Status of STCF Offline Software

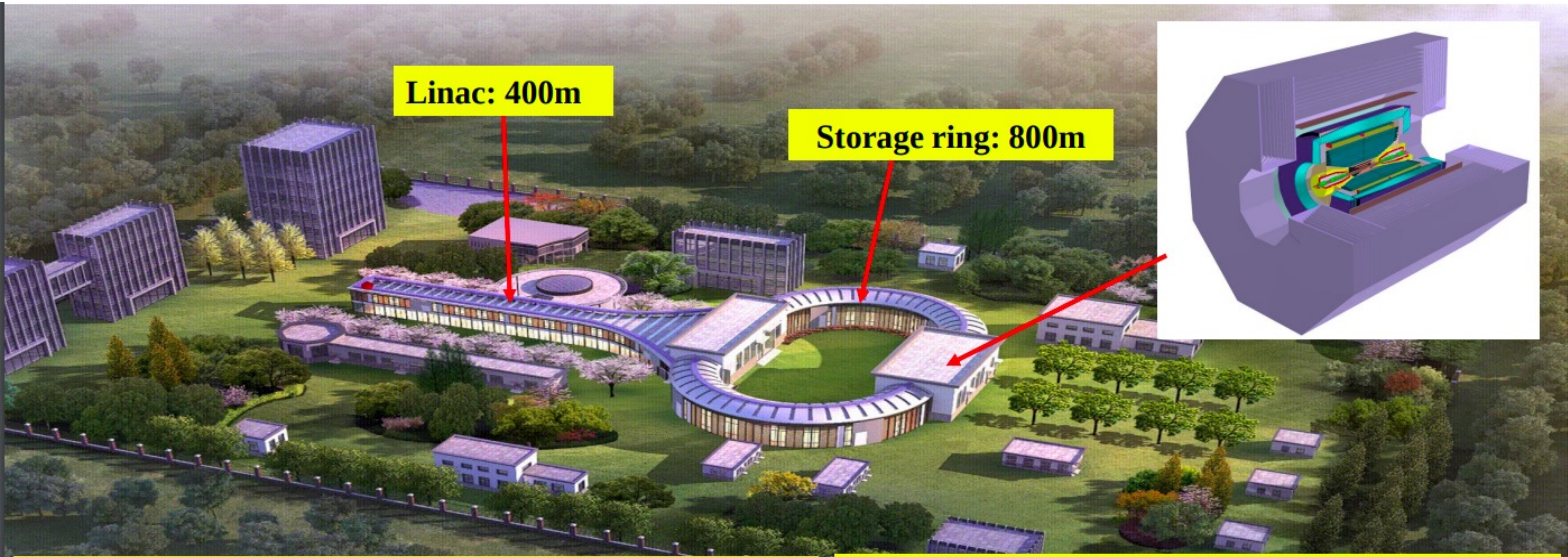
Xingtao Huang[1] , Xiaocong Ai [2]

[1] Shandong University,
[2] Zhengzhou University

On behalf of the STCF Software group

The Super Tau Charm Facility

The future Super Tau Charm Facility



- Peak luminosity $>0.5\times10^{35}$ $cm^{-2}s^{-1}$ at 4 GeV
- Energy range E_{cm} = 2-7 GeV
- Potential to increase luminosity & realize beam polarization
- Total cost: 4.5B RMB

- 1 ab^{-1} data expected per year
- Rich physics program, unique for physics with *c* quark and τ leptons
- Important playground for study of QCD, exotic hadrons, flavor and search for new physics.

Physics requirements

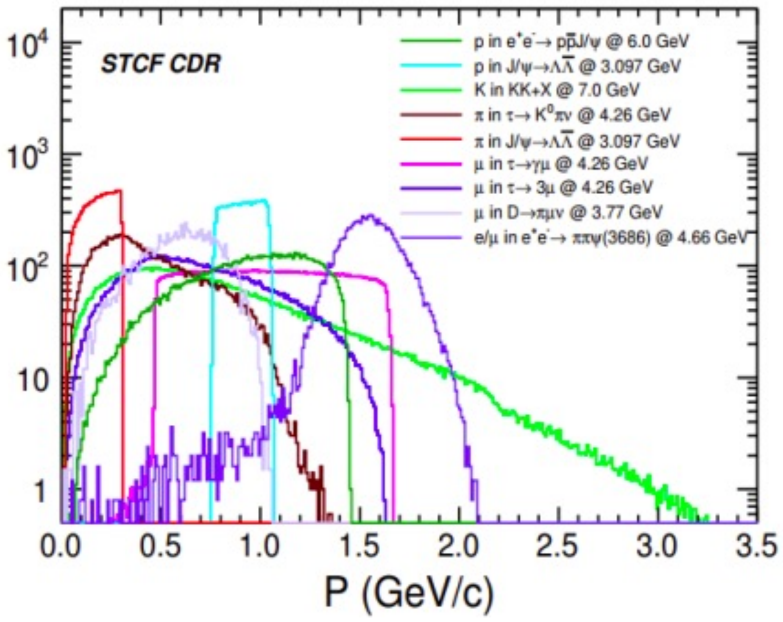
- To achieve physics goals: higher event rate (400k Hz) and background
  - Good efficiency and resolution for tracks (photon)
  - Powerful PID between charged (neutral) particles

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 $\sigma_{\rm pos} \approx 5$ 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	$\sigma_T = \frac{300}{\sqrt{p^3({\rm GeV}^3)}}$ ps

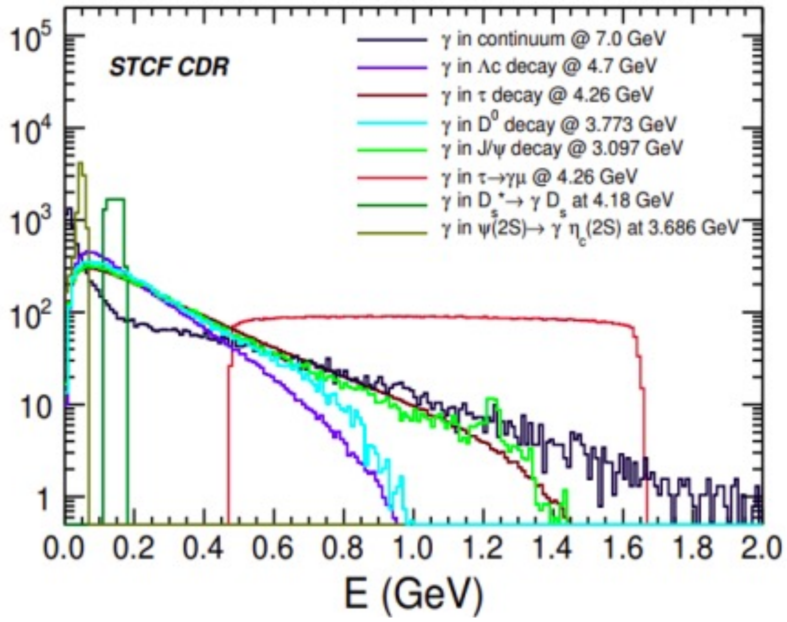
Particles to detect at STCF

- ❖ Charged particles
  - e, μ, K, π, proton (p up to 3.5 GeV, most with p < 2 GeV, lots of particles with p < 400 MeV)
- ❖ Neutral particles
  - γ (energy coverage from 25 MeV to 3.5 GeV)
  - KL, neutron (up to 1.6 GeV)

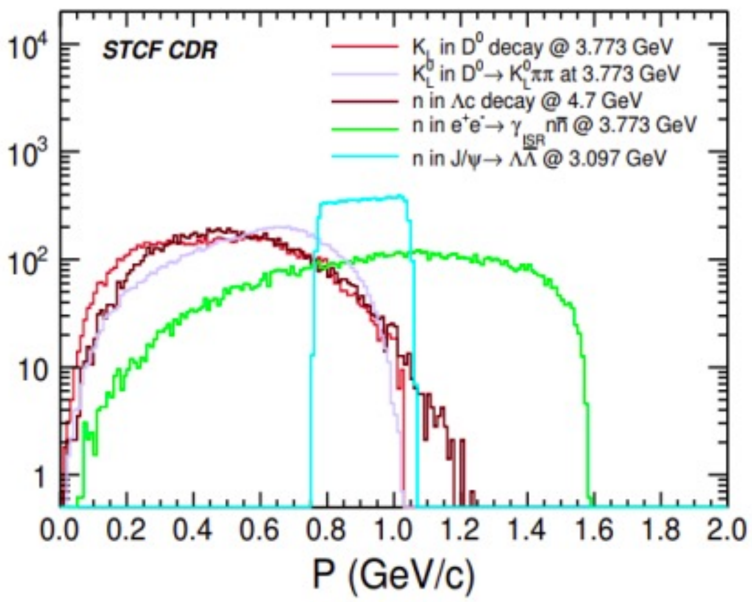
Charged particle momentum



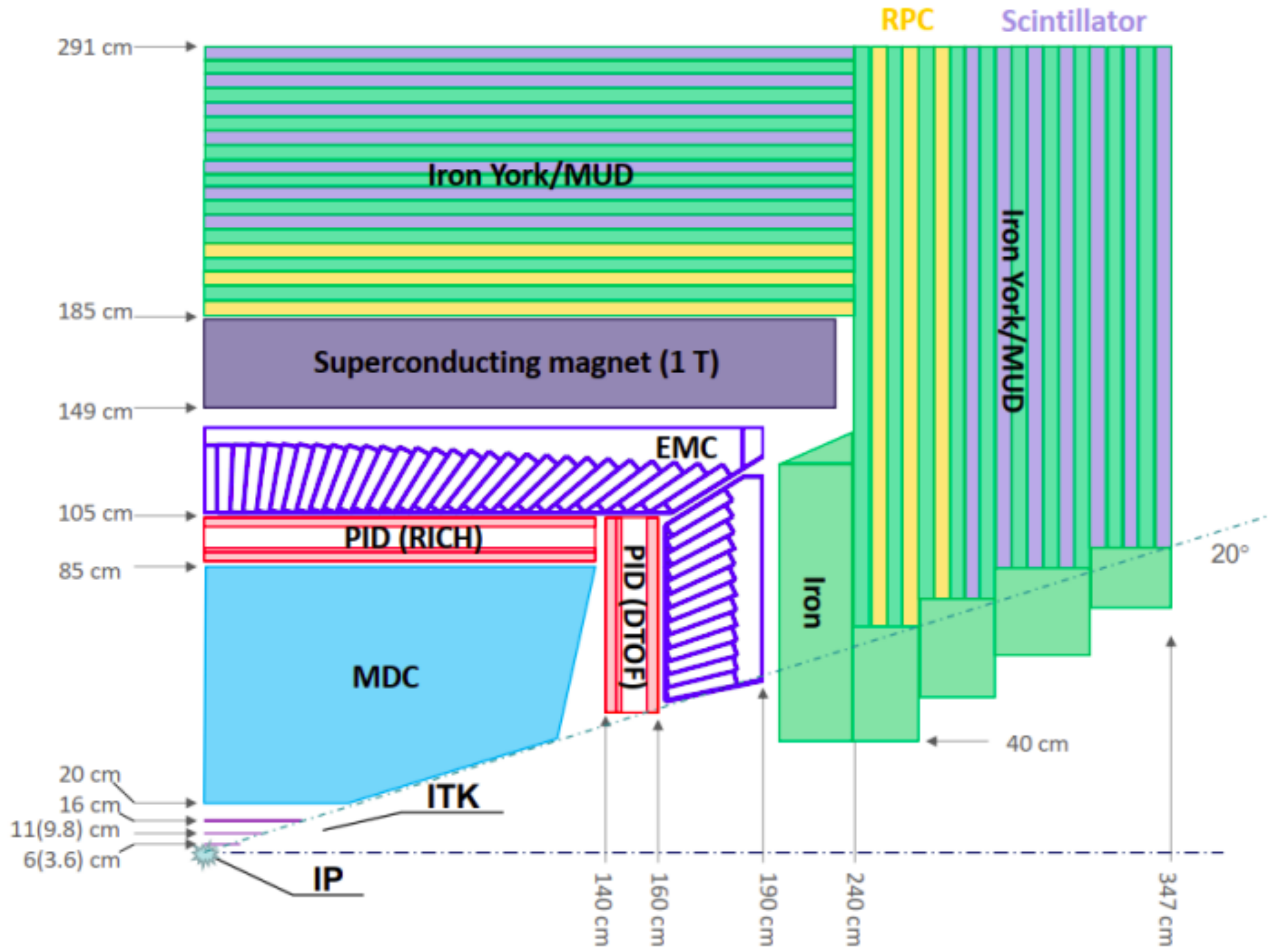
Photon energy



KL and neutron momentum



STCF detector system



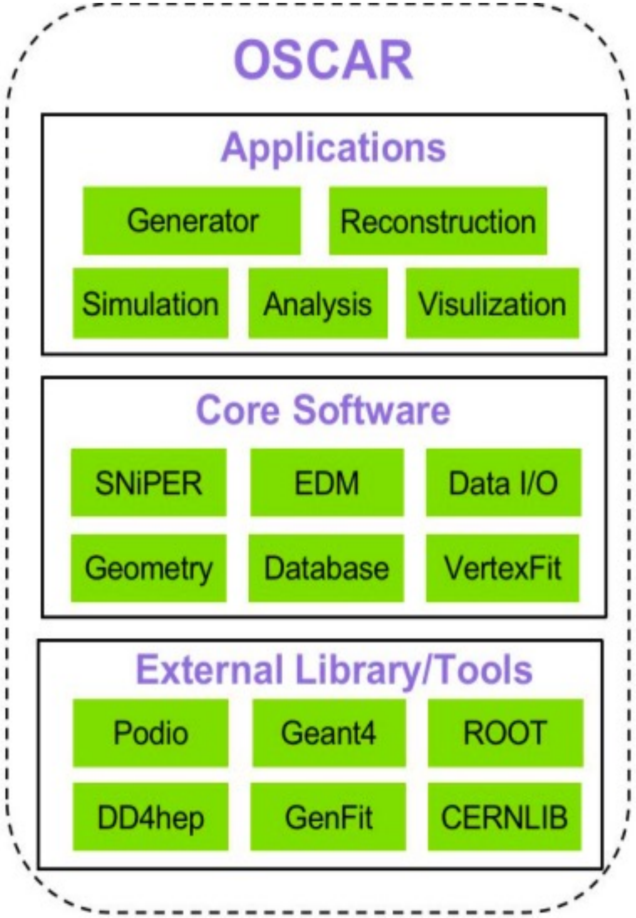
Performance requirements

- ITK:
  - Material < $0.01X_0$, σ_{xy} < 100 μm
- MDC:
  - Material < 0.05 X_0
  - σ_{xy} < 130 μm, $\sigma(p)/p$ < 0.5% at 1 GeV/c
  - dE/dx resolution < 6%
- PID:
  - $3\,\sigma$ π/K separation, PID efficiency > 97% up to 2 GeV
- EMC:
  - σ_E < 2.5%, σ_{pos} ~ 4 mm, σ_t ~ 300 ps at 1 GeV
- MUD:
  - μ efficiency > 95% above 0.7 GeV with $\pi \rightarrow \mu$ misidentification rate <3%

The Offline Software of STCF (OSCAR)

STCF Offline Software System

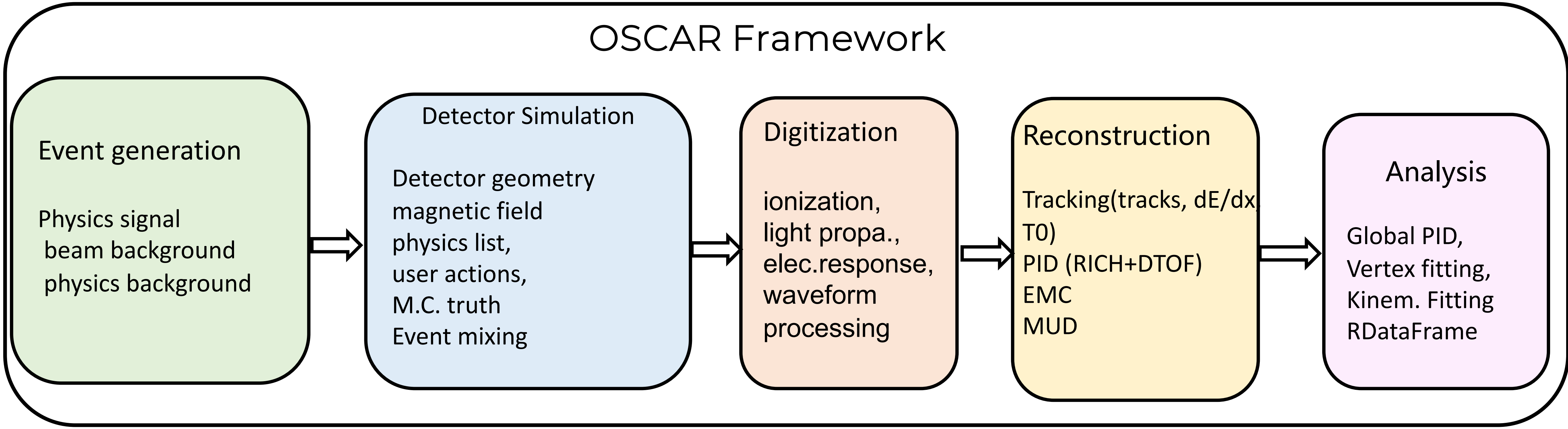
- The Offline Software of Super Tau-Charm Facility （OSCAR） is designed for detector design, MC data production and physics analysis at STCF：
  - Applications: STCF specific application software
  - Core software: common platform for the offline software
  - External libraries and tools
- Core software are developed for common functionalities
  - Event loop control (sequentially and concurrently)
  - Detector data and event data management
  - Common tools for data analysis
  - To support efficient parallel and heterogeneous computing
- Some applications are ported from BESIII



W.H. Huang et al 2023 JINST 18 P03004

Status of STCF Offline Software

- Dedicated Core Software System including the underlying framework, DM system, GMS etc. are developed
- Full chain of detector simulation and physics object reconstruction has been built
- Physics analysis with full detector simulation and reconstructed objects is supported
- Parallelized simulation and reconstruction based on MT-SNiPER under developing

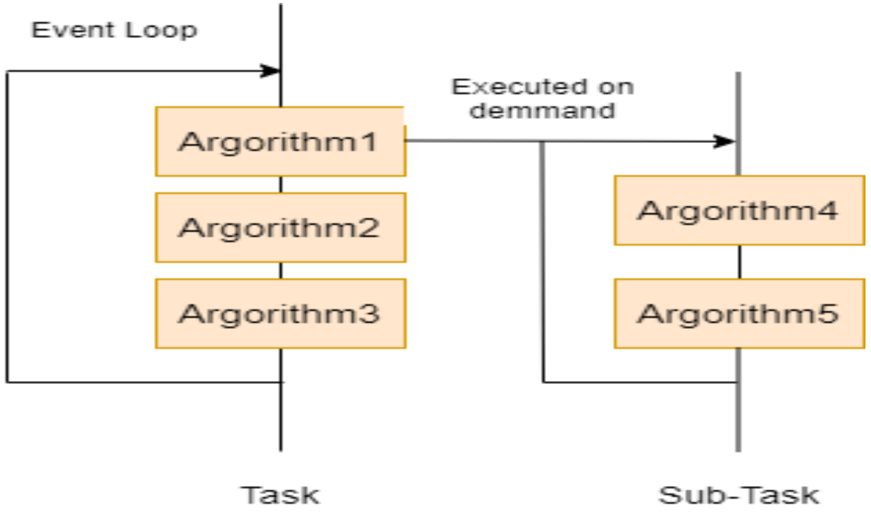


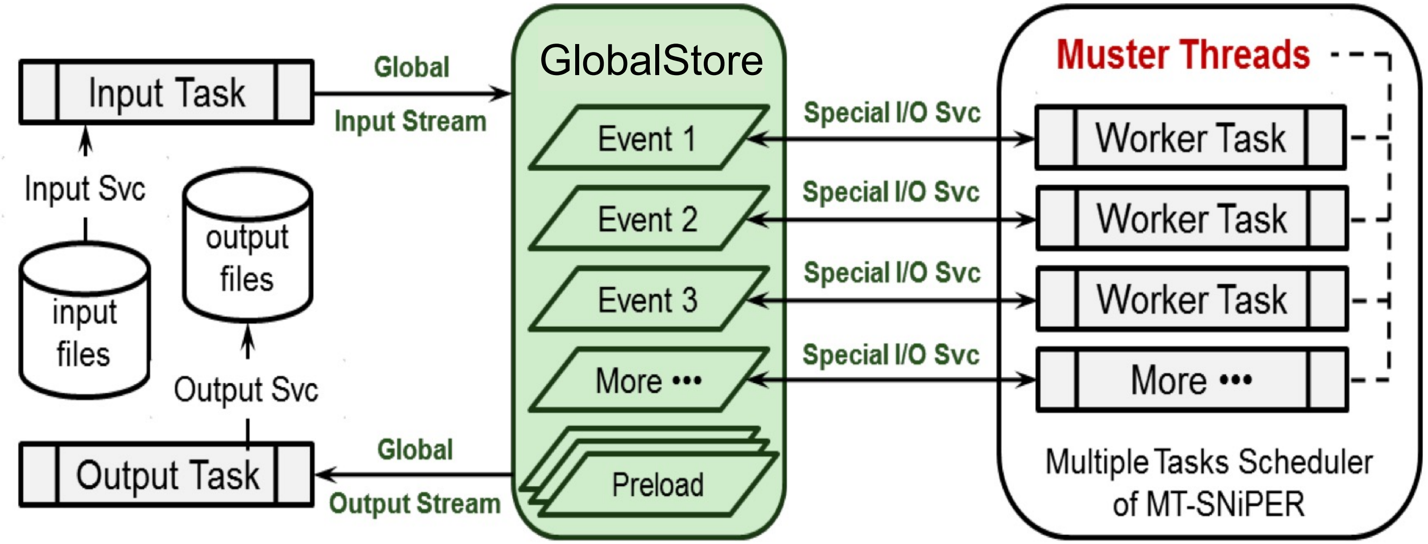
Core Software

Teng Li, Core Software of STCF,
Software & Computing Session, Jan. 17

Underlying framework: SNiPER

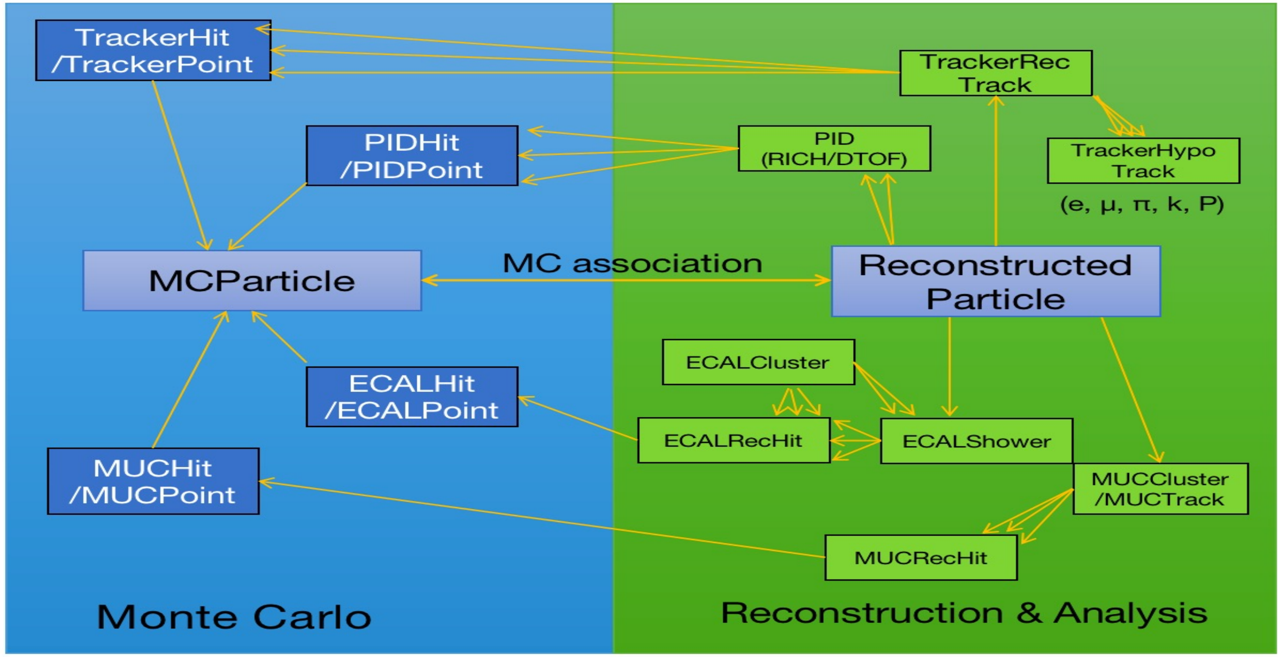
- Developed since 2012, very lightweight and flexible, supporting both non-collider experiments and collider experiments
- Providing basic functions of event loop control, application interface, job configuration, logging etc.
- Providing simple interfaces for building multi-threaded applications, thus supporting both serial and parallel event processing with extension of SNiPER, Muster
- Adopted by JUNO (neutrino), LHAASO (cosmic ray), nEXO (neutrinoless double beta decay) and HERD (dark matter)

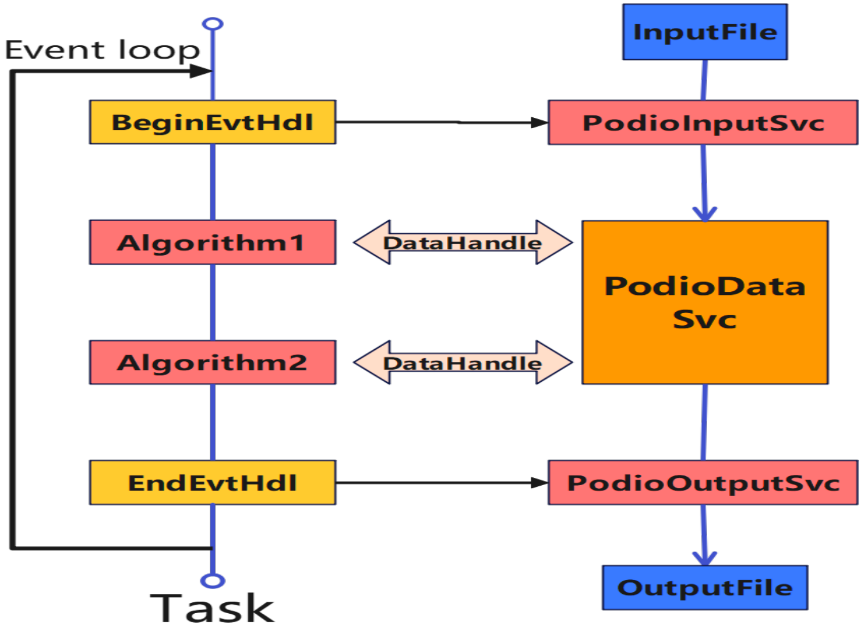




Event Data Model (EDM) and Serial Event Data Management

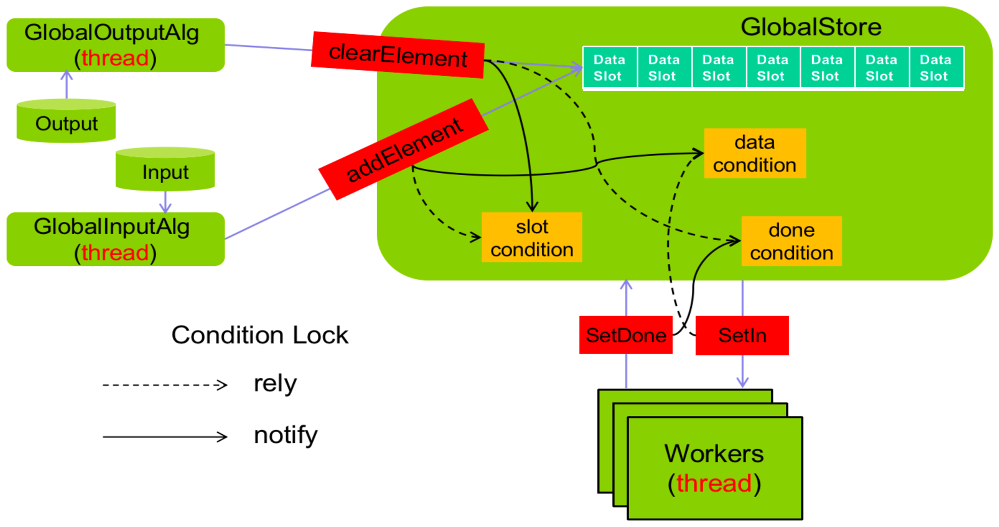
- Very crucial and taking significant effect on performance of the event processing
- Developed STCF EDM based on podio, which is also adopted by EDM4hep
  - Define event data and relationship between data object in YAML file
  - Re-use MCParticle and ReconstructedParticle from EDM4hep as the core index
  - Good support for multithreading
- Extended SNiPER data management system based on podio

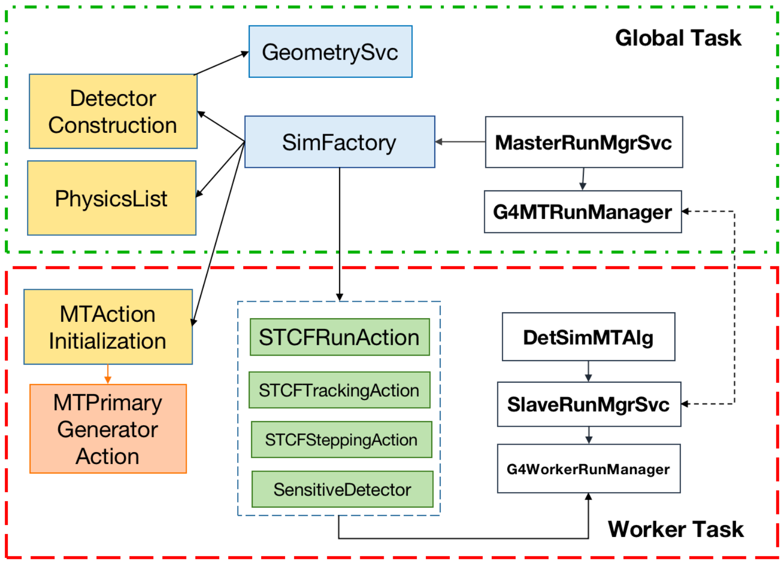


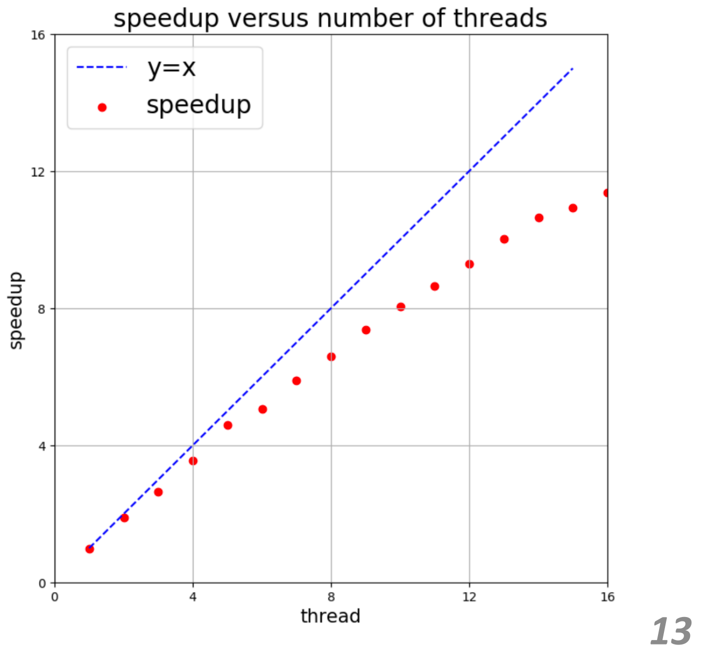


Parallelized Event Data Management

- Developed parallelized Event Data Management for parallelized event processing
  - GlobalStore to cache multiple events (each within one data slot)
  - Two dedicated threads for input and output respectively
  - SNiPER Muster (Multiple SNiPER Task Scheduler) works as a thread scheduler
- Parallelized detector simulation and reconstruction are developed
- Track-level and algorithm-level parallelism are under R&D

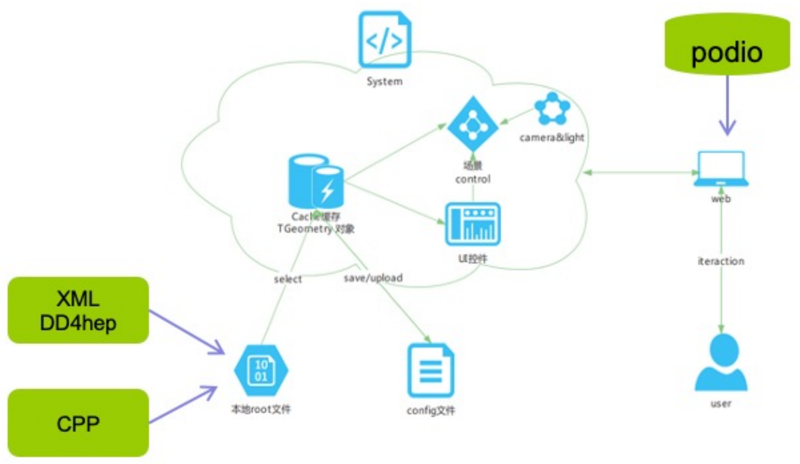


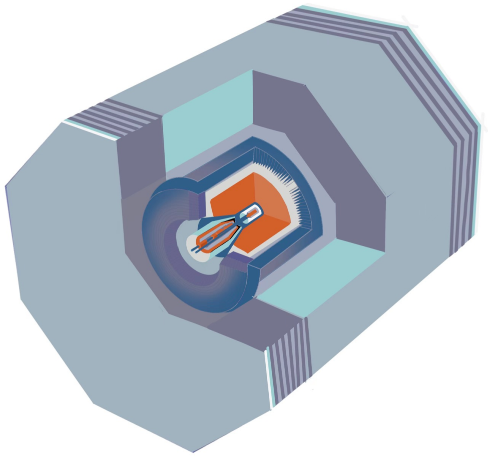


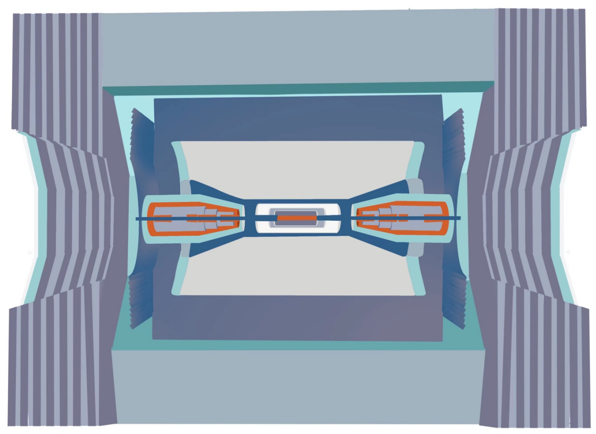


Detector and Event Display

- A common geometry and event display system is being developed
  - Based on Web3D technology and the open-source JSROOT
  - 3D engine and graphic library based on Three.JS
  - Using the Vue.js HTML5 development framework to implement the Web interface
  - Reducing 3D motion lag by the multi-threading capabilities of Web Worker framework



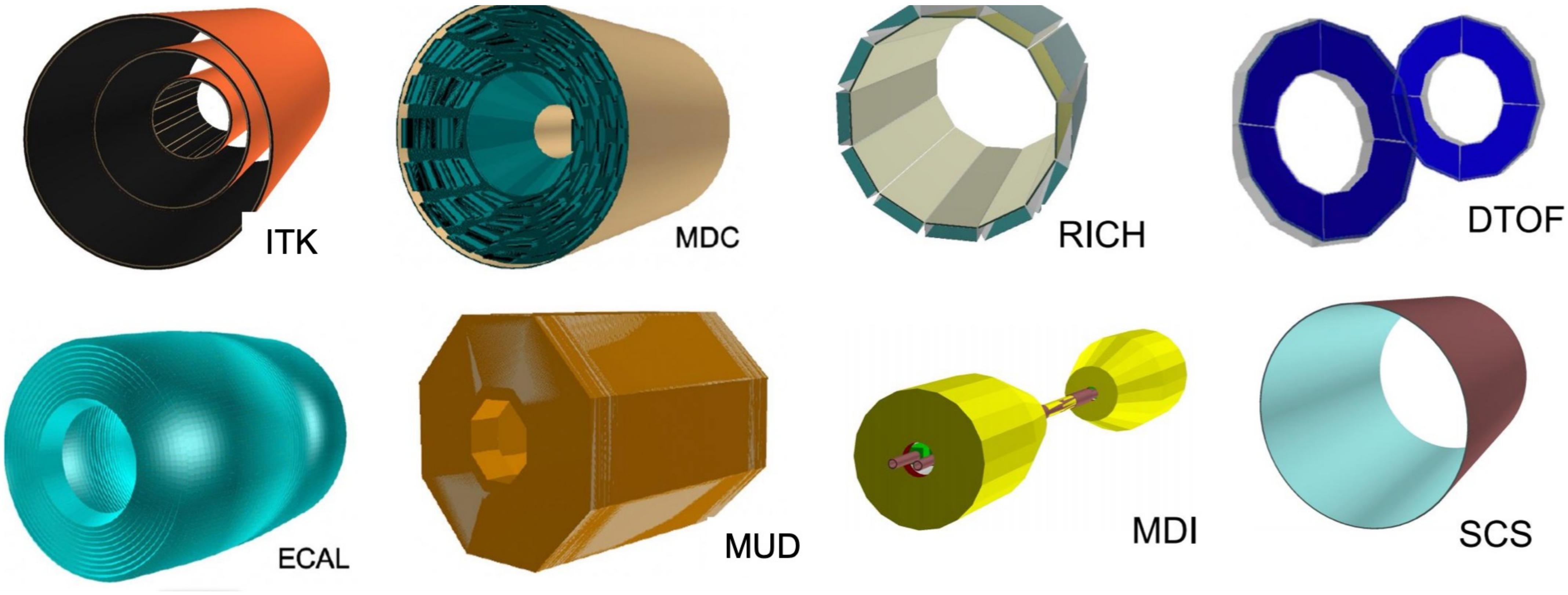




Detector Simultion

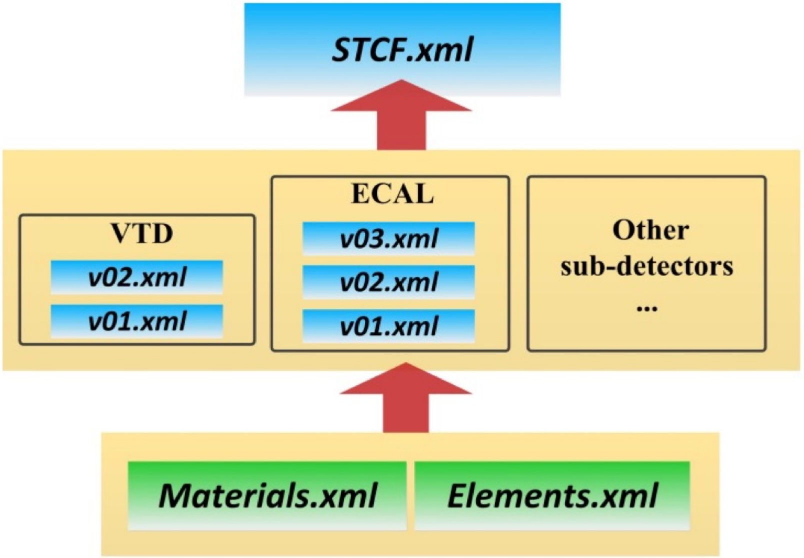
Detector Description

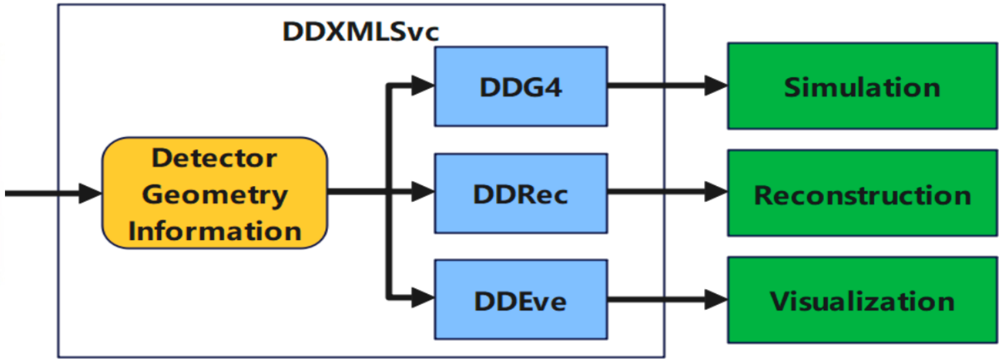
- The full STCF Detector is described with DD4hep
- Each sub-detector is implemented with a single compact file
- Very convenient to optimize detector geometry according to detector experts



Geometry Management System

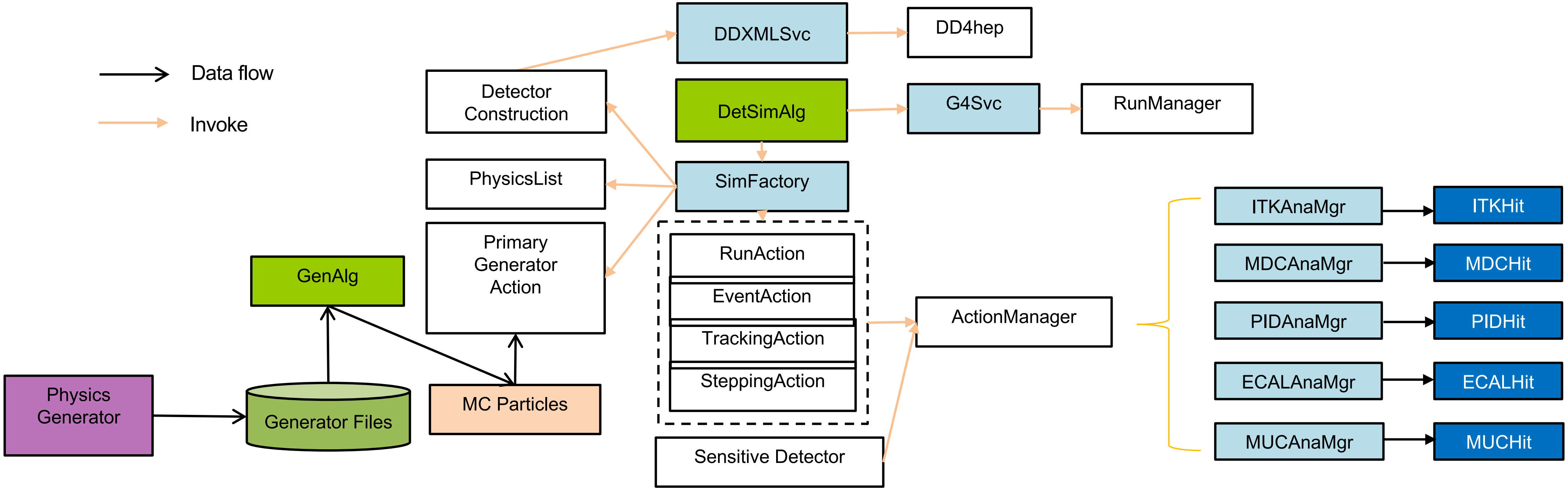
- Developed geometry management system for all applications
  - detector simulation, reconstruction and event display
  - The version number is used for different design options
- Ensure consistent detector information between different applications with a single source of detector description
  - DDG4 for delivering detector geometry to Geant4
  - DDRec for delivering detector geometry to reconstruction algorithms
  - DDXMLSvc: the unified interface to DD4hep, including DDG4 and DDRec





Detector Simulation Chain

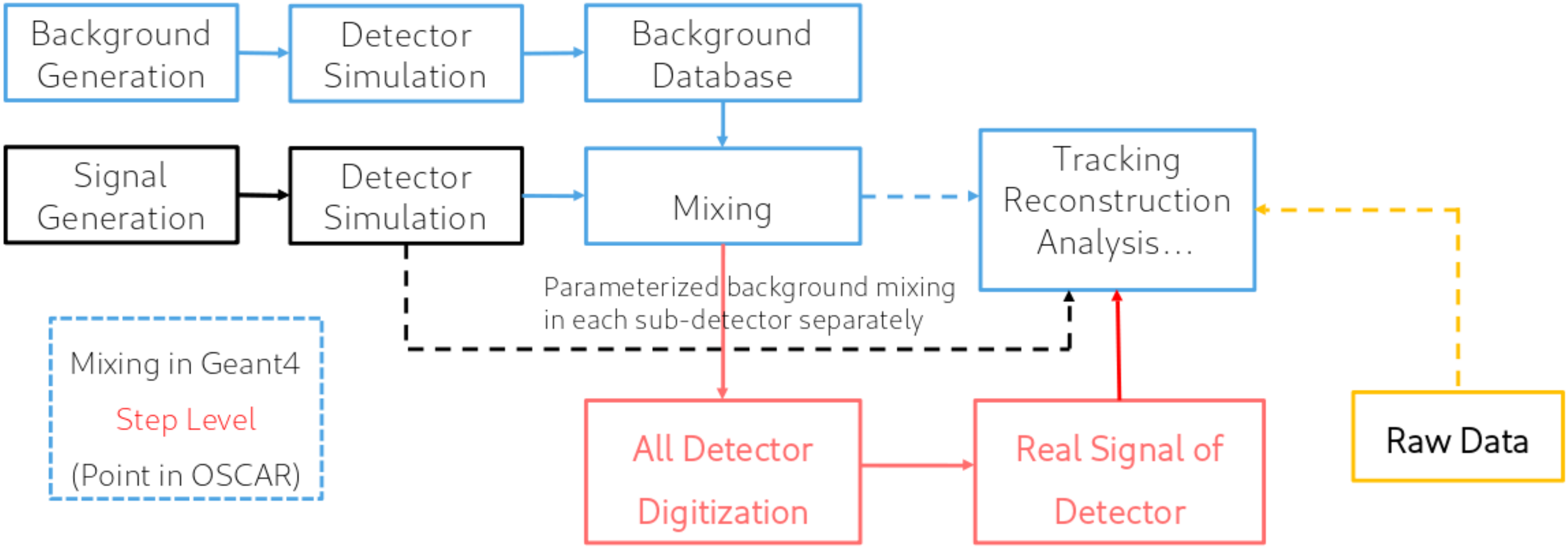
- Full chain of detector simulation from generator to simulated information is built
- Providing flexible configuration for different purposes of detector simulation
  - Generator for different physics topics i.e. Babayaga, KKMC, Phokhara, DIAG36, BBBrem …
  - Geometry for different detector design options
  - User actions for recording MC truth information



Background generation and event mixing

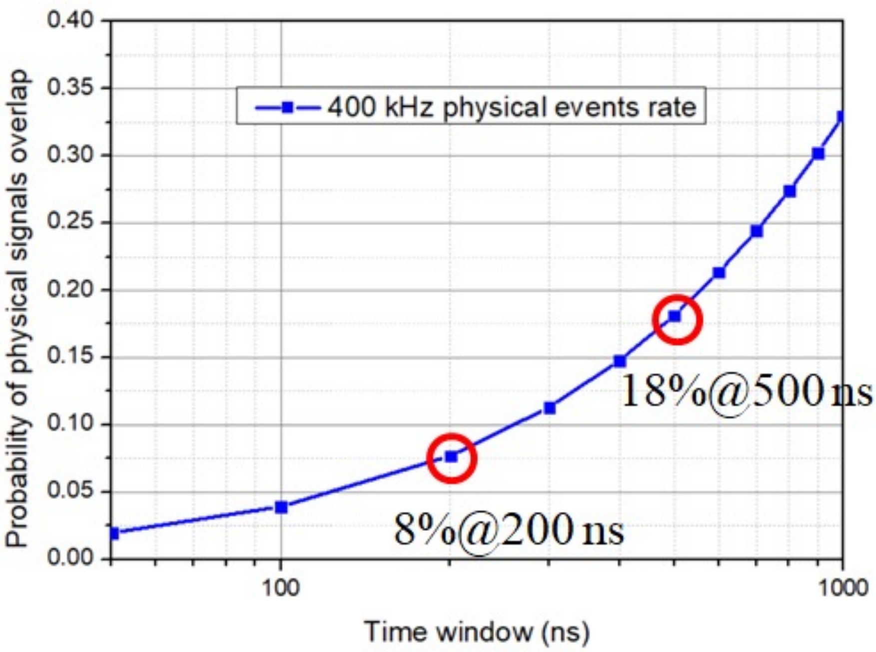
Yupeng Pei, STCF beam background sim. and impl.,
Software & Computing Session, Jan. 16

moving to ‘realistic’ simulation

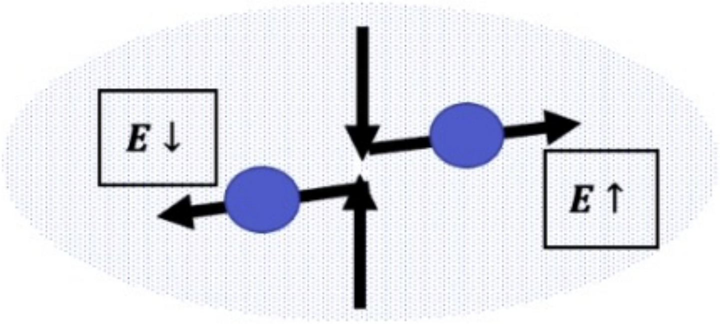


Background Production

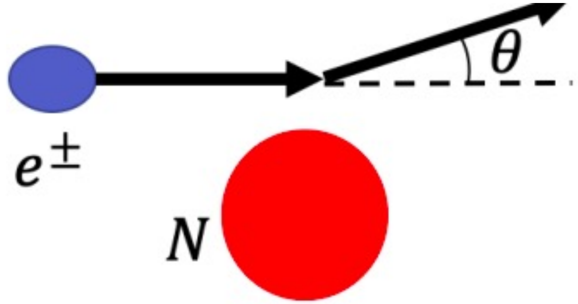
- Maximum event rate at STCF: 400kHz
  - Maximum time window in sub-detectors : 1000 ns
  - Probability of events overlapping in1000 ns: 32.9%
  - Multiple physics events can exist in one event, i.e. pileup



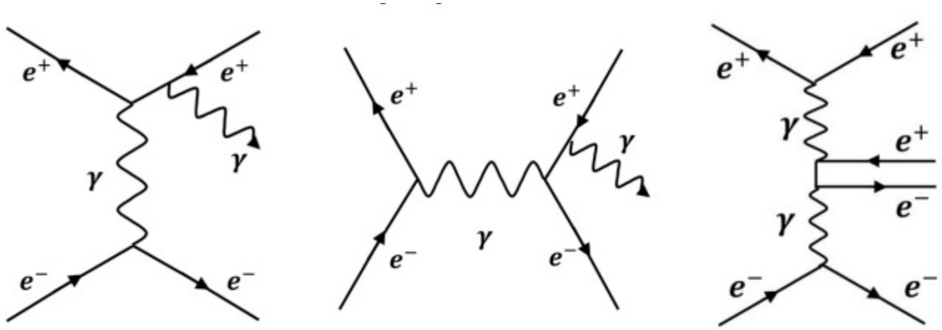
- Main Backgrounds: Touschek, beam-gas, luminosity-related
- Setup background simulation software and produced background data samples



Touschek



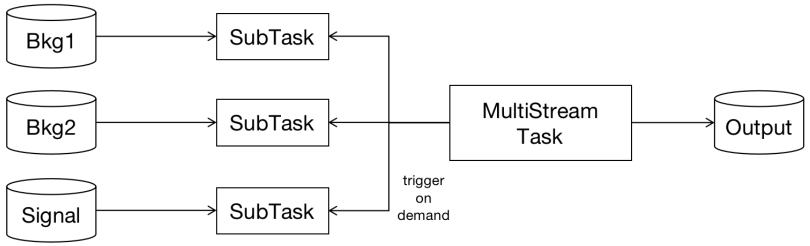
Beam-gas



Luminosity-related

Event Mixing Algorithm

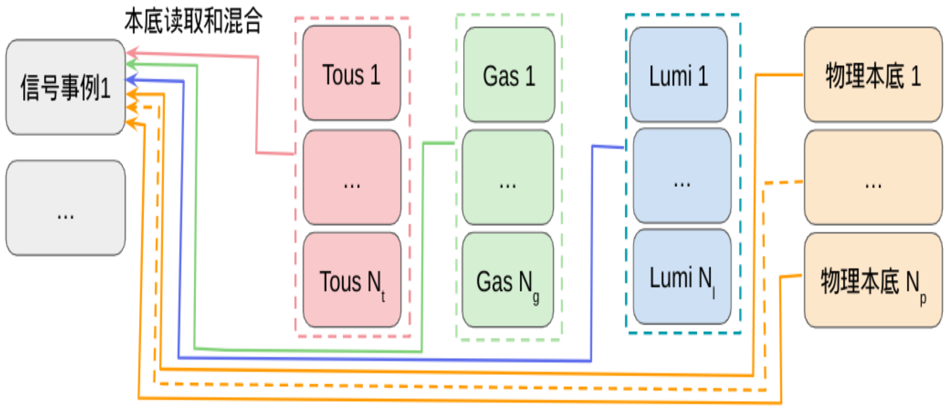
- Developed mixing algorithm based on multi-stream functions provided OSCAR



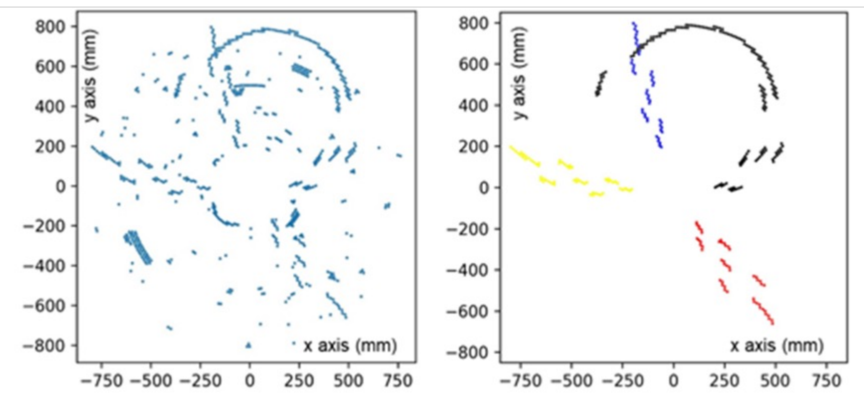
- Providing flexible configuration for event mixing
  - Signal, e.g. $e^+e^- \to \pi^+\pi^- J/\psi(\to \mu^+\mu^-/e^+e^-)$
  - Backgrounds: Touschek, beam-gas, luminosity-related, reading from background database
  - Underlying physics background: $e^+e^- \to anything$ at 4.26 GeV

- Mixed data used for trigger simulation and optimizing of reconstruction algorithms

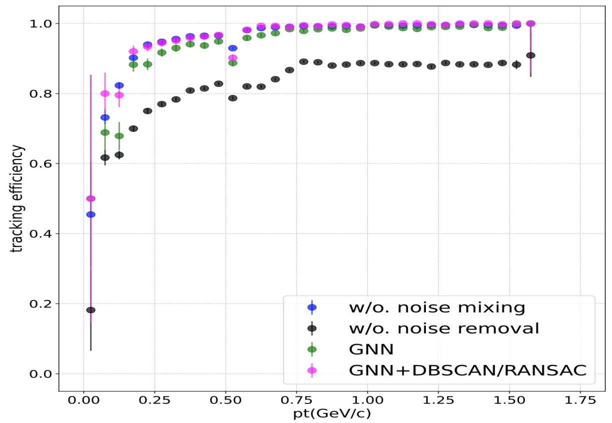
Strategy for event mixing



Trigger simulation



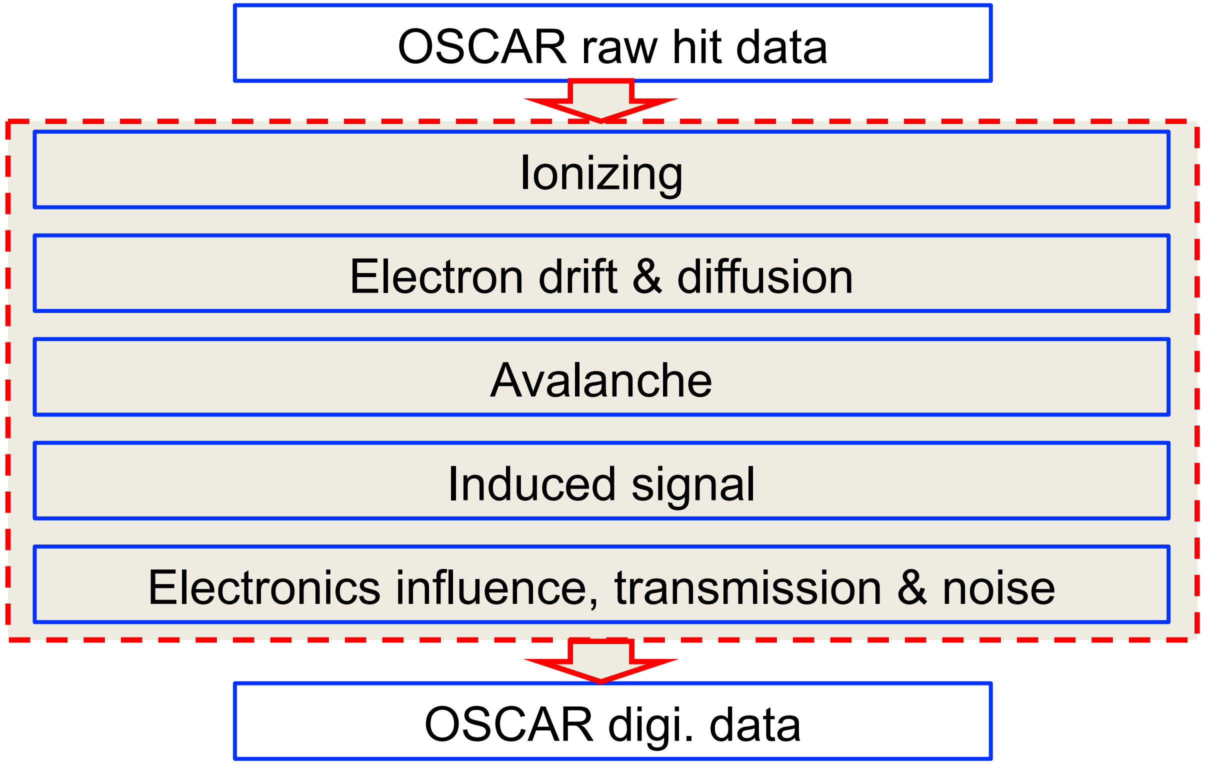
Tracking via GNN noise removal



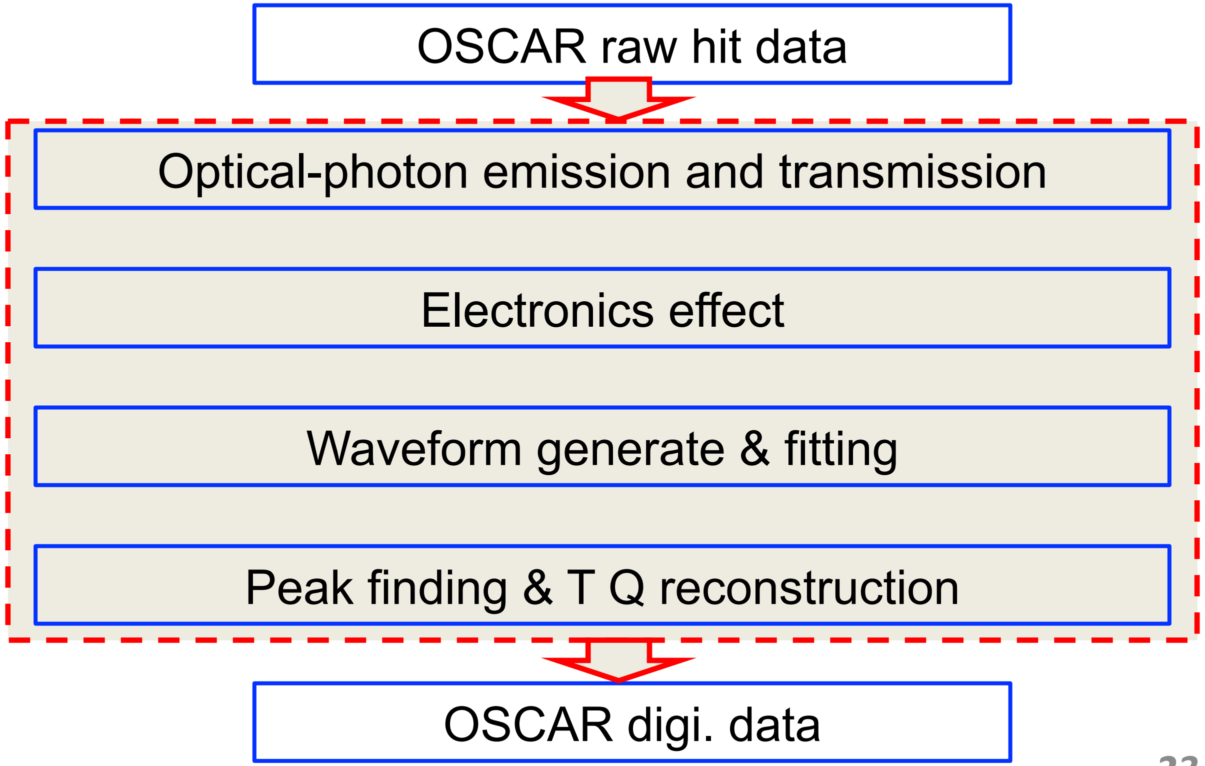
Digitization

- Developed a unified digitization framework for all sub-detectors on OSCAR
- Each sub-detector implemented its digi. algorithms and produced the digi. information
- The reconstruction algorithms are under tuning with the digi. information as their inputs

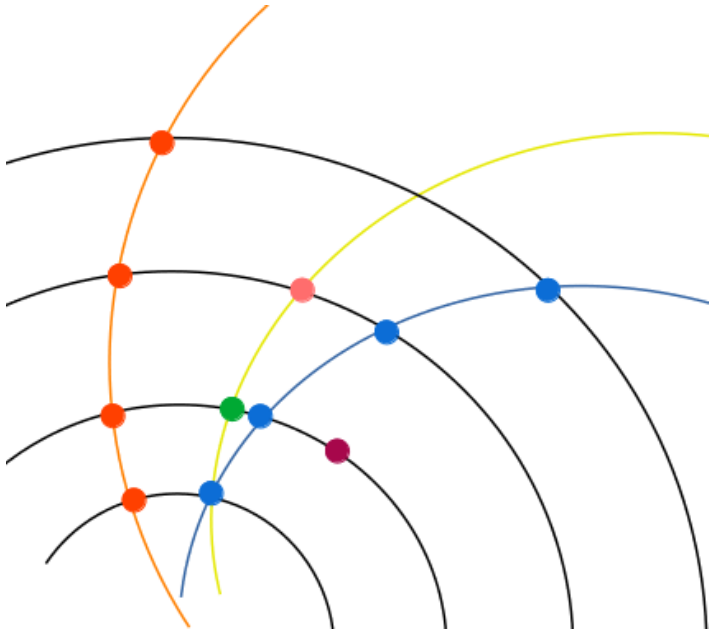
Gaseous detectors:



Scintillator detectors:

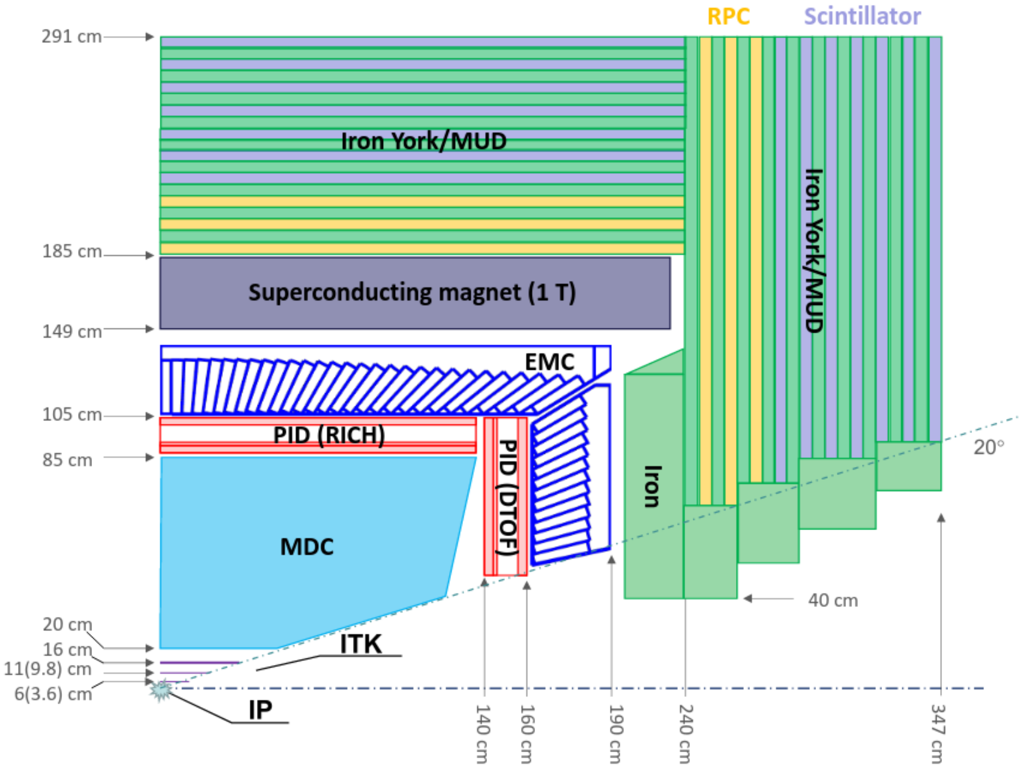


Tracking at STCF

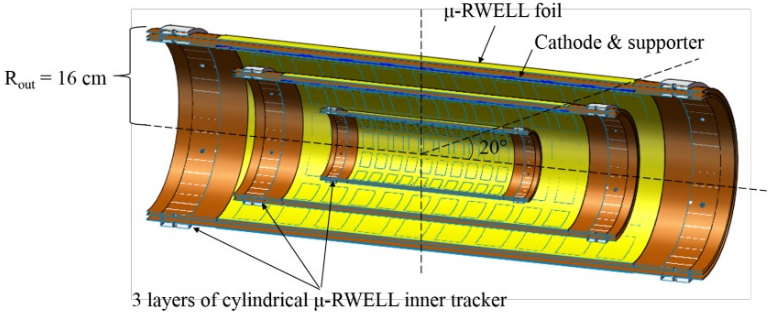


STCF tracking system

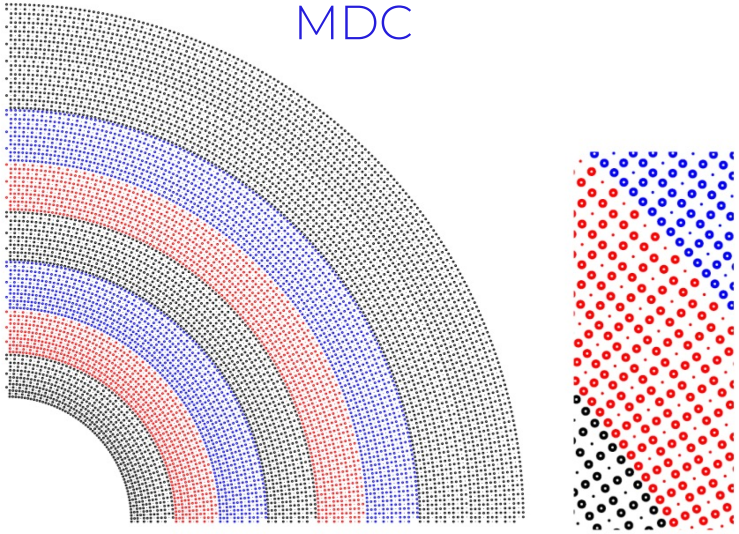
- The baseline tracking system includes uRWELL-based Inner Tracker (ITK) and Main Drift Chamber (MDC)
  - ITK: 3 layers, $\sigma_{r\text{-}\varphi} \times \sigma_z \approx$ 100 um x 400 um
  - MDC: 48 layers, $\sigma_{\text{drift dist}} \approx$ 120~130 um



uRWELL-based ITK



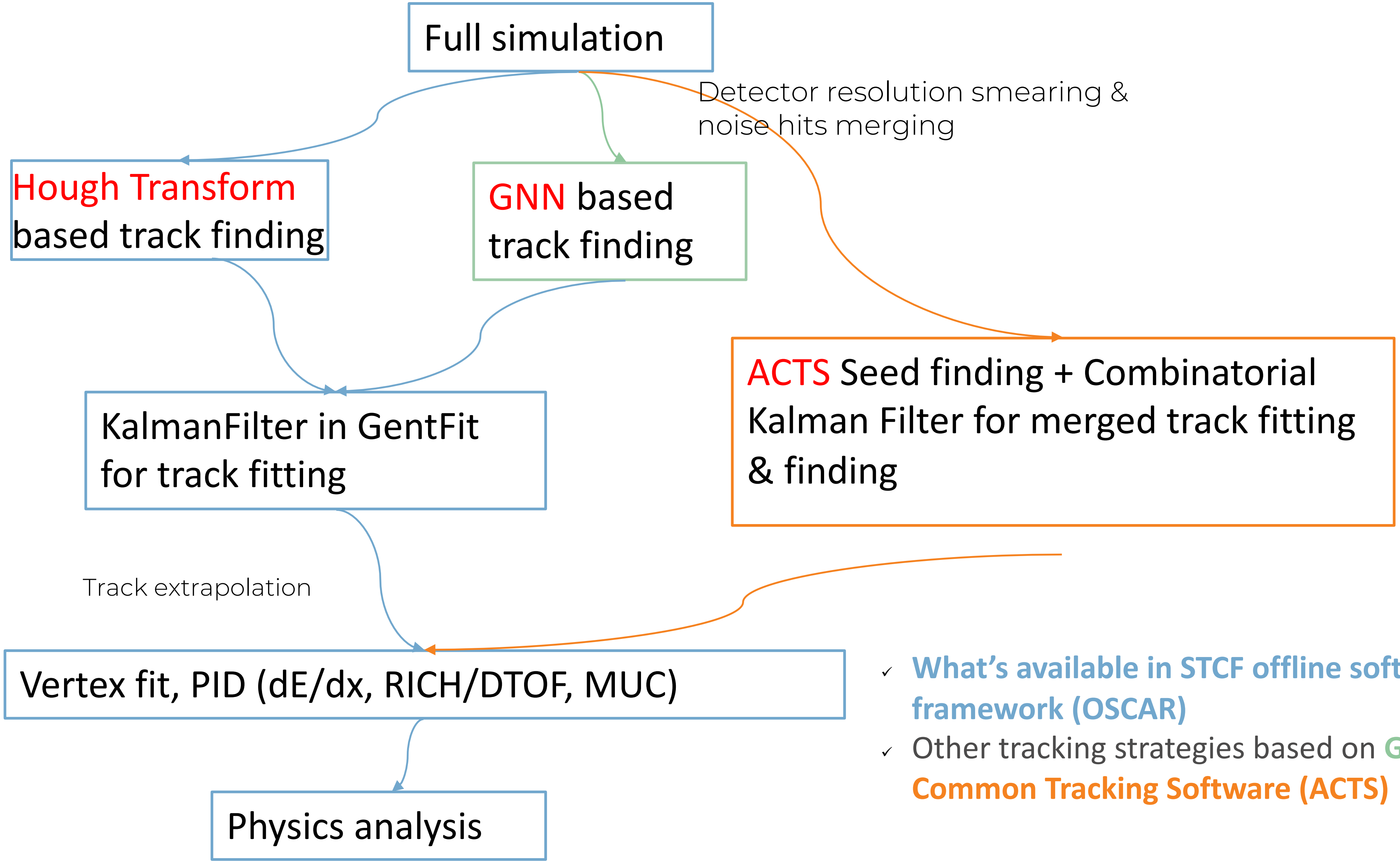
MDC



(a) (b)

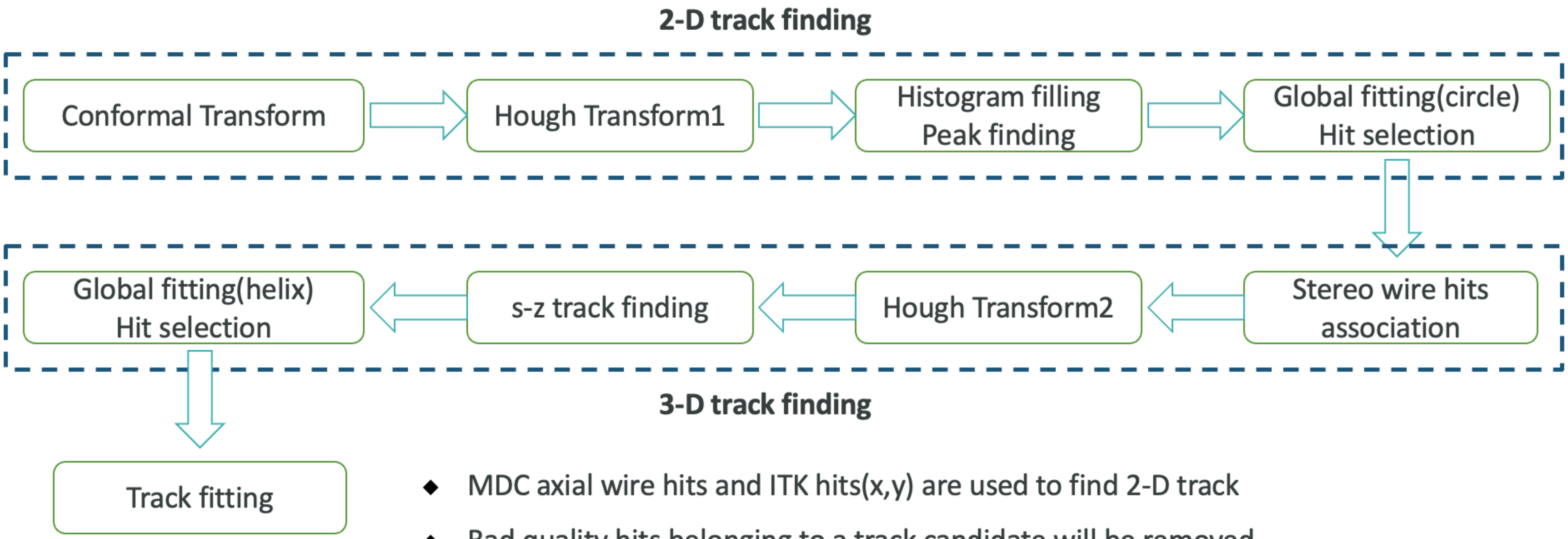
Figures from STCF CDR (arXiv: 2303.15790)

STCF tracking landscape



- ✓ **What's available in STCF offline software framework (OSCAR)**
- ✓ Other tracking strategies based on GNN and **A Common Tracking Software (ACTS)**

Tracking based on Hough Transform



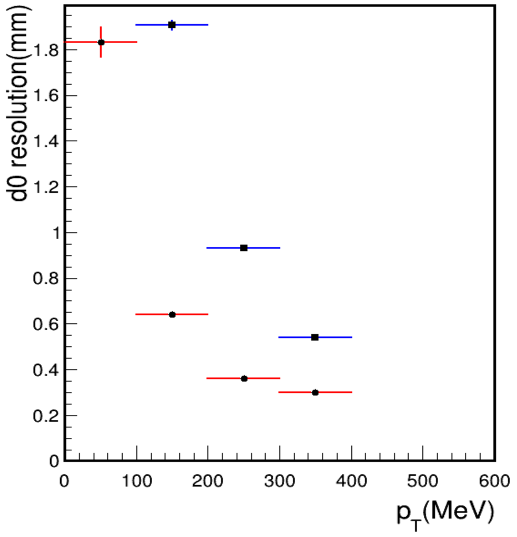
- MDC axial wire hits and ITK hits(x,y) are used to find 2-D track
- Bad quality hits belonging to a track candidate will be removed
- The trajectory is straight in the s-z space → similar to the 2-D tarck finding

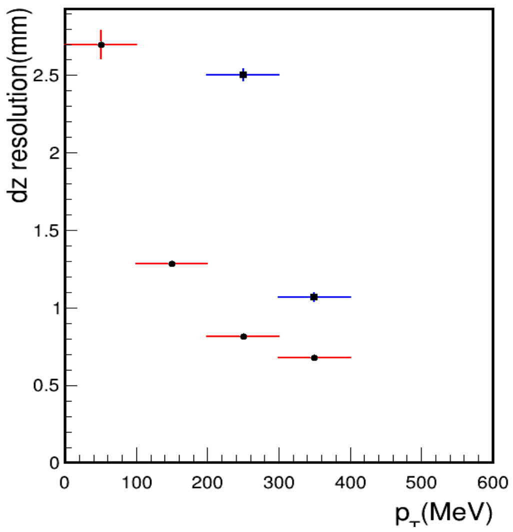
Tracking performance of Hough Transform

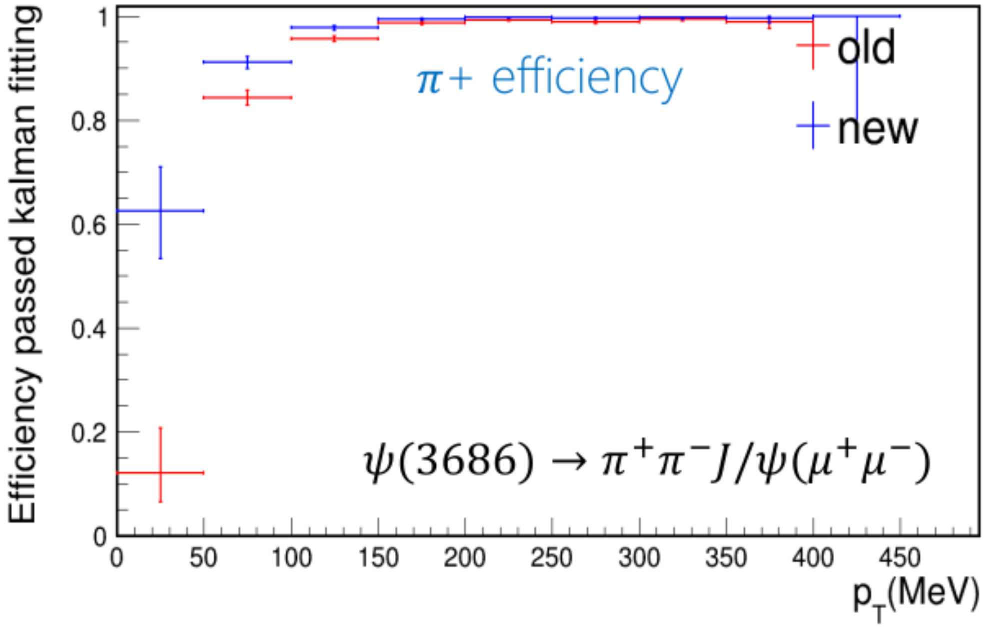
Hang Zhou, Tracking Software for STCF, Software & Computing Session, Jan. 16

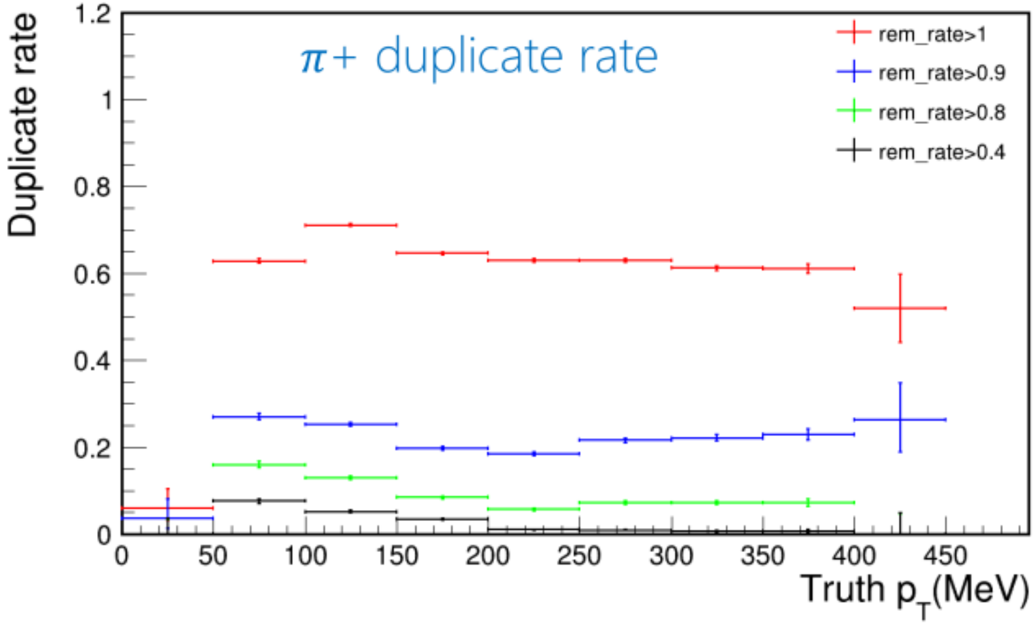
- >92% tracking performance event for π with p_T in [50, 100] MeV
- Duplicate tracks removed using shared hits info (<1% duplicate rate with p_T > 200 MeV)

d0, dz resolution of π+





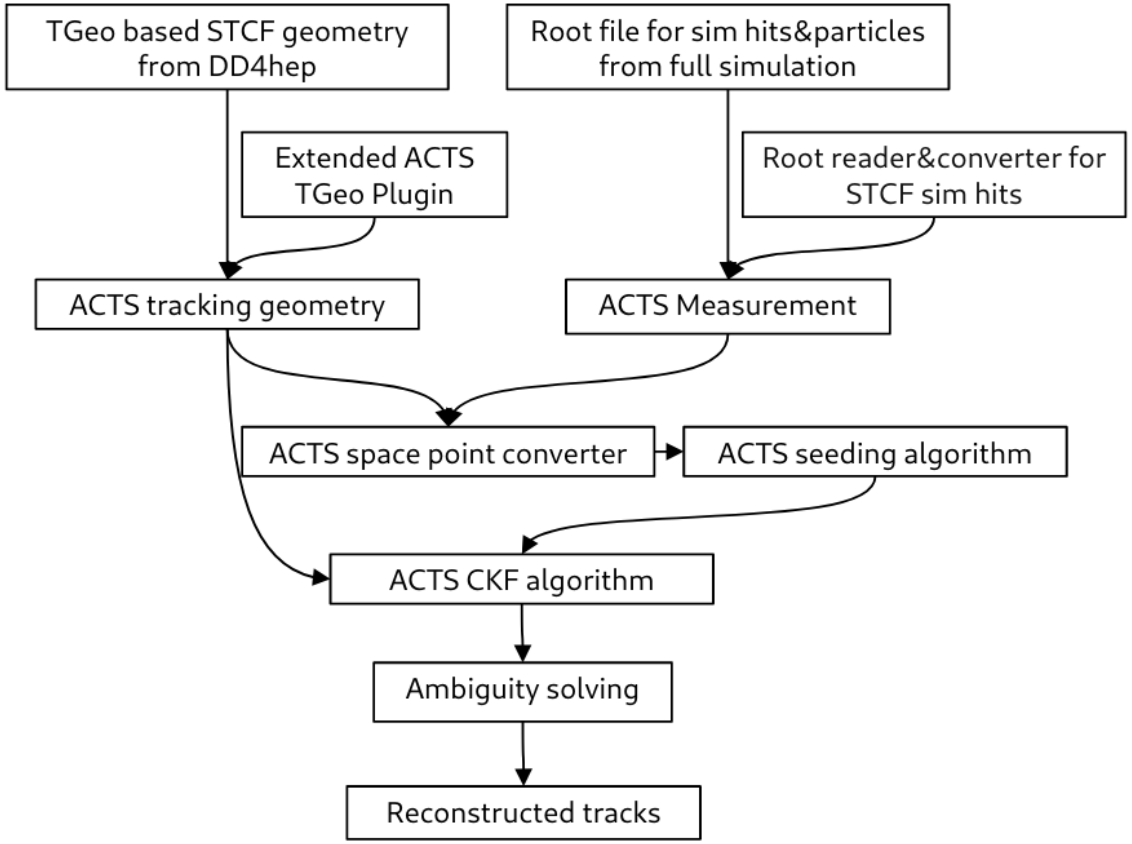




Tracking based on ACTS

Xiaocong Ai, ACTS app. for gaseous tracking detector,
Software & Computing Session, Jan. 16

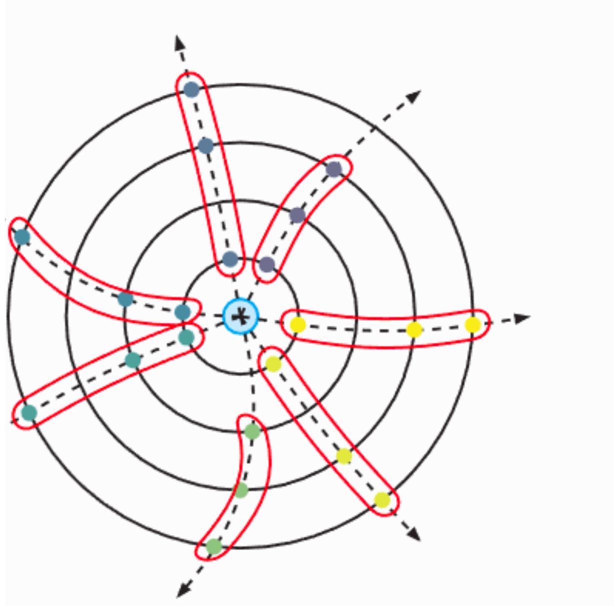
Application strategies of ACTS for STCF

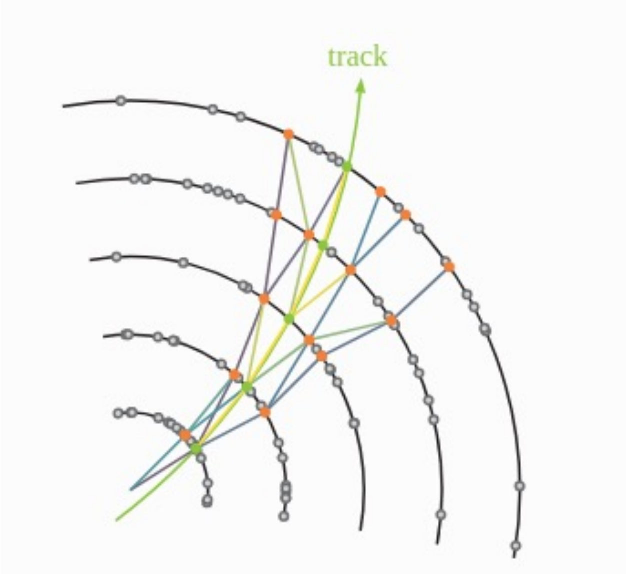


Seeding
(find seeds using hits on ITk layers)

→

Combinatorial Kalman Filter (CKF)
(simultaneous track fitting + finding)

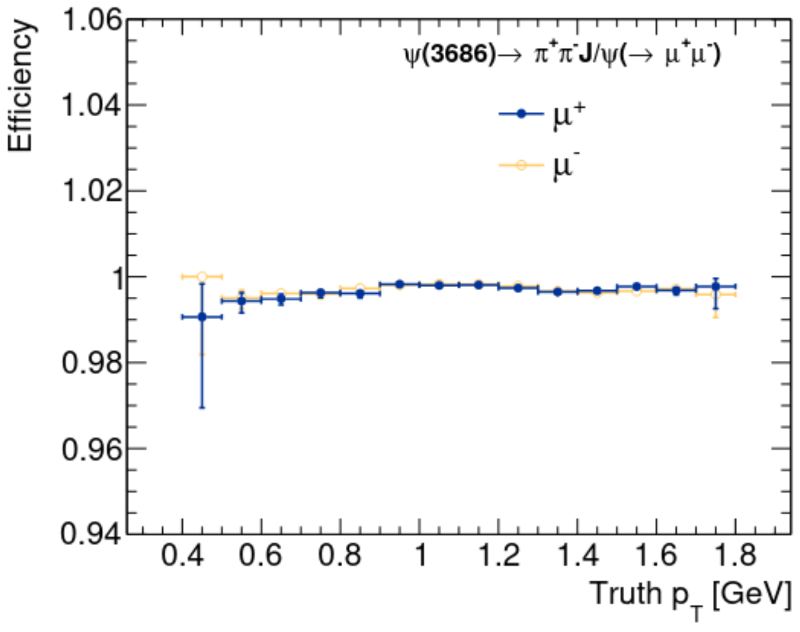


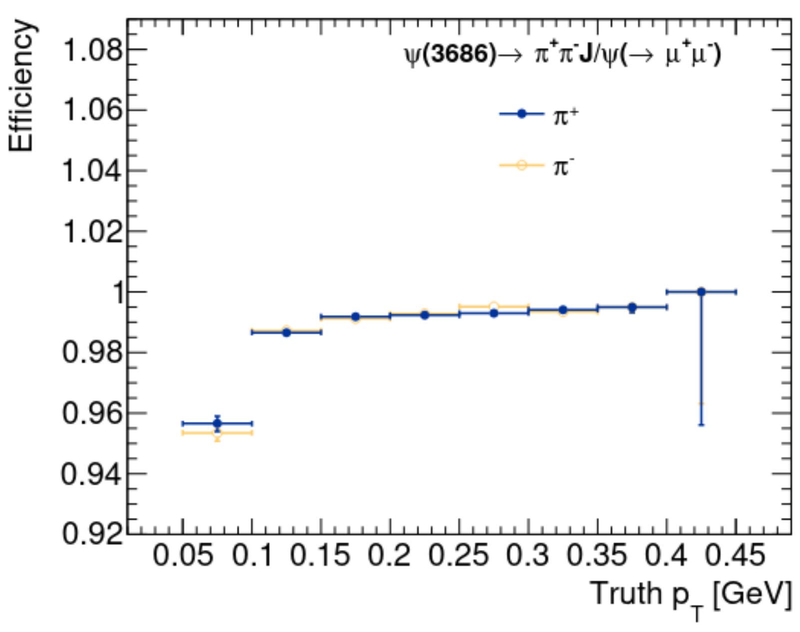


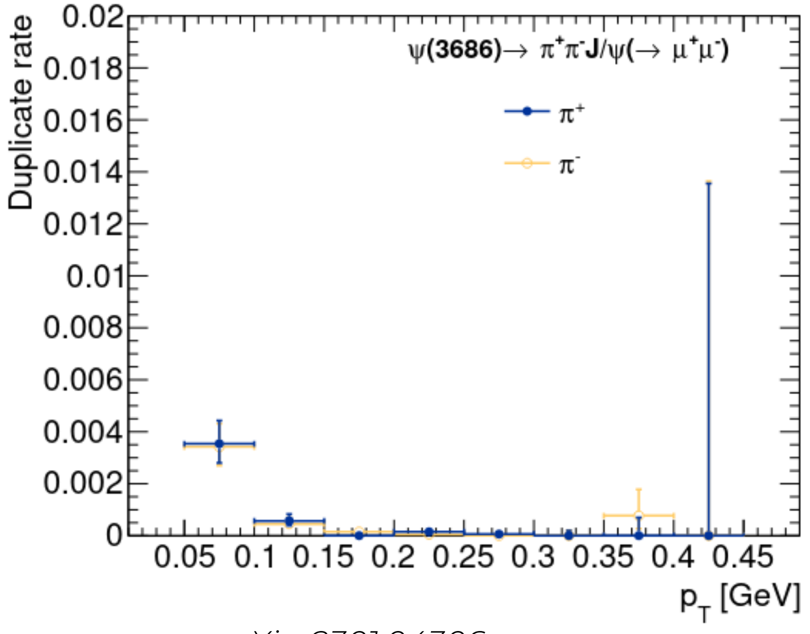
Figures from ACTS readthedocs

Tracking finding performance with ACTS

- Above 99% efficiency for $p_T > 400$ MeV
- 95% efficiency for pion with p_T in [50, 100] MeV
- <0.5% duplicate tracks for $p_T < 130$ MeV due to duplicate seeds for looping tracks
- Negligible fake tracks (<0.01%)



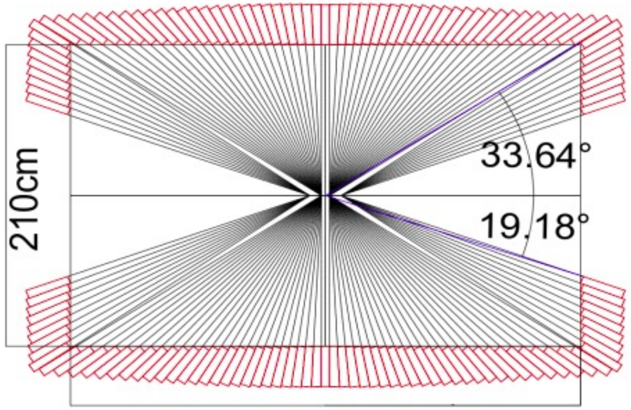




arXiv: 2301.04306

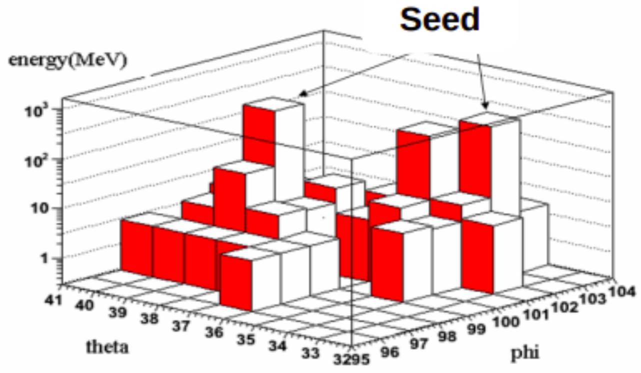
Photon reconstruction

Reconstruction method



Crystal arrangement diagram

Visualized by DD4Hep



π^0 Cluster (two photons)

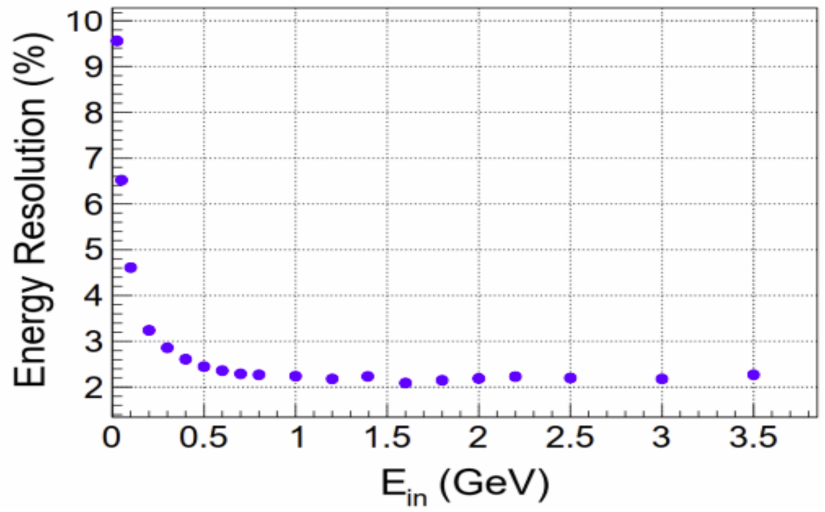
ECALRecAlg

- Cluster search
- Seed search
- Clusters split into Showers
- Calculation and correction of shower energy and position

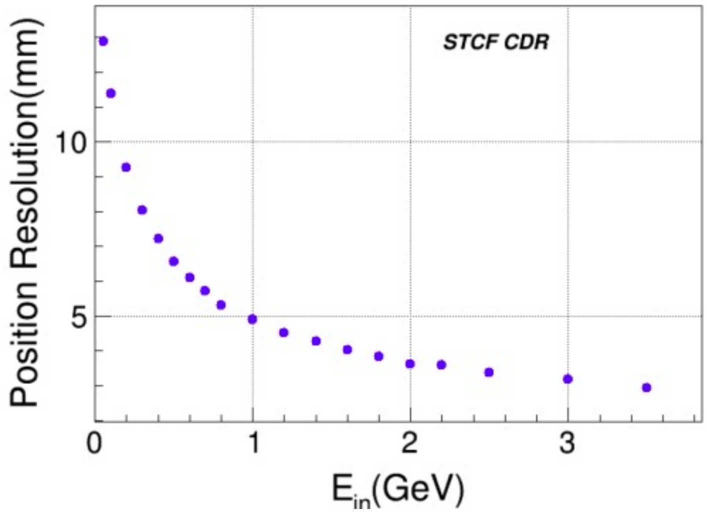
Peformance of ECAL reconstruction

Bo Wang, ECAL Software for STCF, Software & Computing Session, Jan. 17

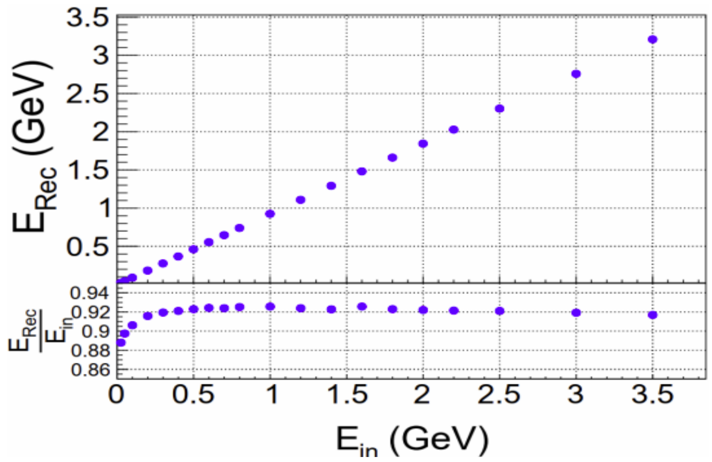
- ❖ Energy resolution is 2.2% at 1 GeV



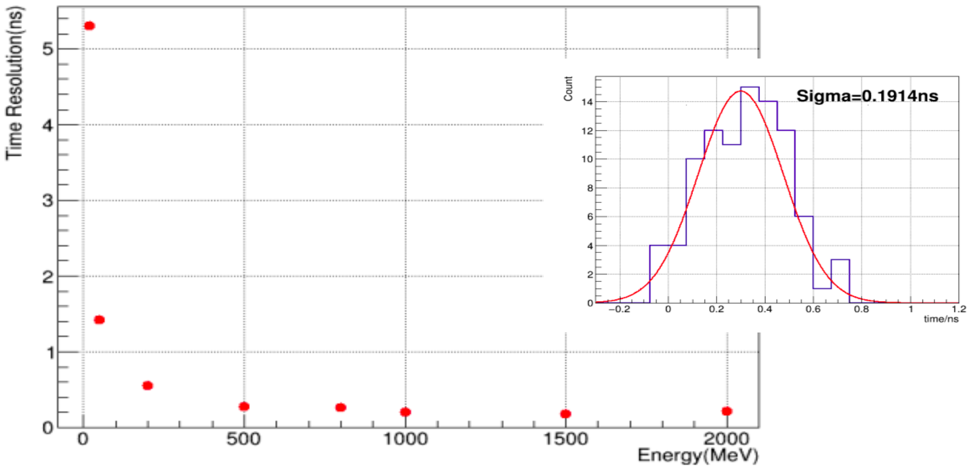
- ❖ Position resolution is 4.9 mm at 1 GeV



- ❖ Good energy linearity between 25 MeV – 3.5 GeV



- ❖ Time resolution is 200 ps based on waveform fitting

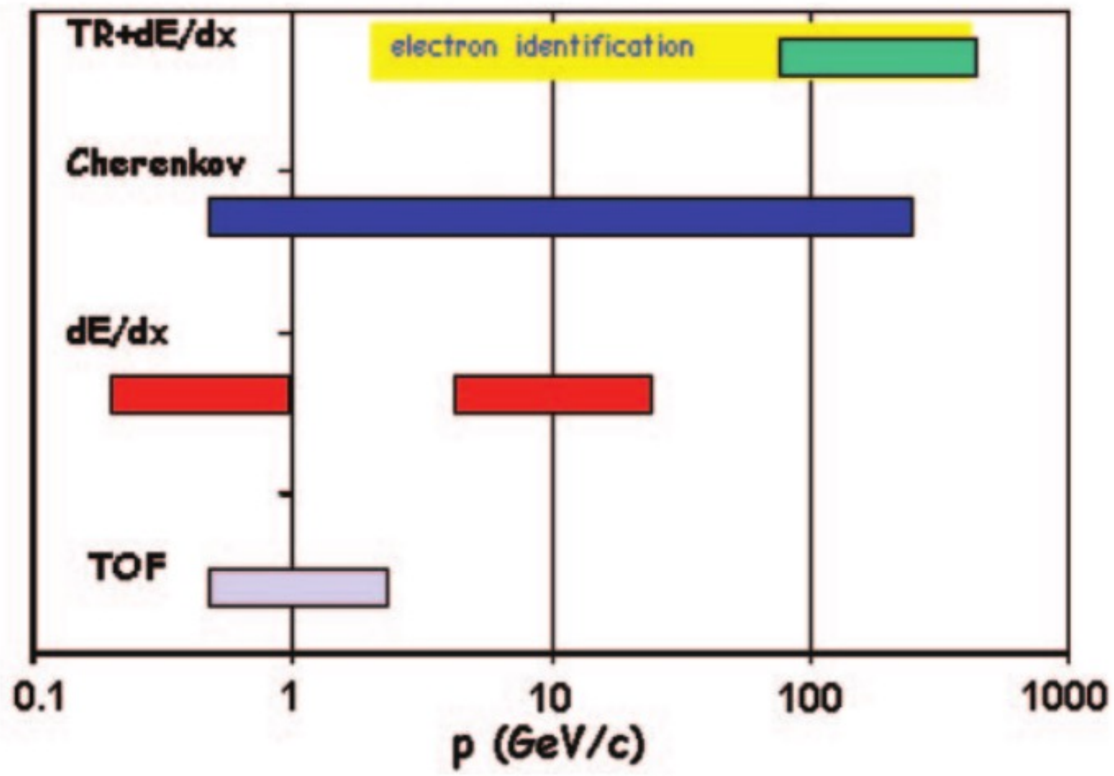


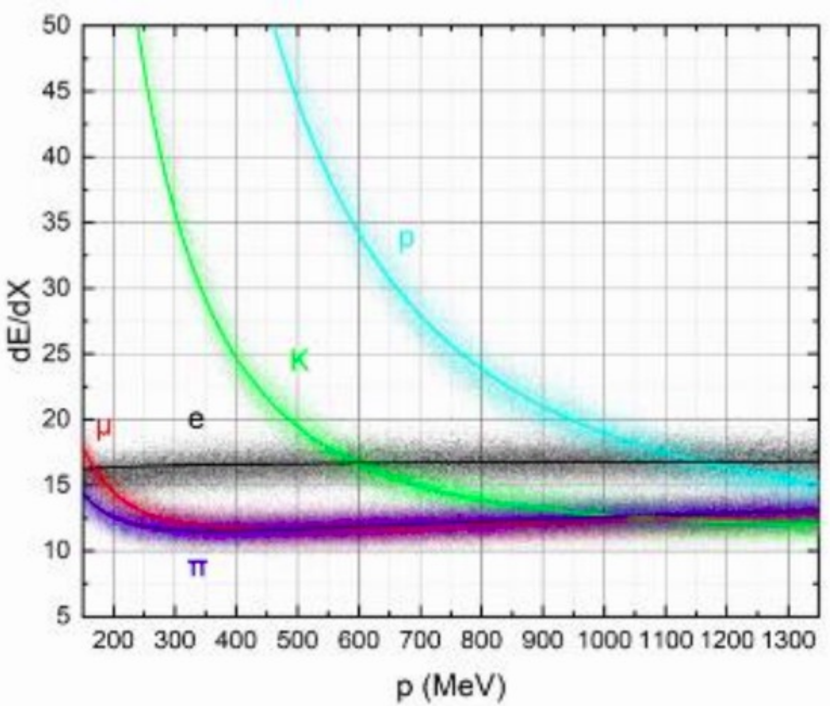
PID at STCF

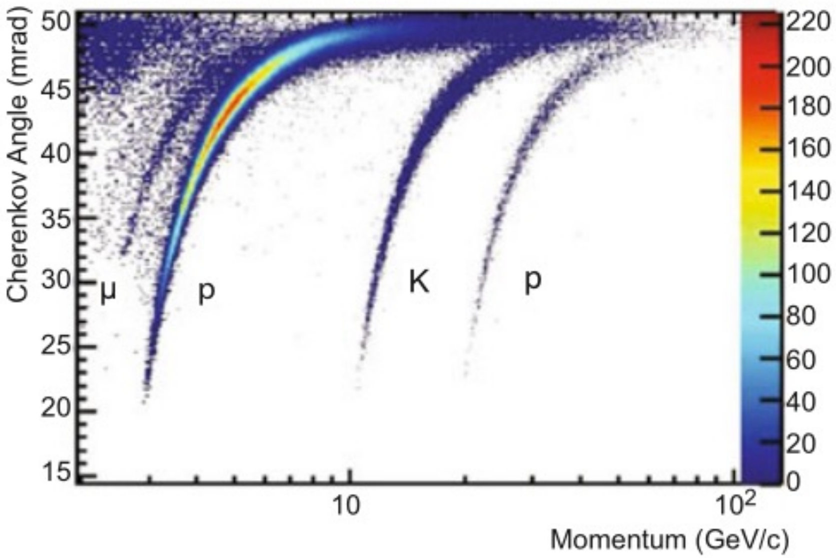
- ➔ π/K separation based on PID detectors
- ➔ PID based on global information!

PID methodology

- Ultimately based on the difference of velocity between particles!

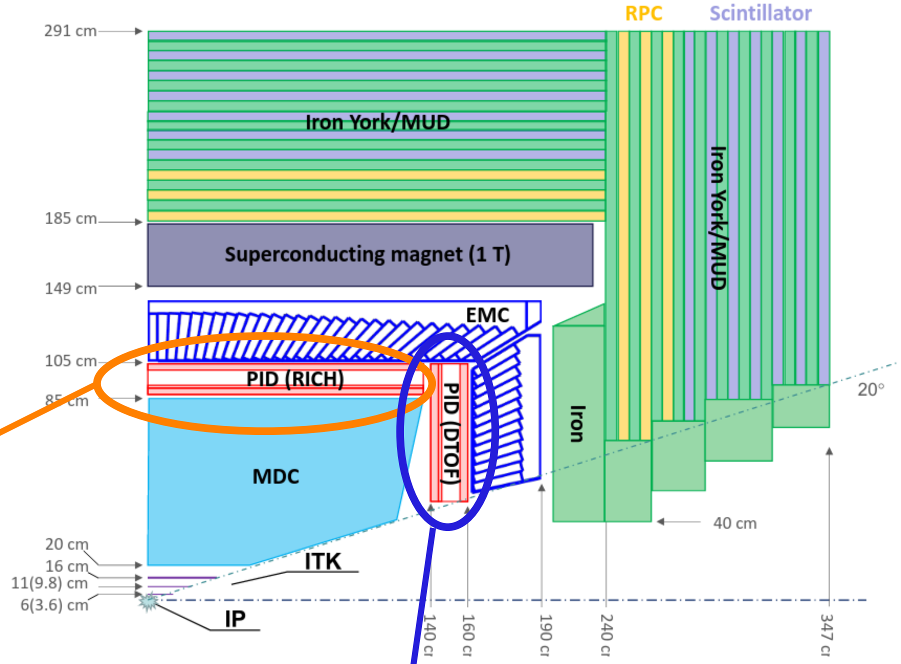


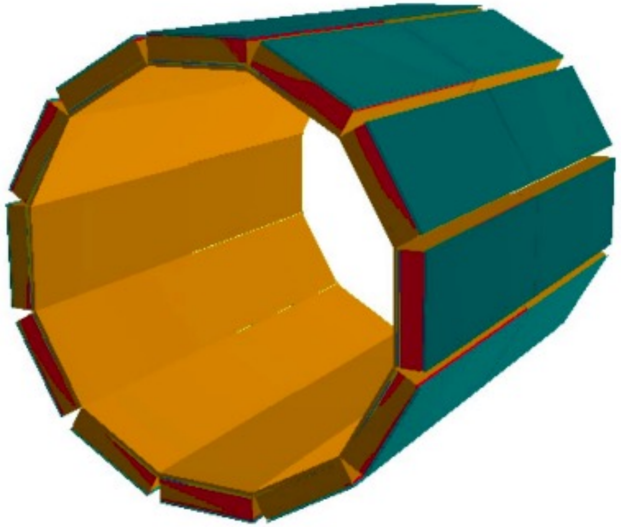


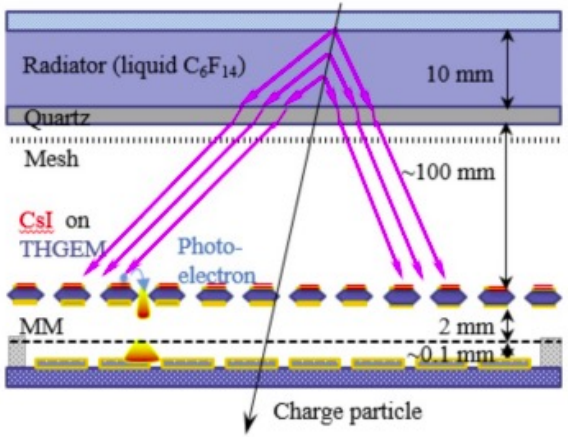


PID detectors at STCF

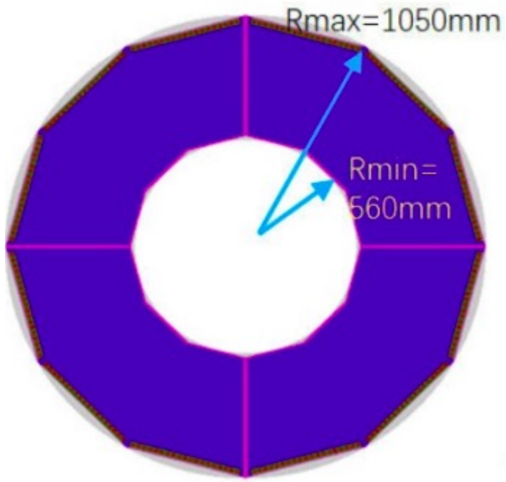
- Cherenkov-based technology for PID detectors in both barrel (RICH) and endcap (DTOF)

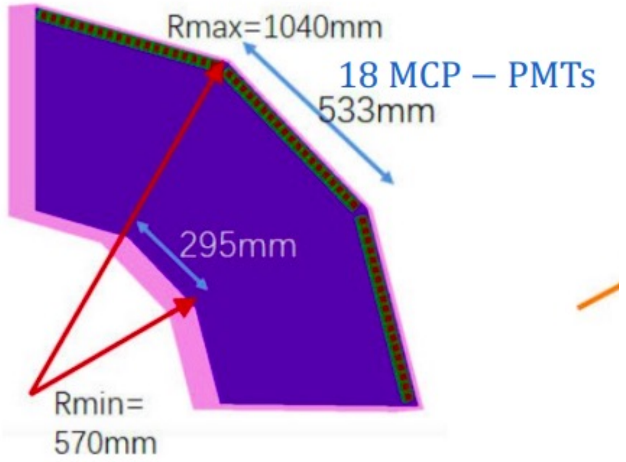






THGEM + MM





RICH PID performance with likelihood method

- The photon collected in each anode pads follows the **Poisson distribution**

$$pdf_{i,h} = Poisson\left(N_i + 10^{-3}, avg_{i,h} + 10^{-3}\right),$$

num of γ in anode pad

avg of γ in simulation

10^{-3} bkg level

- Likelihood of h hypothesis:

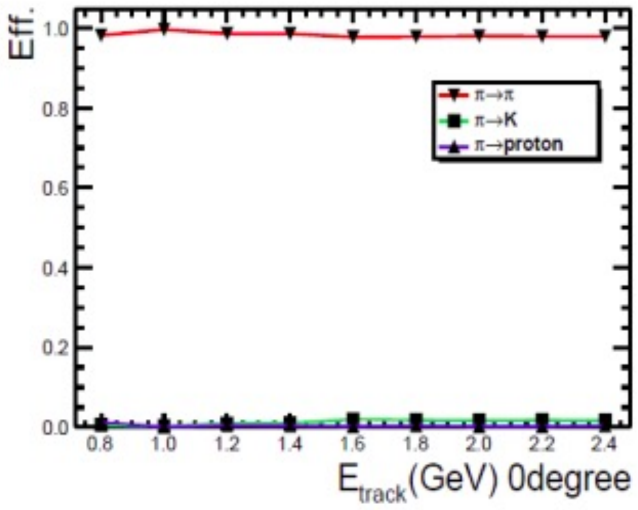
$$\ln \mathcal{L}_h = \Sigma_i^{npads} \ln pdf_{i,h}$$

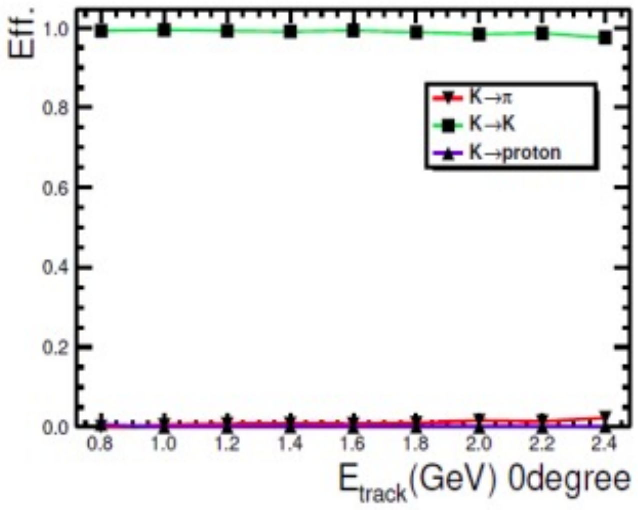
- π, K separation:

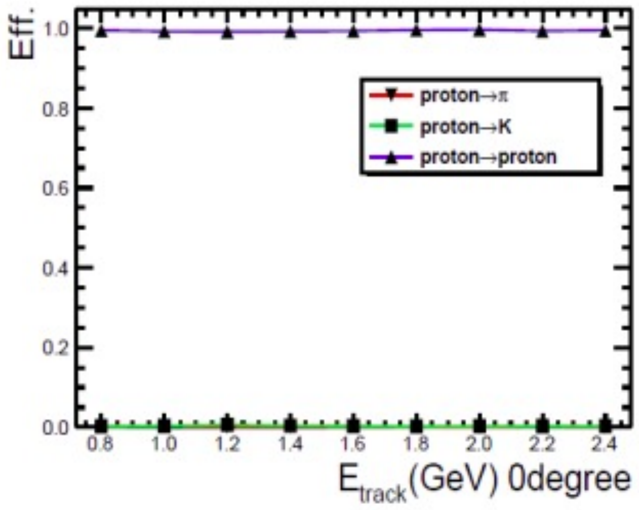
$$DLL = \Sigma_i^{npads} \ln \frac{pdf_{i,\pi}}{pdf_{i,K}}$$

PDF of average photon number with different particle assumptions is required

- Performance
  - K/π 3.3σ separation up to 2.0 GeV/c
  - K/P 4.4σ separation up to 2.0 GeV/c
- PDF generation based on Geant4 is very time-consuming. Cherenkov map based on analytic method has been developed







DTOF PID performance with timing method

Yutong, DTOF Software for STCF,
Software & Computing Session, Jan. 17

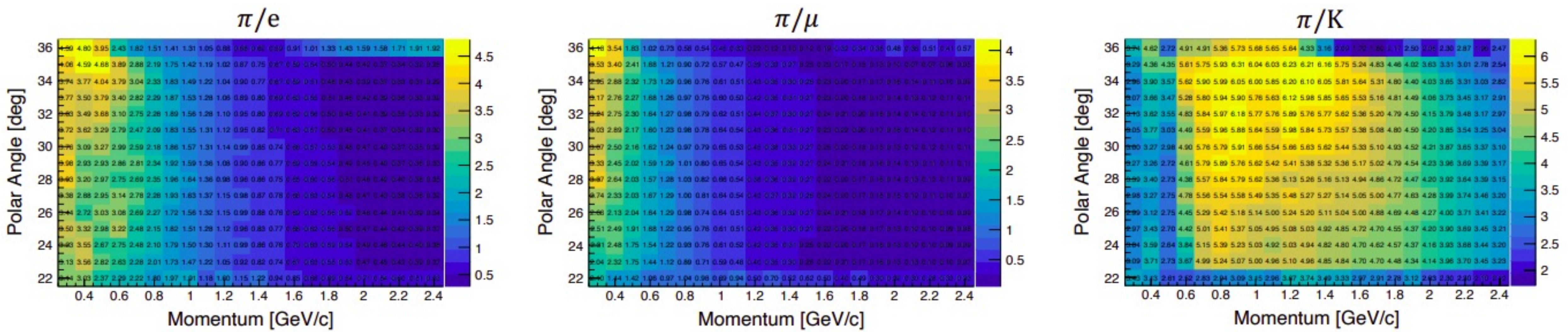
$$\mathcal{L}_h = \prod_{i=0}^{N_{p.e.}} f_h(TOF_i)$$

signal bkg

$$f_h(t) = Gaus(TOF_{rec} | TOF_{hypo}, \sigma) + 0.05$$

where $TOF_{hypo} = \frac{LOF}{c\beta_{hypo}}$

- $\sigma_t \sim 95\ ps$ by single photon-electron
- $\sigma_t \sim 50\ ps$ by multi-photon-electrons
- 3.95σ π/K separation at $p = 2.0$ GeV/c



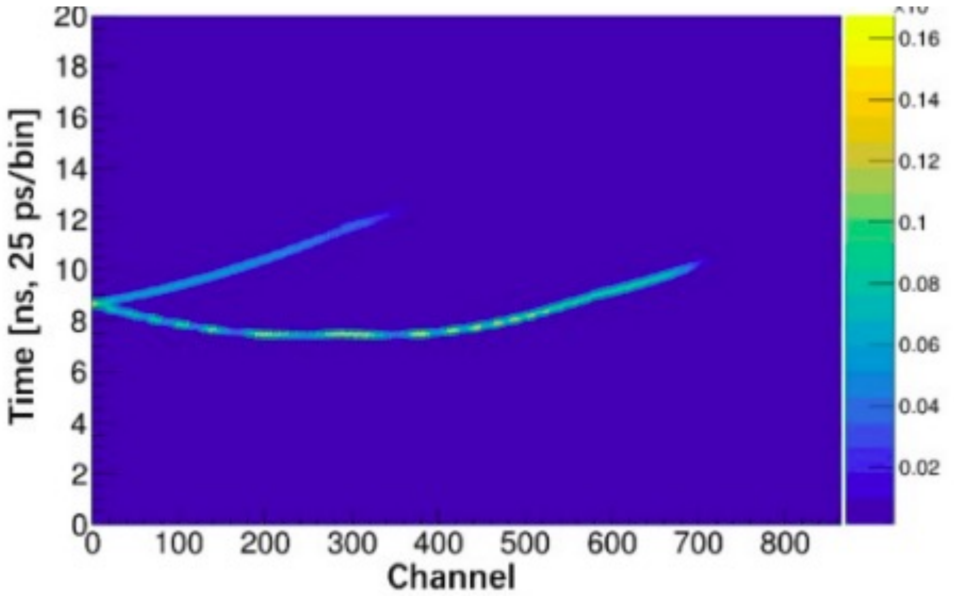
DTOF PID performance with imaging method

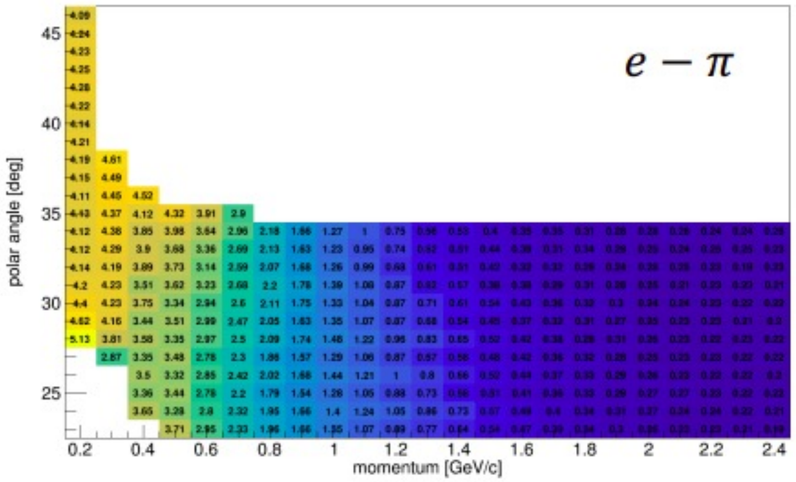
$$\mathcal{L}_h = p_h(N_{p.e.}) \prod_{i=0}^{N_{p.e.}} f_h(x_i, t_i)$$

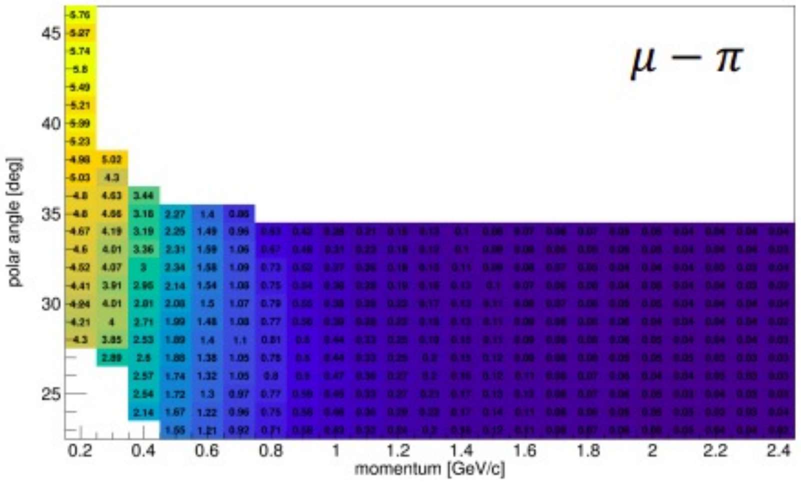
$$p_h(N_{p.e.}) = \sum_{n=0}^{N_{p.e.}} Poisson_h(n, N_e) \times F_{bkg}(N_{p.e.} - n)$$

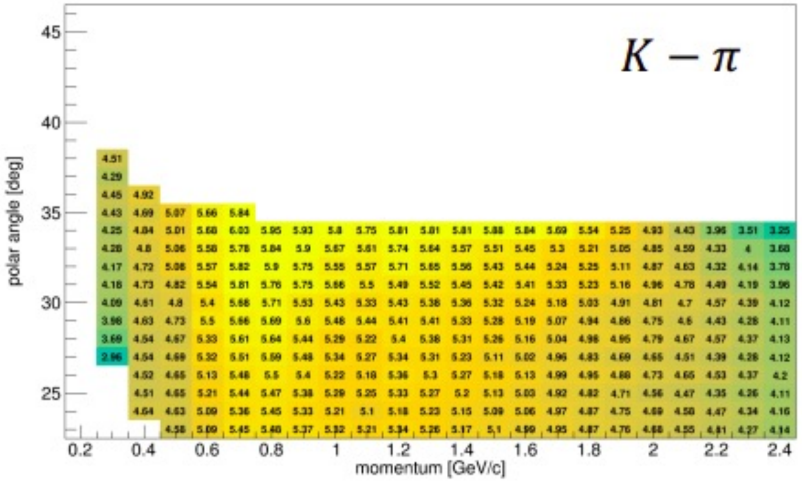
$$f_h(x,t) = S_h(x,t) + const_{bkg}$$

Improve π/K separation $\sim 4.7\sigma$, at $p = 2.0$ GeV/c







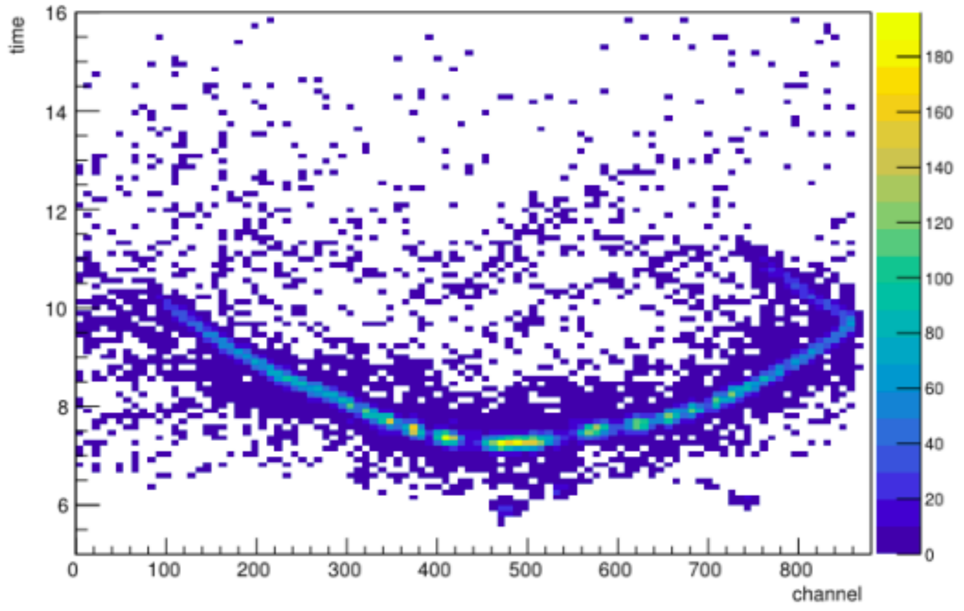


DTOF PID performance with CNN method

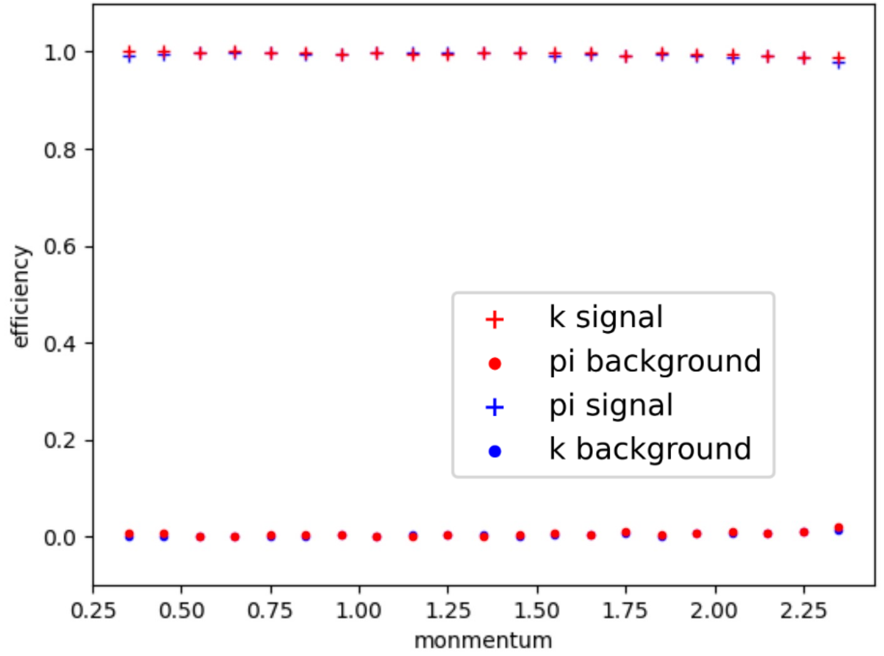
Yucong, ML based PID for STCF, Software & Computing Session, Jan. 17

- A CNN (based on EfficientNetV2S) which utilizes both timing and spatial information of the DTOF hits is developed
- Test set accuracy : 99.46%

Stage	Operator	Stride	#Channels	#Layers
0	Conv3x3	2	24	1
1	Fused-MBConv1, k3x3	1	24	2
2	Fused-MBConv4, k3x3	2	48	4
3	Fused-MBConv4, k3x3	2	64	4
4	MBConv4, k3x3, SE0.25	2	128	6
5	MBConv6, k3x3, SE0.25	1	160	9
6	MBConv6, k3x3, SE0.25	2	256	15
7	Conv1x1 & Pooling & FC	-	1280	1



X-label: the hit position of Cherenkov photon by PMT
Y-label: the arrival time of Cherenkov photon by PMT
Value: the number of photons within this bin



PID at STCF

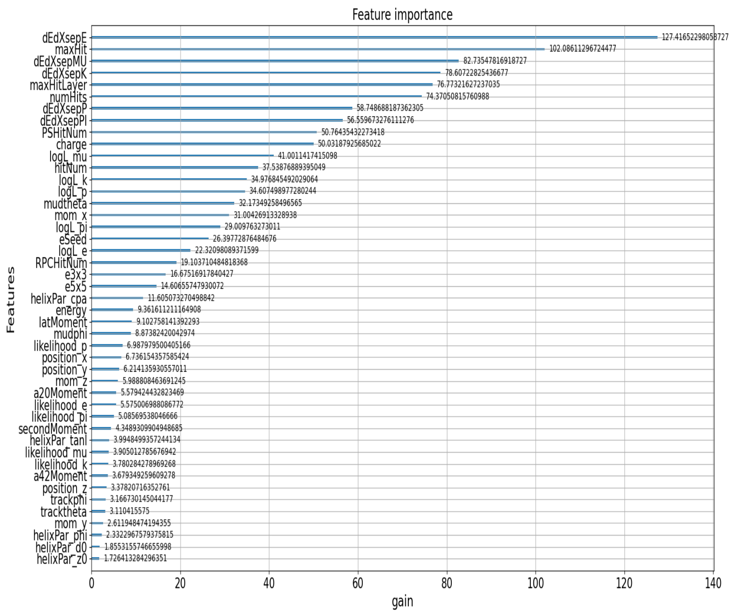
→ π/K separation based on PID detectors

→ **PID based on global information!**

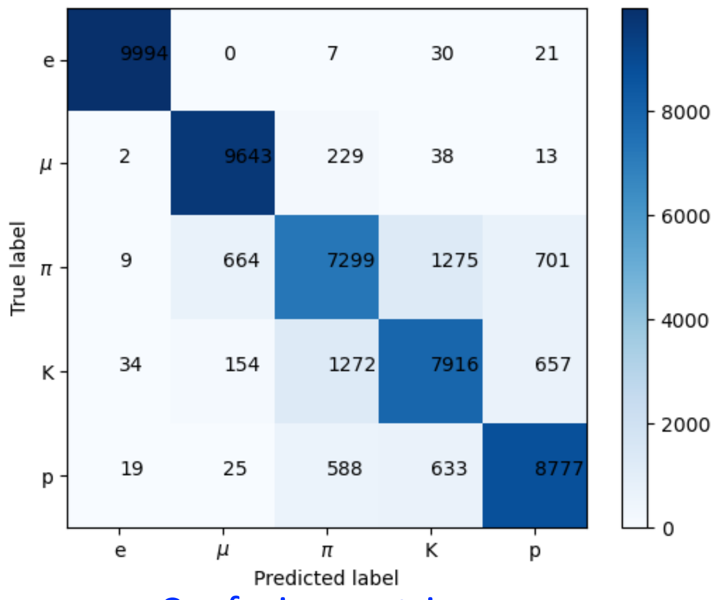
Global PID method

Yucong, ML based PID for STCF,
Software & Computing Session, Jan. 17

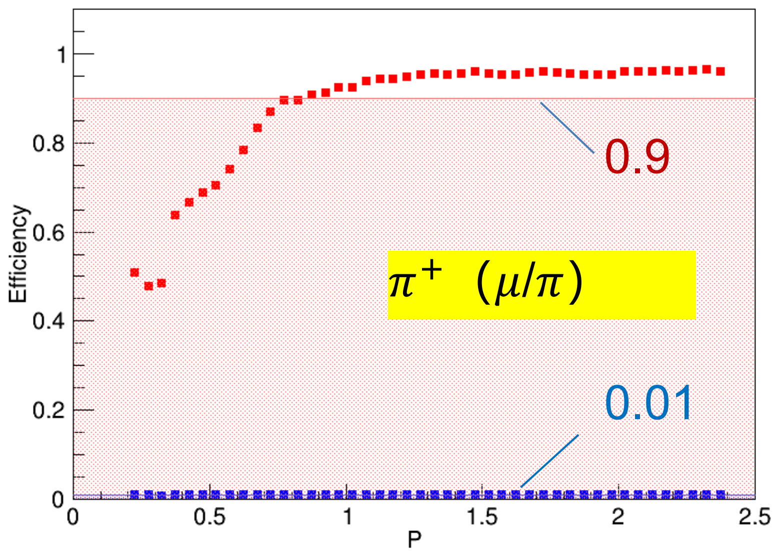
- A global PID algorithm is developed based on ML and selected 45 features from all sub-detector
  - Tracker/dEdx/RICH/DTOF/ECAL/MUD
- Trained the ML model based on XGBoost for better separation power in different modes: e/μ/π/K/P; π/K/p; e/π/K; π/K; μ/π
- The model and algorithm are provided for physics analysis within OSCAR



Features Selection



Confusion matrix



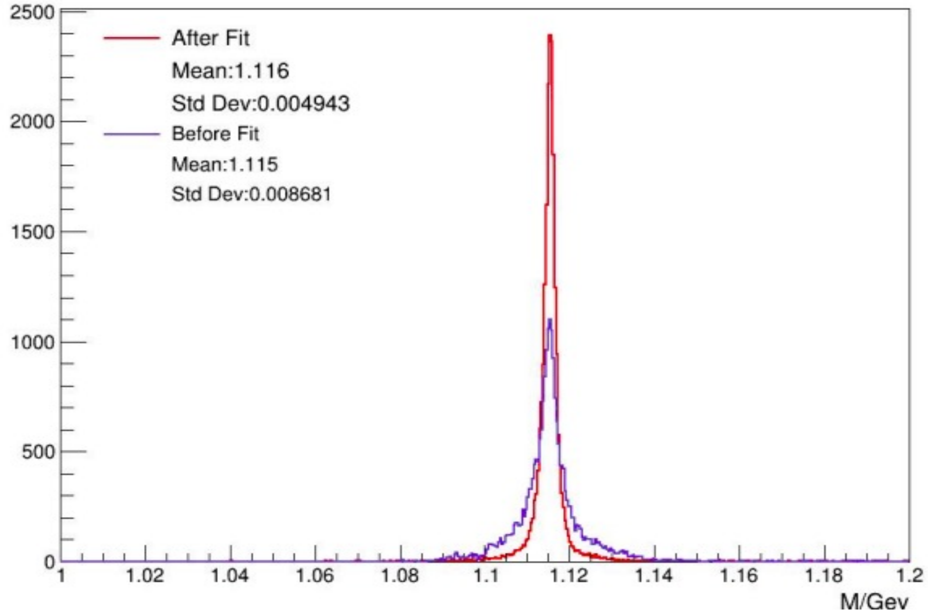
Pion efficiency & mis-identification

Analysis tools at STCF

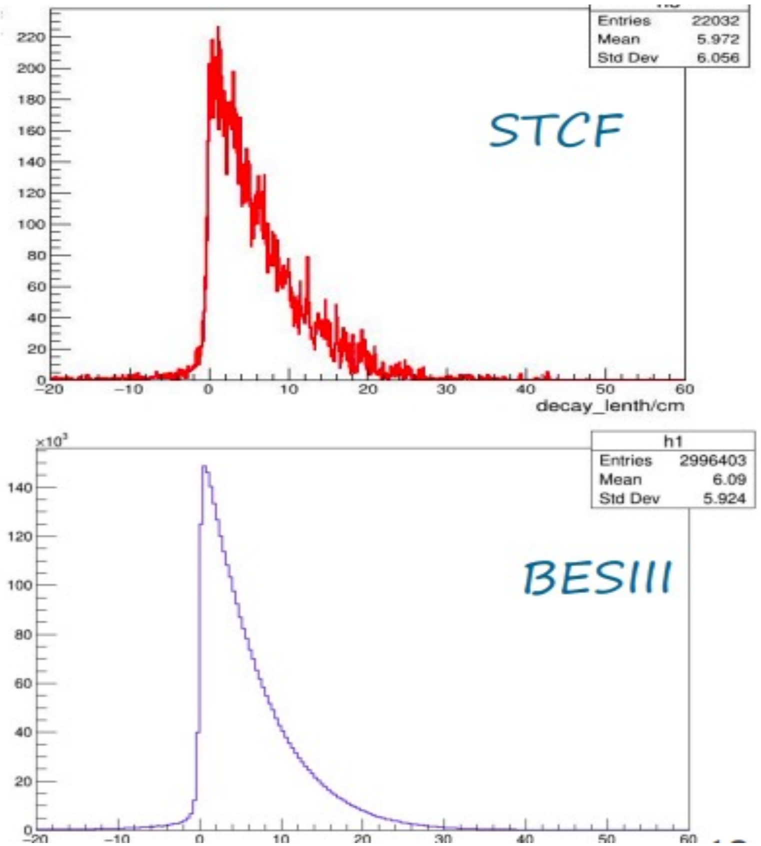
Vertex fitting

- The Vertex fitting and Kinematic fitting tools are developed based on the ones of BESIII after some modifications according to STCF
  - Two methods are used :Lagrange multiplier and Kalman filter
- Both are validated with STCF M.C. data and provided for physics analysis

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



Invariant mass of Λ



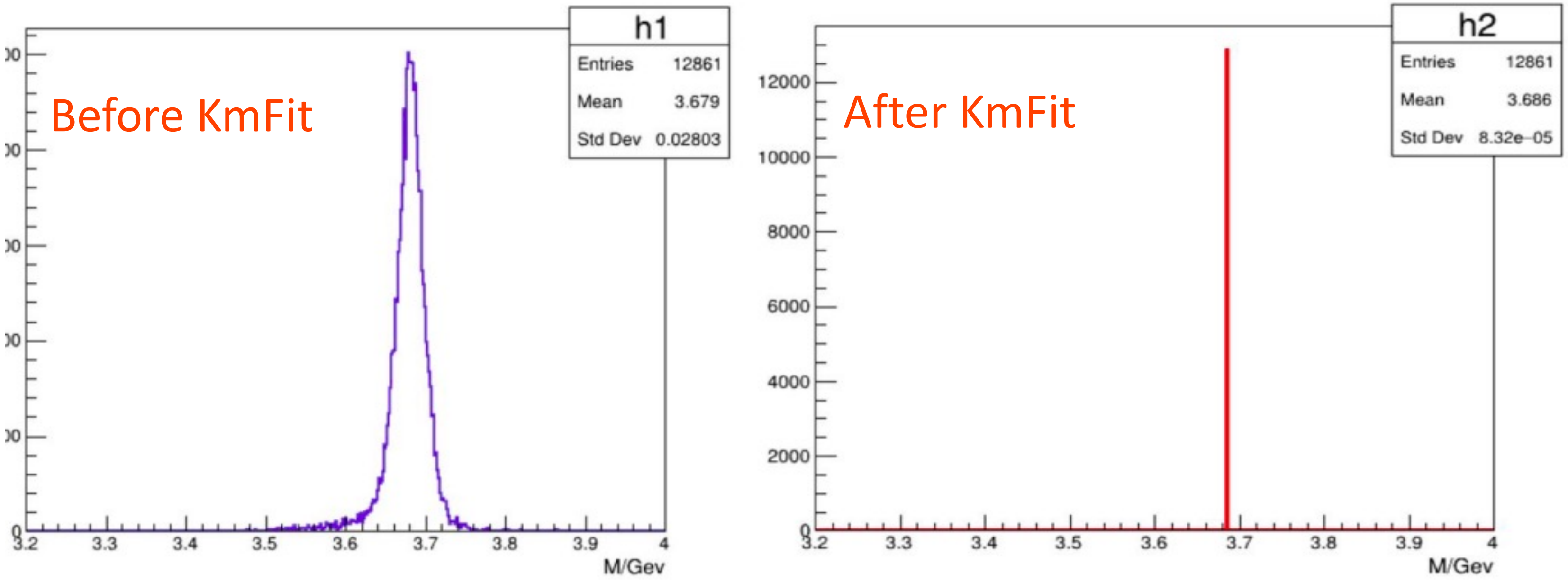
Decay length

44

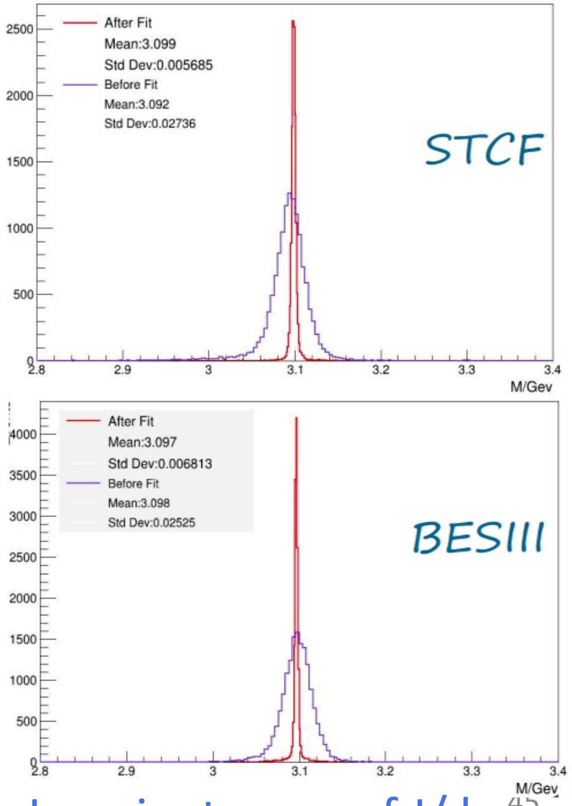
Vertex fitting

- The Vertex fitting and Kinematic fitting tools are developed based on the ones of BESIII after some modifications according to STCF
  - Two methods are used :Lagrange multiplier and Kalman filter
- Both are validated with STCF M.C. data and provided for physics analysis

$$\psi(2S) \rightarrow \pi^+\pi^- J/\psi(\rightarrow \mu^+\mu^-)$$



Invariant mass of ψ(2S)



Invariant mass of J/ψ

Summary

- STCF is an important and unique platform for probing physics at τ-charm sector
- The STCF offline software system (OSCAR) is being developed based on SNiPER since 2018, and now partially based on Key4hep (podio, dd4hep, edm4hep)
  - The core software is extended to meet requirements of STCF i.e. high luminosity
  - The full chain of the event simulation+ background mixing + digitization+ reconstruction +analysis is built
  - Ready to start physics analysis within OSCAR
- Future plan
  - Optimization of recon. algorithms, especially with digitization information after event mixing
  - Focusing on improving OSCAR’s computing performance to meet high luminosity requirements

Thanks for your attention !