ/π PID efficiency



(n, KL)/γ PID efficiency

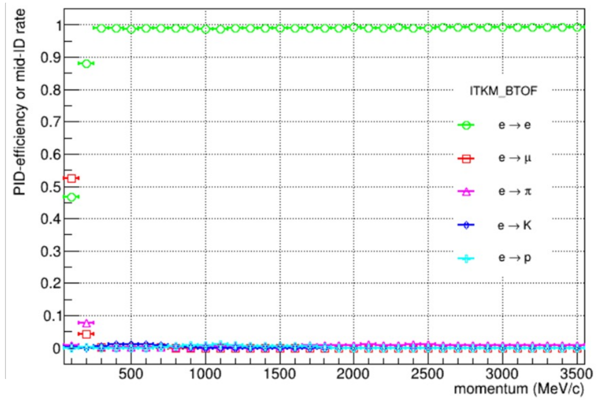


More details in Yulin Liu's talk

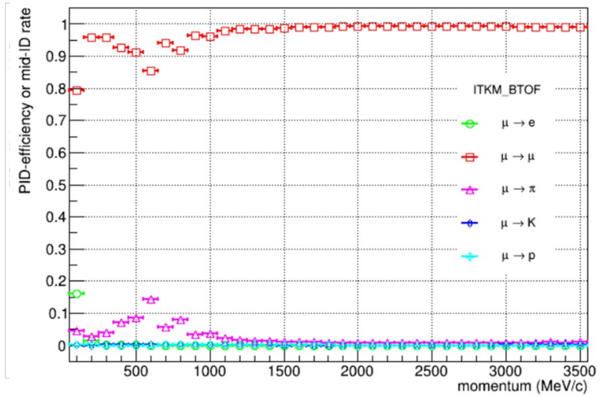
Traditional global PID performance

- A **combined likelihood with weight** was applied to integrate all PID information.
- Different PID information is assigned **different weights** in **different pairwise particle comparisons**.
  - $\mathcal{L}_h = \prod_{det} \mathcal{L}_{det,h}^{w_{det}}$
  - $w \propto \frac{\bar{\mu}_{\Delta logL}}{\bar{\sigma}^2_{\Delta logL}} \propto \frac{N_\sigma}{\bar{\sigma}_{\Delta logL}}$

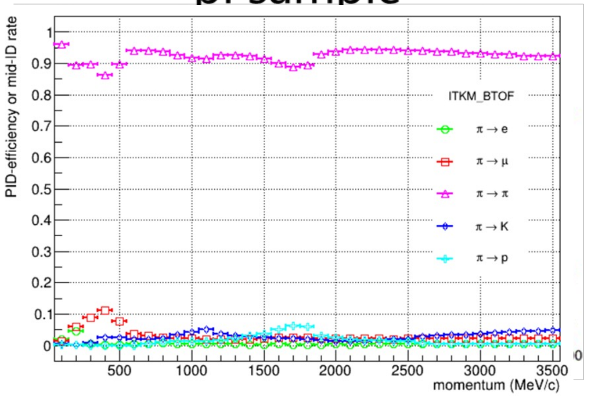
e sample



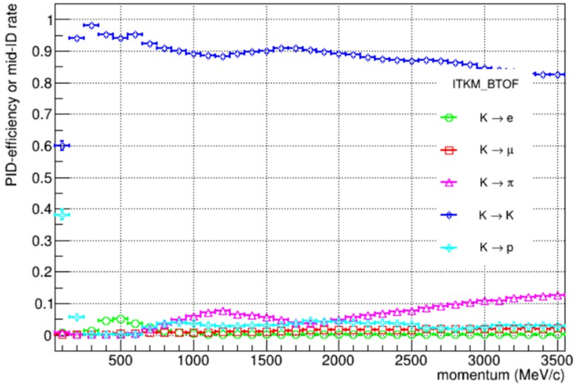
mu sample



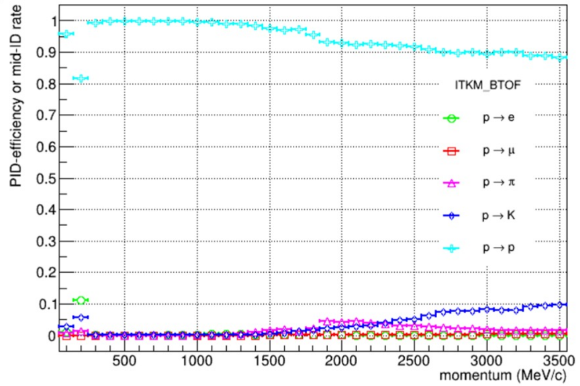
pi sample



K sample

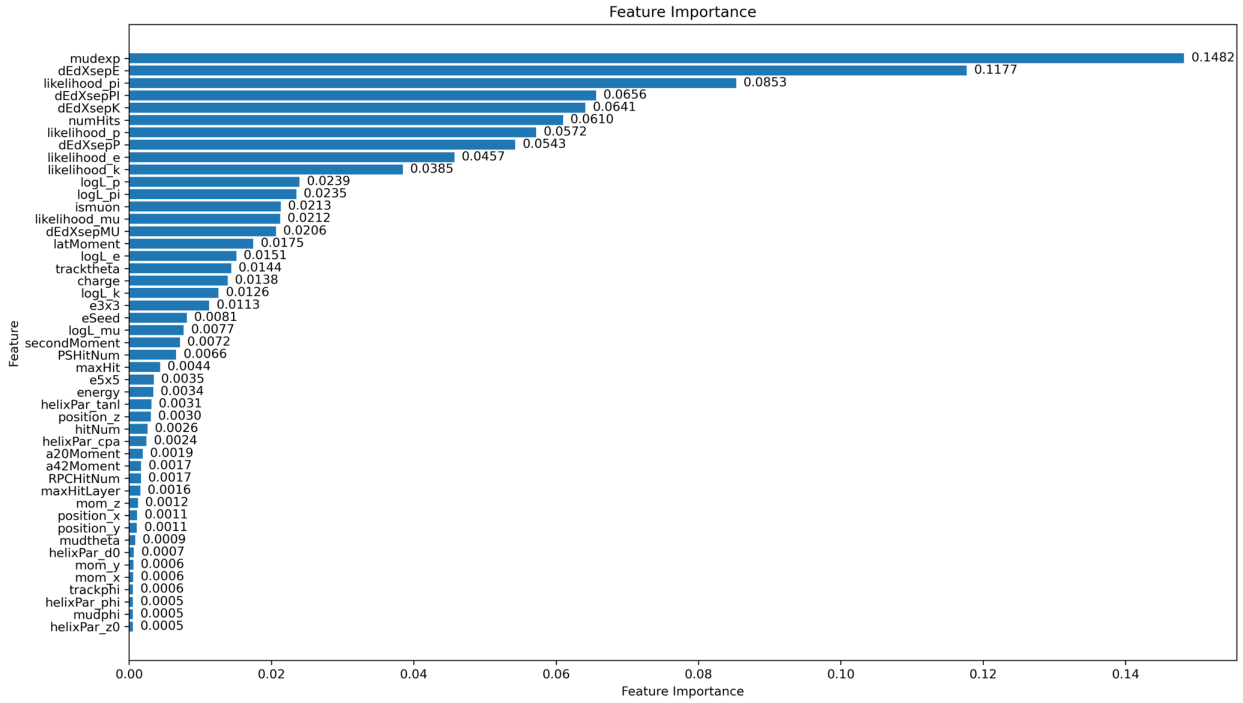


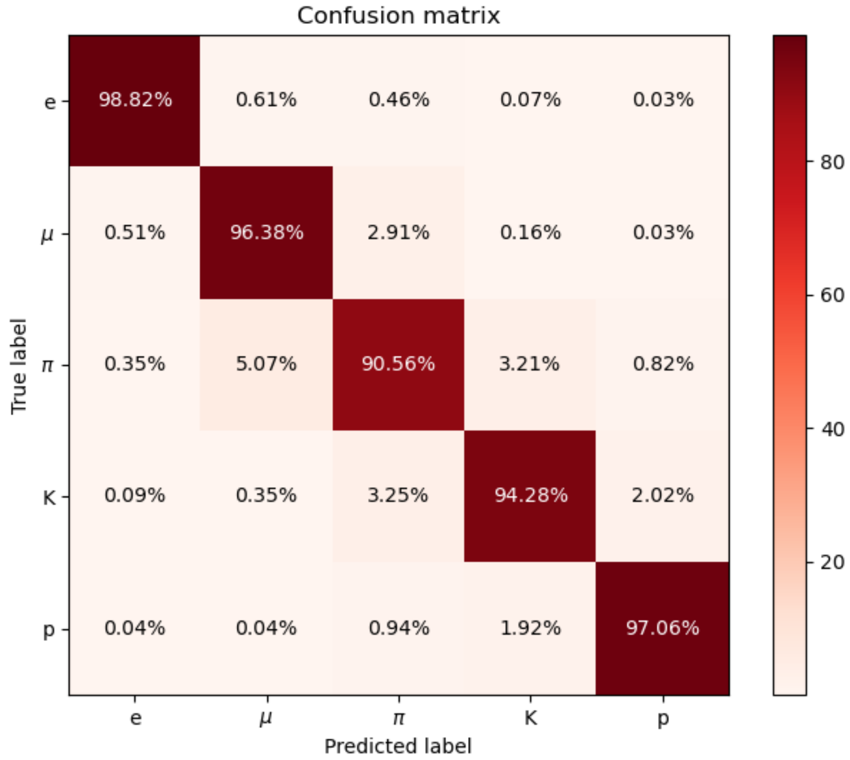
p sample



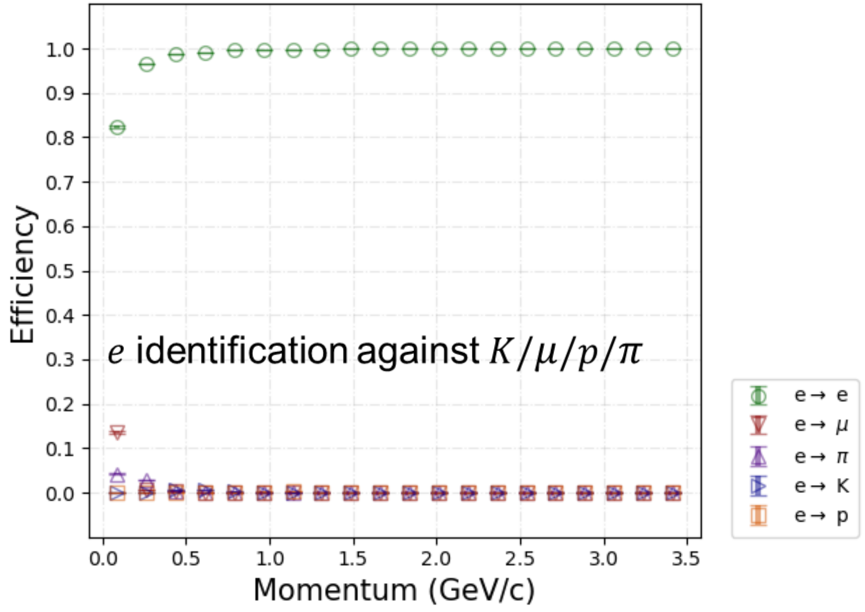
BDT-based global PID performance

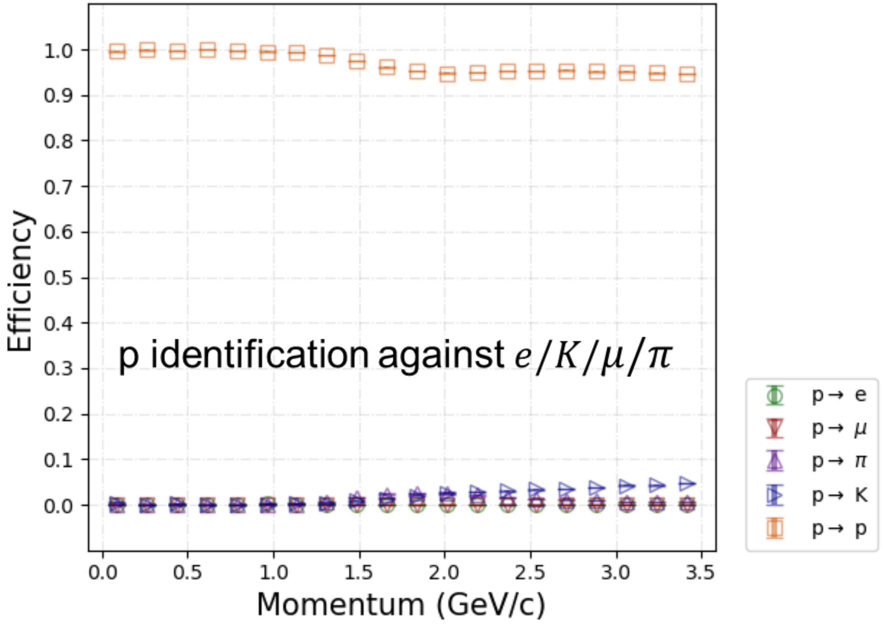
- Combining Tracker/dEdX/BTOF/DTOF/ECAL/MUD rec info (**47 features**) using BDT
  - ~100% efficiency for e (momentum above 0.5GeV) and μ (momentum above 1.3 GeV)
  - >95% efficiency for proton (~100% with momentum below 1.5GeV)
  - >90% efficiency for π and K in most regions

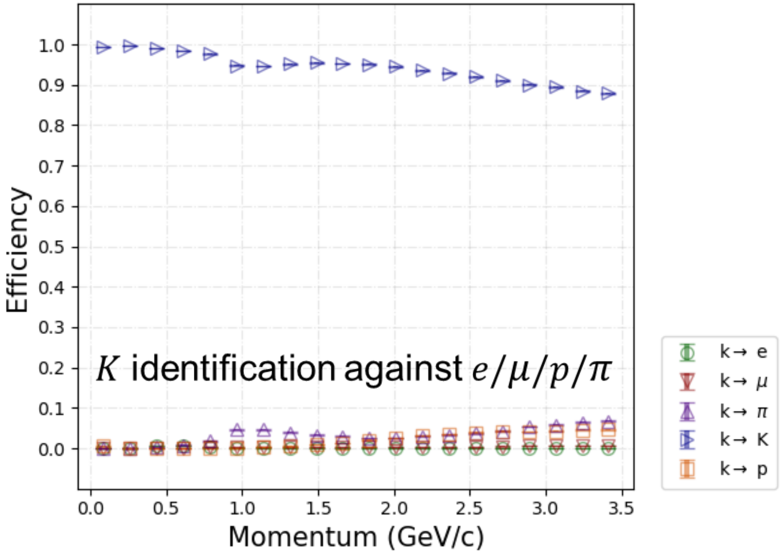


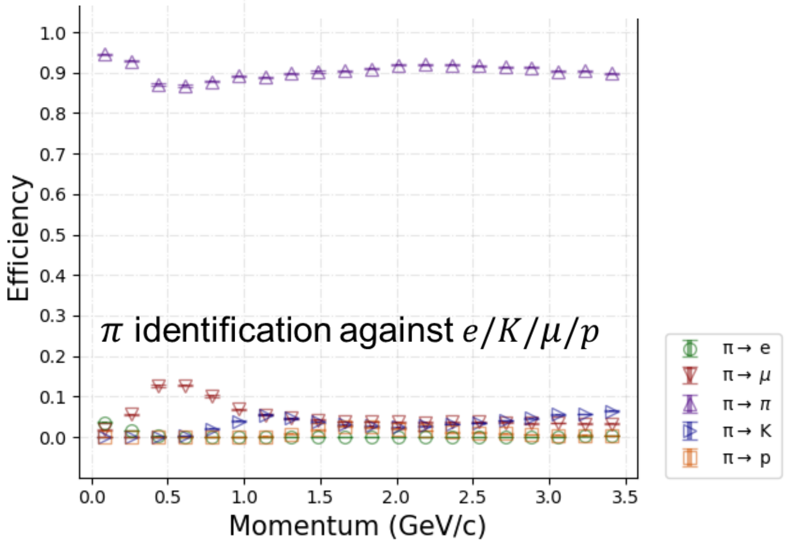


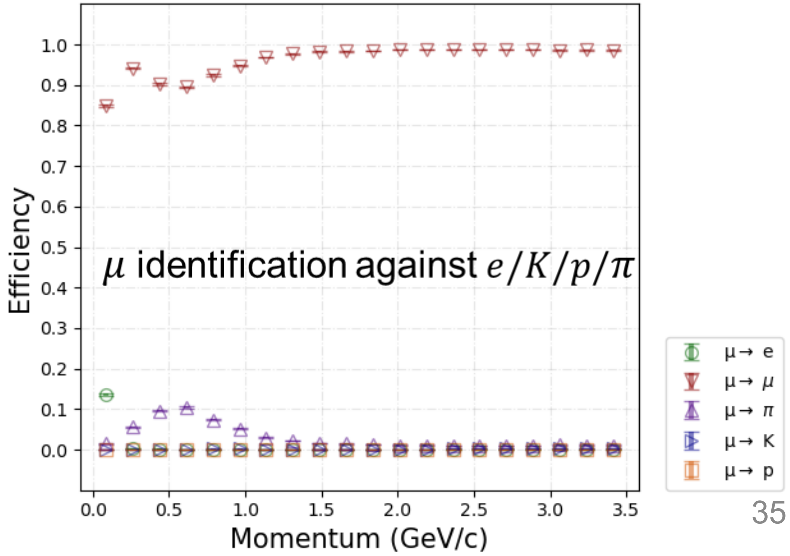
BDT-based global PID performance









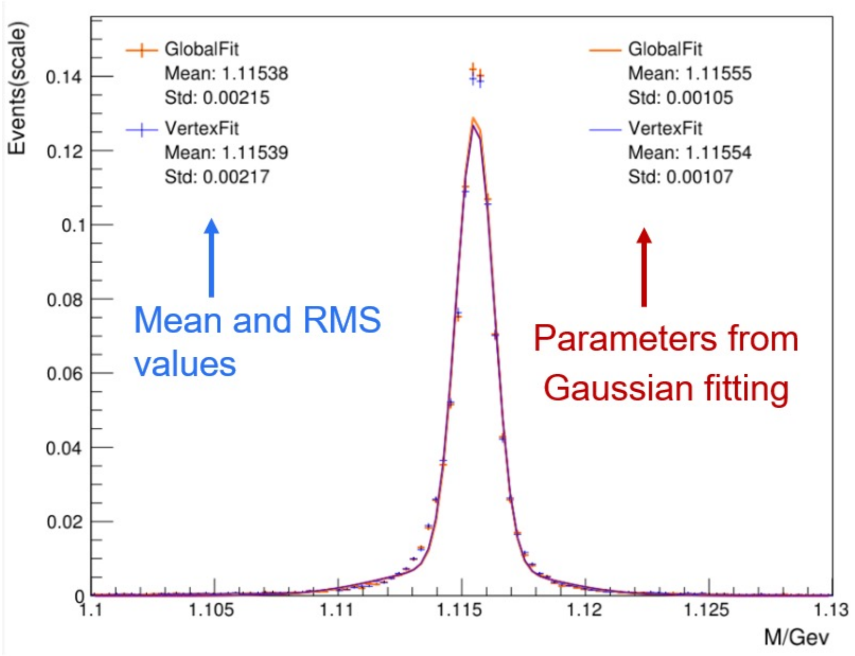


Analysis tools

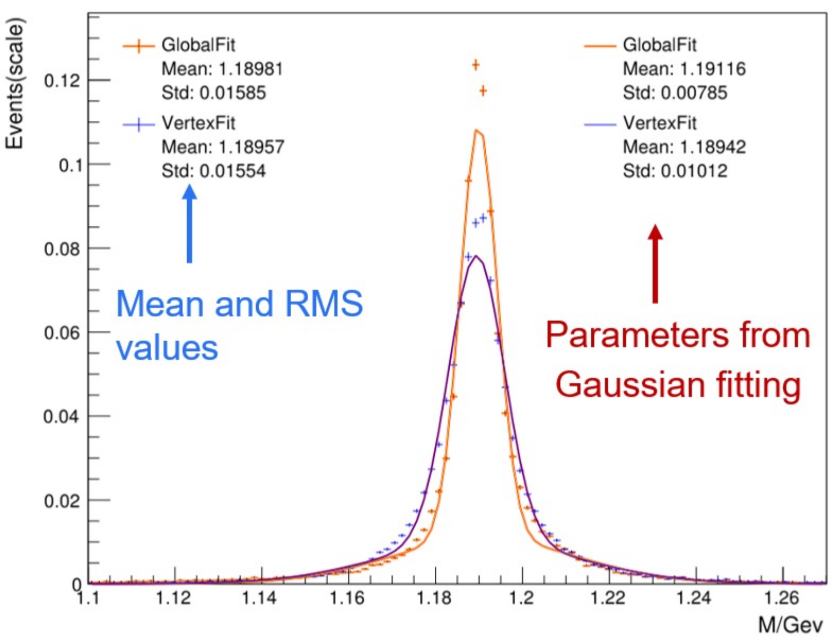
Vertex fit performance

- Designed and developed the **GlobalFit** package for STCF based on the global vertex fitting algorithm of **Belle II**, showing better performance than previous **vertex fit** package (i.e. **VertexFit**) imported from **BESIII**
  - Slightly better mass resolution $\Lambda/\bar{\Lambda}$ for and 1% higher efficiency
  - Improved $\Sigma/\bar{\Sigma}$ resolution by including constraint for decayed photon vertex in the fit

$J/\psi \rightarrow \Lambda\bar{\Lambda} \rightarrow (p\pi^-)(\bar{p}\pi^+)$



$J/\psi \rightarrow \Sigma^+(Pr)\bar{\Sigma}(\pi_0(rr)\bar{P})$



DTag migration

- The DecayChain-based DTagAlg and the dedicated DTagTool have been migrated to STCF
- Optimization according to the STCF event data model
- Consistency results between BESIII and STCF

$$\psi(3770) \rightarrow \boxed{D_0(K^-\pi^+\pi^+\pi^-)}\overline{D_0}(K_{s0}(\pi^+\pi^-)\pi^+\pi^-\pi^0(\gamma\gamma))$$

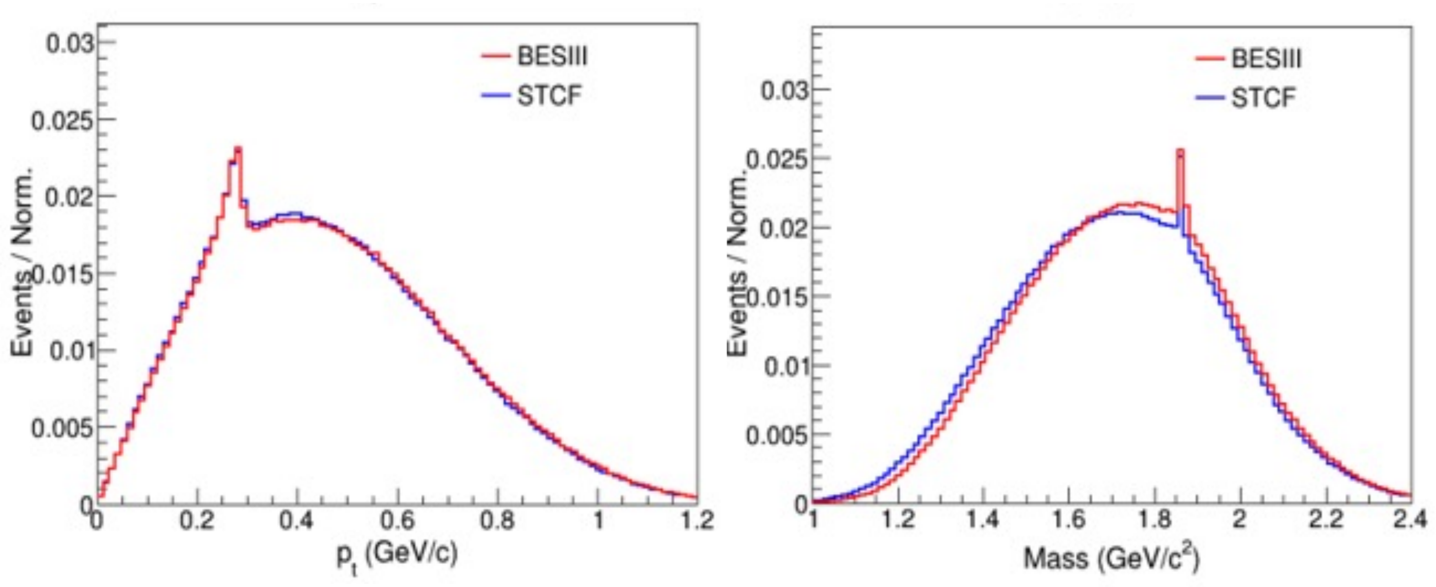


Figure. Transverse momentum and mass distributions in ψ(3770) center-of-mass system

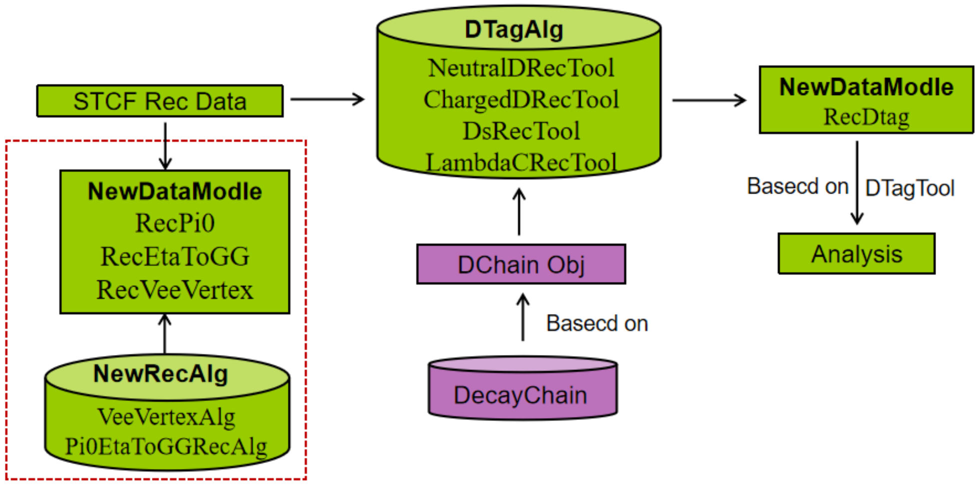
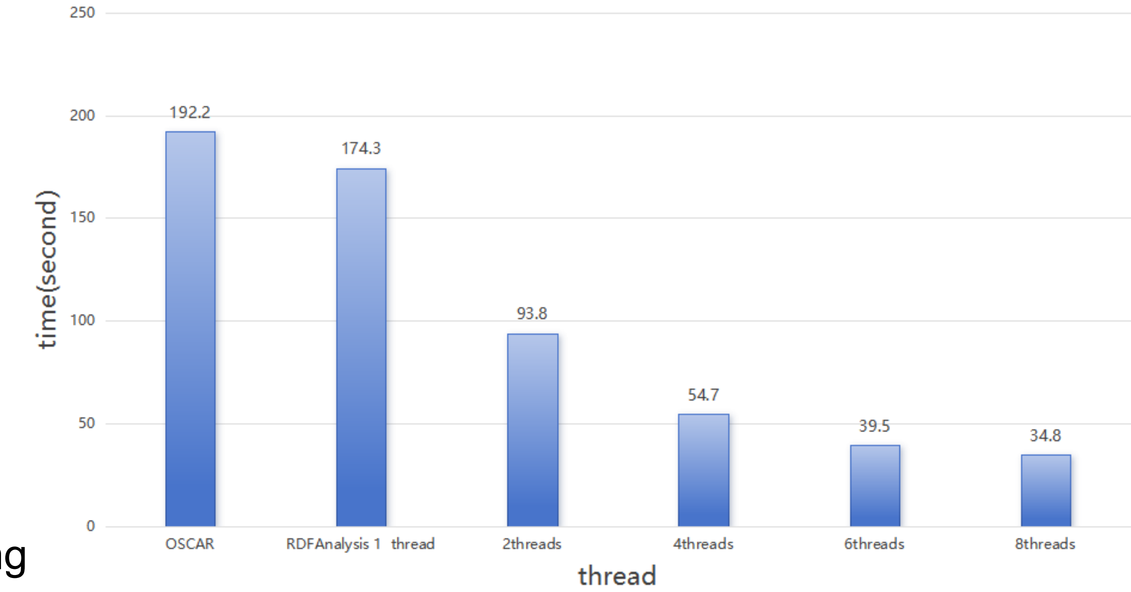
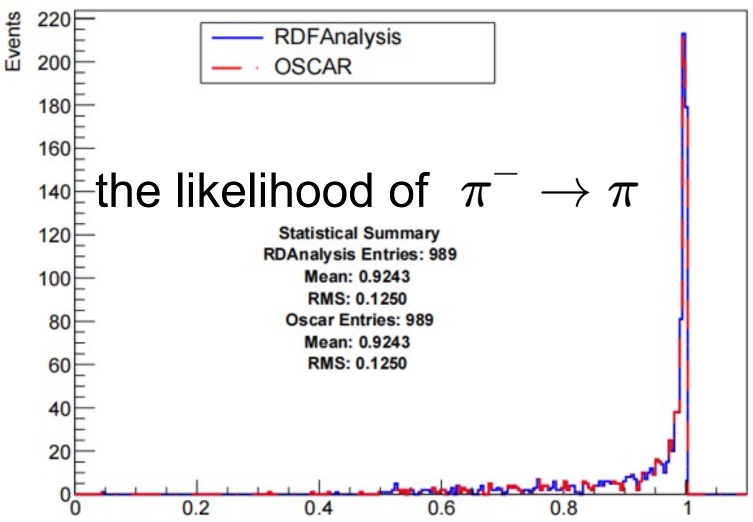


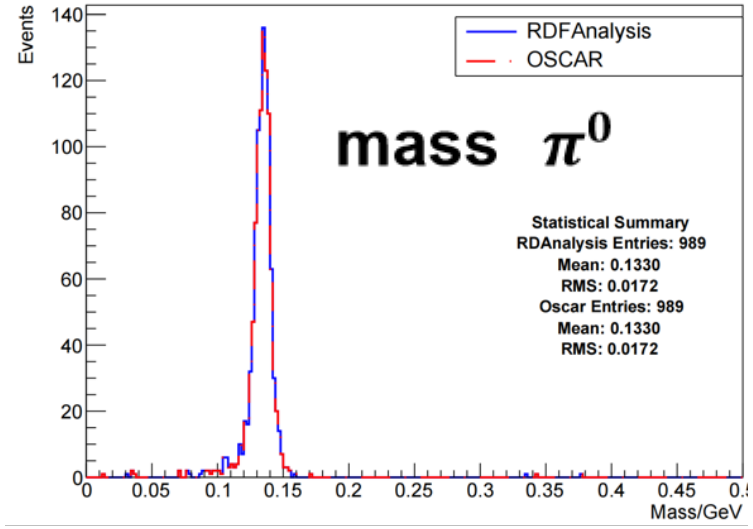
Figure. Structure of DTagAlg

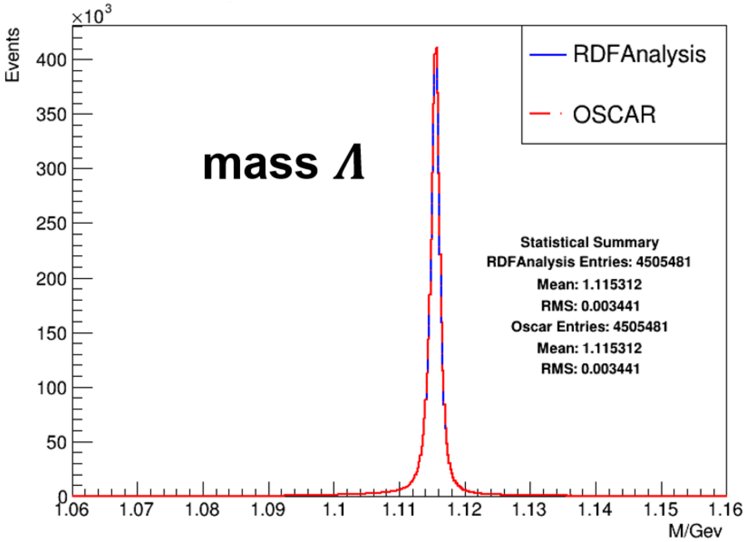
RDataFrame analysis

- RDataFrame-based analysis interface is developed to ease and speed up analysis in OSCAR
- GlobalPID is integrated into the framework and is validated by 2000 $J/\Psi \rightarrow \rho\pi^0 \rightarrow \pi^+\pi^-\gamma\gamma$
- The physics analysis outputs are validated using 10 million $J/\psi \rightarrow \Lambda\bar{\Lambda}$ events
- CPU performance analysis for 0.2 million events
  - RDataFrame-based anlaysis shows better performance utilizing multicores







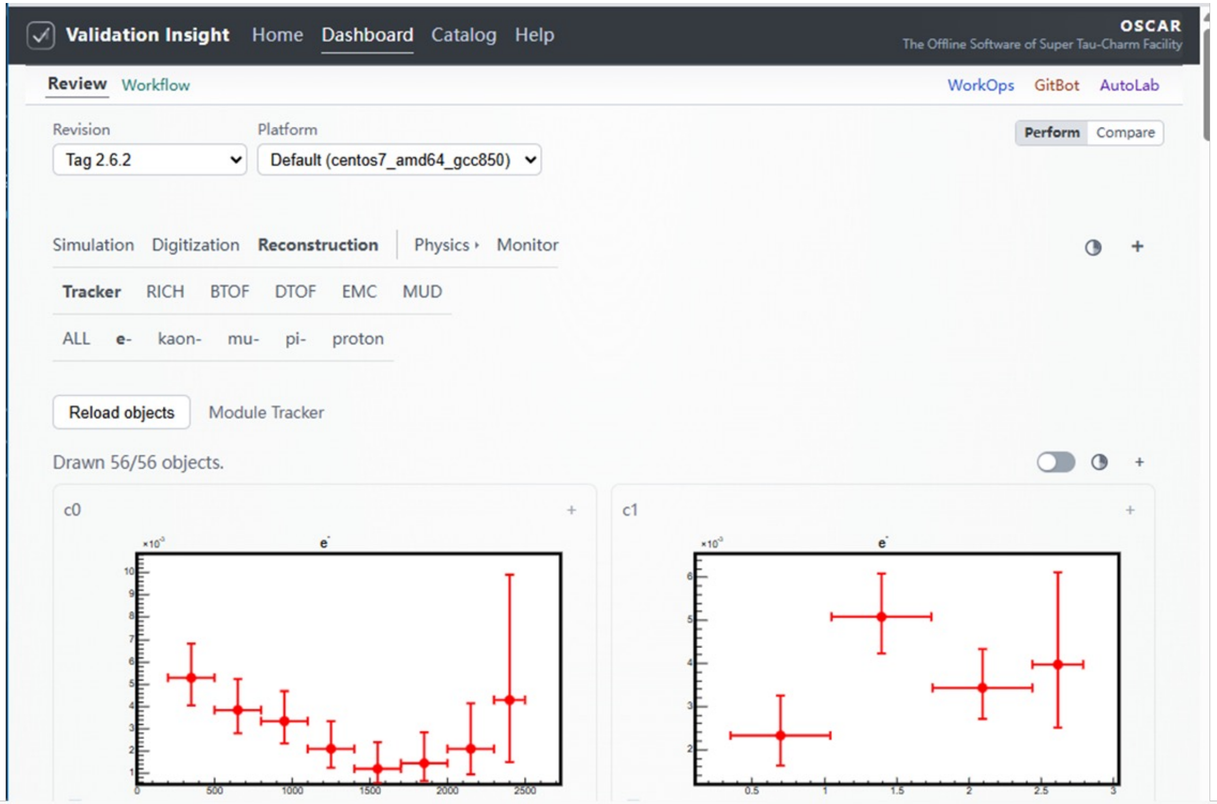


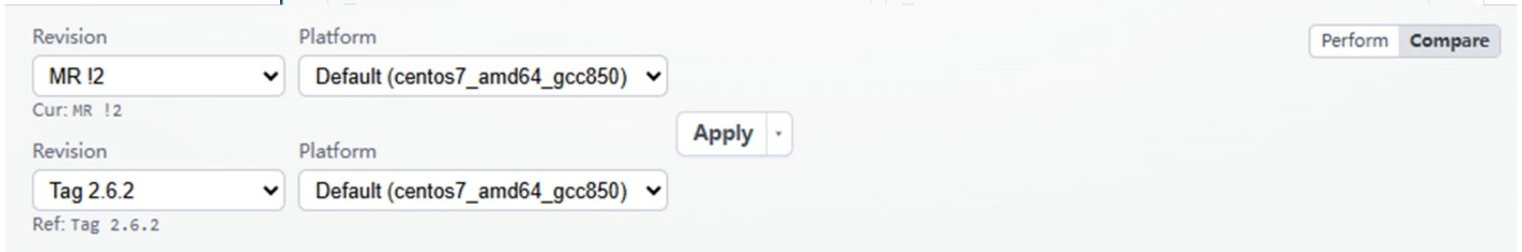
More details in [Ying Yang's talk]

Software validation

Validation System

- Automated validation system developed for OSCAR — provide full-chain of physics validation and performance profiling.
- Provide web-portal to review results and compare different releases.
- Developed AI-assisted plot review and batch screening of validation results.
- Fully linked with Gitlab to support
- AI-assisted code review and validation

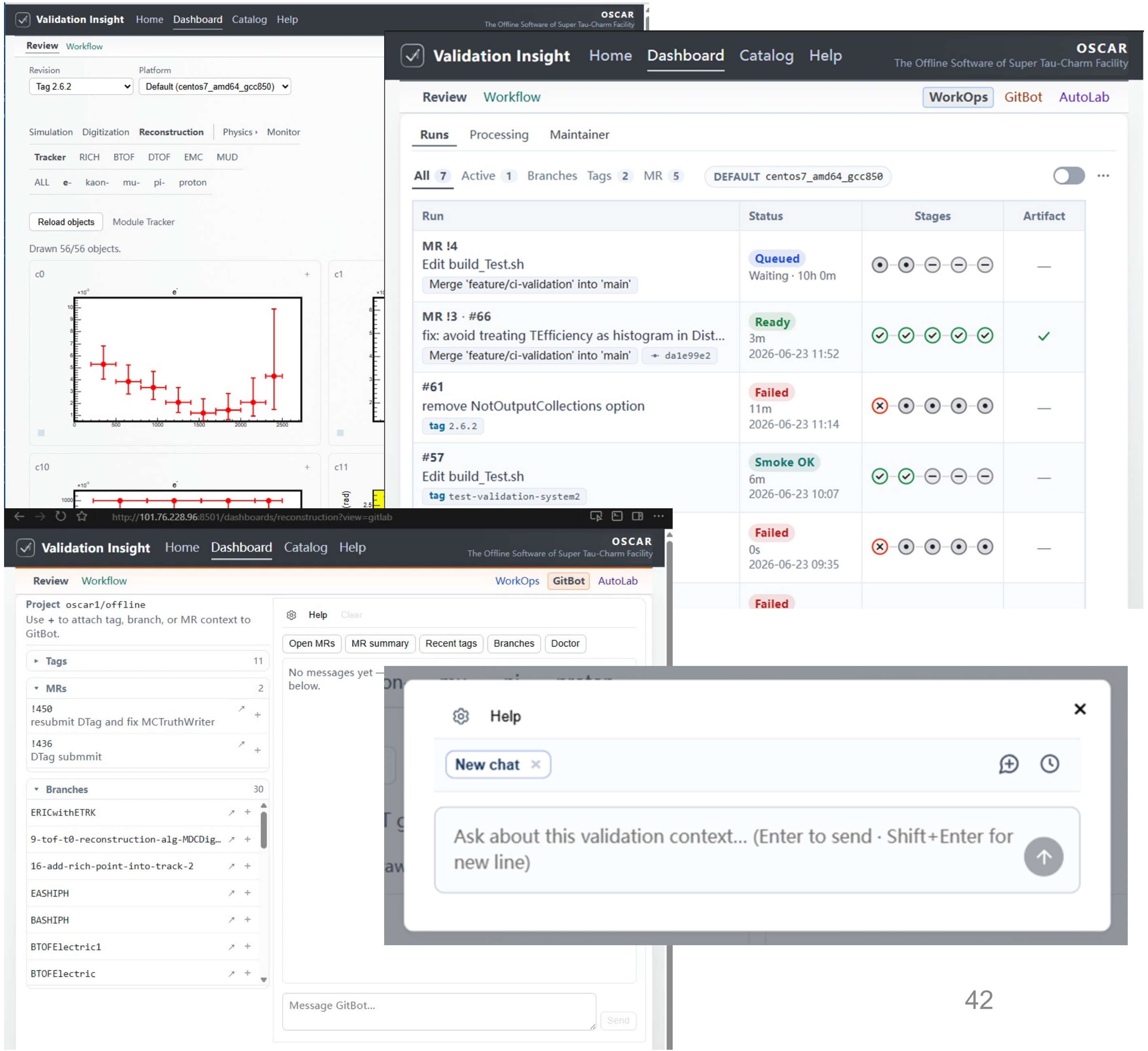




Web Dashboard Functionalities

- Browse and compare validation performance plots (between different releases).
- Track validation run status.
- Code review
- AI Interpretation:
  - AI-assisted interpretation of plots in validation context.
  - MR comments and software ref lookup on GitLab.
  - Batch analysis: Batch digest and grading of validation plots.

Experiment-agnostic: can be easily ported to other experiments



Summary

STCF offline software status

Tasks				ITKW	ITKM	MDC	BTOF	RICH	DTOF	EMC	MUD
Simulation											
Bkg mixing											
Digitization											
Reconstruction	Charged tracks										
	dE/dx										
	Photon										
	T0										
PID	sub-detector PID										
	global PID	ML-based	Charged								
			Neutral								
		Likelihood-based	Charged								
			Neutral								
Analysis	Vertex fit, Kinematic fit										
Detector optimization											
Event display											
Calibration&Alignment (dedicated group formed recently)											
CPU&disk optimization											

- good state
- under optimization
- Under development
- Irrelevant

Physics simulation using OSCAR

Physics goal	Processes	Contributors
Baryon CPV and EDM	$J/\psi \to \Lambda\bar{\Lambda}$	Mingyu Yu (SDU), Jianyu Zhang (UCAS)
	$\Xi^+\bar{\Xi}^-$	Yue Xu (FDU)
τ physics (QE, EDM, rare decay)	$\tau \to \pi(\pi\pi^0)\nu$	Mingyi Liu et al (USTC)
	$\tau \to 3\mu$	Yanlin Liu et al (SDU)
	Radiative decay	Haolin Li (ZZU)
Charm mixing, strong phase	C-even correlated $D\bar{D}$	Yinghao Wang (LZU, UCAS) Qiang Lan (USC)
CPV in Λ_c^+ decays	$\Lambda K^+, \Sigma^0 K^+, \Sigma^+ K_S^0$	Hongjian Wang (LZU)
CP and CPT in $K_S^0 \to \pi^+\pi^-$ decay	$J/\psi \to K_S^0 K\pi$	Huibin Zhang (UCAS)
CLFV decay	$J/\psi \to \gamma l l'$	Zhibo Tao (JLU)
	$D^0 \to \bar{\Lambda} + ALP$	Xinru Tang (USC)
Two-photon physics	$e^+e^- \to e^+e^- hadrons$	Jie Wang (CUG)
	$\eta_c \to \gamma\gamma$	Xinmei Jing (UCAS)
Tau mass	$e^+e^- \to \tau^+\tau^-$ near threshold	Hailin Song (USTC)
Strong phase	$D^0 \to K^-\pi^+\pi^0$	Cheng Wang (LZU)
α_s	$e^+e^- \to \tau^+\tau^-, \tau \to l\nu\nu, \tau \to \pi\pi^0\nu_\tau$	Shanshan Li (HNU)

Topic	Processes	Contributors
Charmonium-like, tetra- and penta-quarks	$e^+e^- \to X(3872) \to \pi\pi J/\psi$	Nan Lu et al (USTC)
	$e^+e^- \to P_c\bar{p} \to J/\psi p\bar{p}$	Zhuojuan Dong (SYSU)
	$e^+e^- \to J/\psi\eta_c$	Zikang Chen (SYSU)
Collins FFs	$e^+e^- \to KK, \pi\pi, K\pi$	Teng Li (SDU)
Hadron EMFFs	$e^+e^- \to p\bar{p}$	Wenhui Tian (SYSU)
	$e^+e^- \to n\bar{n}$	Yuhe Huang (USTC), Kai Liu (LZU)
From factors in hyperon decay	$\Sigma^+ \to \gamma p$	Zheyan Zhou (USTC)
	$\Lambda \to pe\nu$	Junxian Zhou (FDU)
	$\Xi^- \to \Lambda e\nu$	Xu Han (USTC)
	$\Sigma^+ \to pee$	Xiaoshuai Qin (SDU)
CKM matrix	$D \to \pi\mu\nu$	Qingyuan Huang (UCAS)
	$D \to \mu\nu$	Jiahui Qiao (HNU)
	$D \to inclusive$	Jiayuan Su, Hongrui Liu (HUST)
CKM matrix	$D \to \pi\mu\nu$	Qingyuan Huang (UCAS)

Summary

- STCF offline software has much matured since last conference providing **full functionalities of simulation + digitization + reconstruction + analysis + AI-supported performance validation**
  - Supporting tens of **physics feasibility studies** and **detector optimization** towards STCF TDR
  - Efforts for detector calibration and alignment has started
- Both traditional and **ML-based algorithms** are deployed for tracking and PID (and fast EMC simulation)
  - Tracking, photon and PID performance can fulfill CDR requirements
  - Still room for improvement in difficult phase space (e.g. low p_T/displaced tracks), and in CPU consumption
- Developed RDataFrame-based analysis to facilitate fast physics analysis
- The performance with optimized detector geometry is being finalized

Thanks for your attentions!

Many thanks to all people of the software group!
And many thanks to the physics simulation group for the inputs and feedbacks!