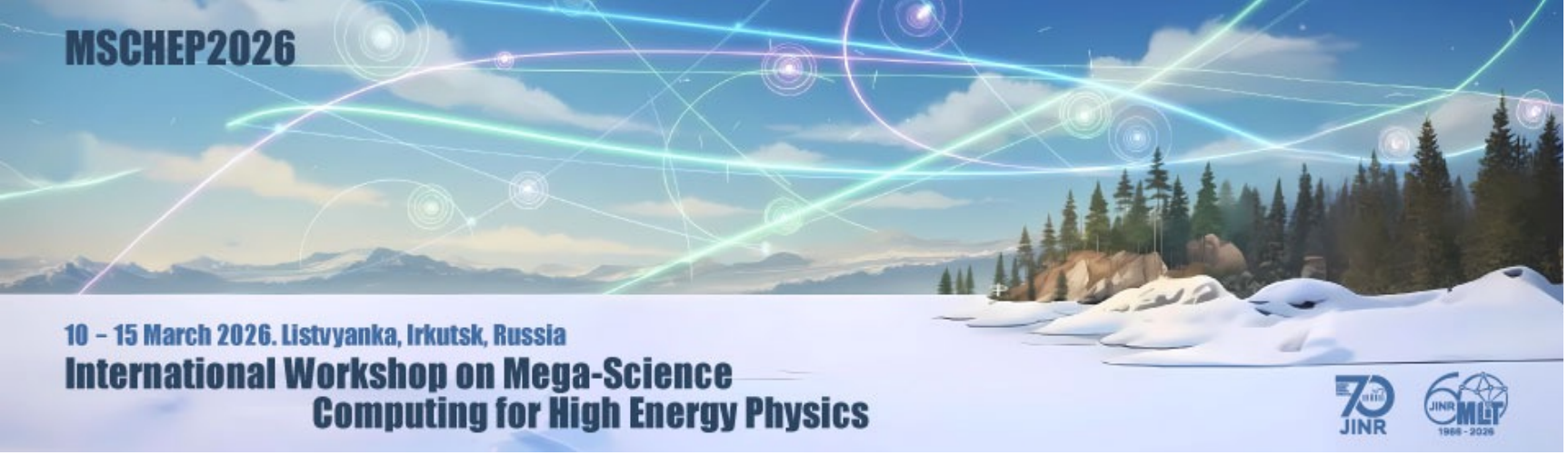


超级陶粲装置
Super Tau-Charm Facility

Design and Development of STCF Offline Software

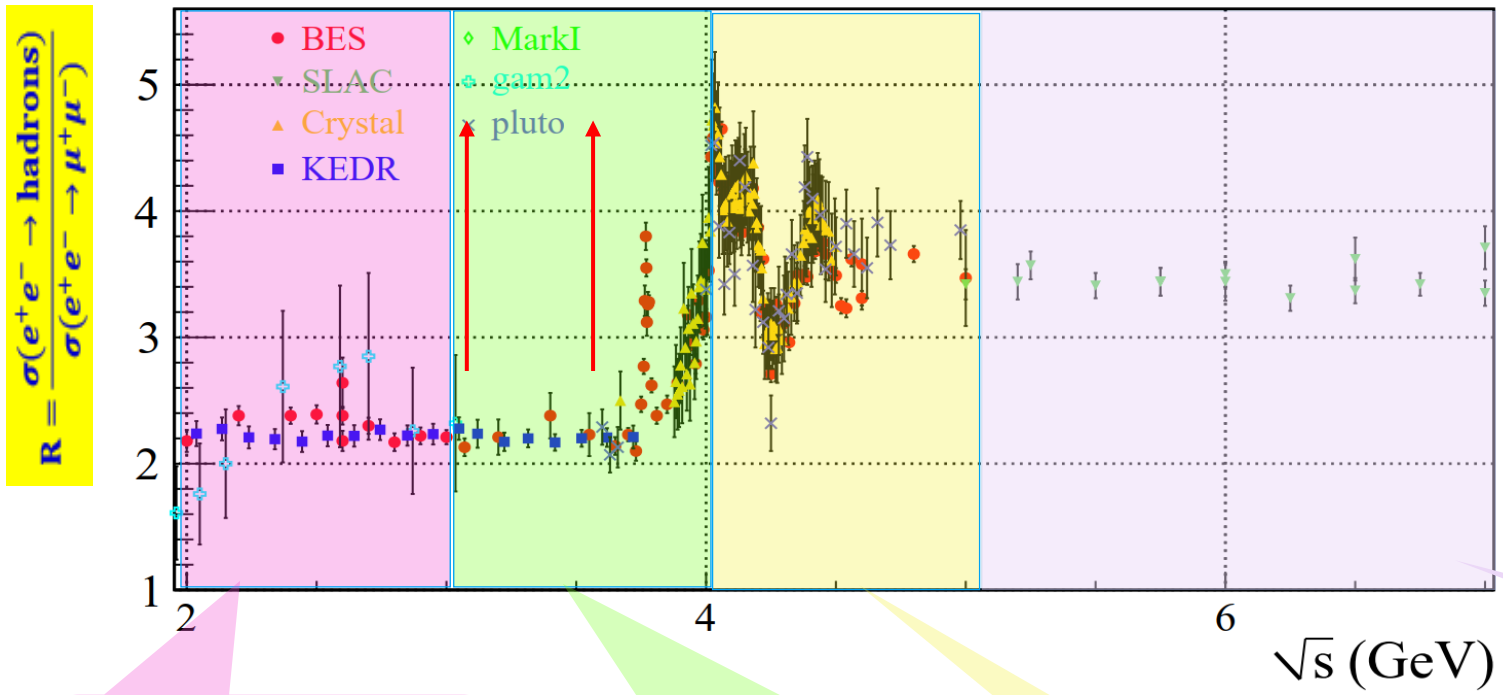
Xingtao Huang, Xiaocong Ai

On behalf of STCF Offline Software Group

Mar.10-16, 2026

Introduction

Tau-Charm Energy Region



Physics flagships

- Exploring QCD and confinement
- Testing fundamental symmetries
- Making precision measurement of key physical parameters

- Hadron form factors
- $Y(2170)$ resonance
- Mutltiquark states with s quark
- R value / g-2 related

- Light hadron spectroscopy
- Gluonic and exotic states
- Processes of LFV and CPV
- Rare and forbidden decays
- Physics with τ lepton

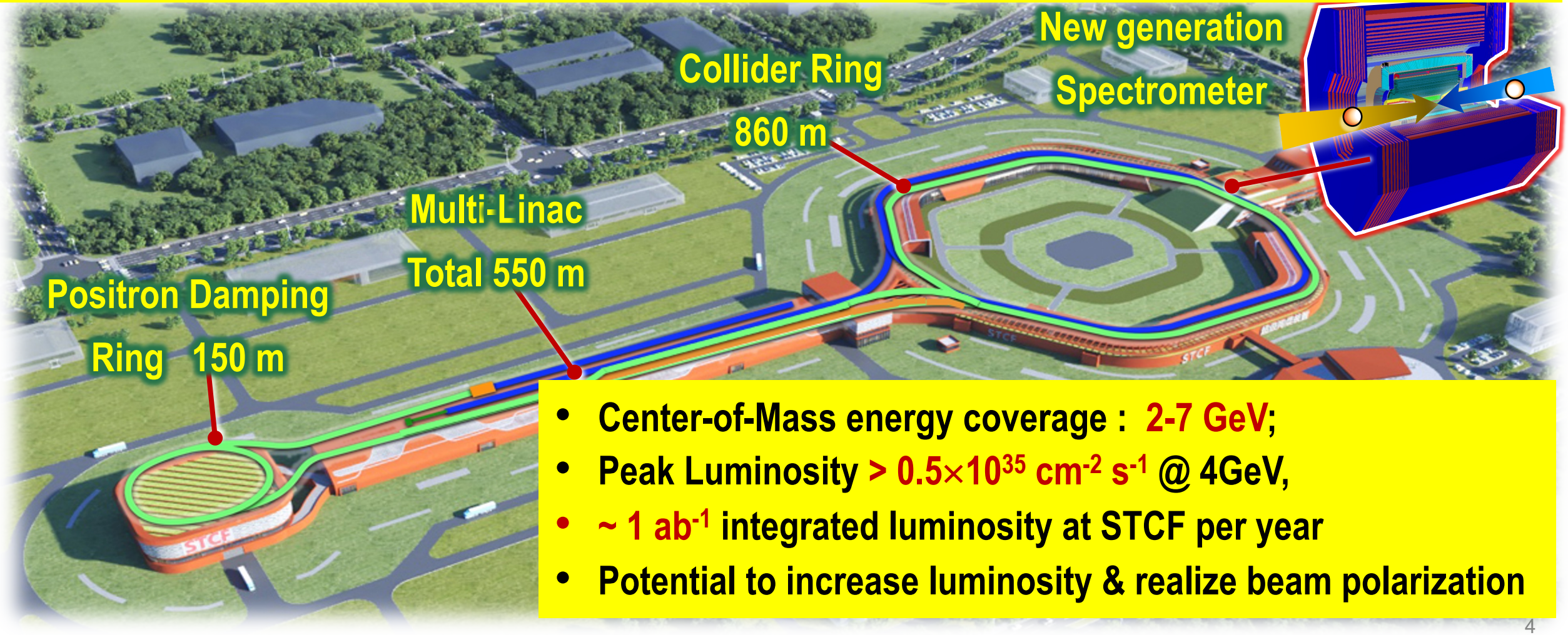
- XYZ particles
- Physics with D mesons
- f_D and f_{D_s}
- $D^0 - \bar{D}^0$ mixing
- Charm baryons

- Complete XYZ family
- Hidden-charm pentaquarks
- Search for di-charmonium states
- More charmed baryons
- Hadron fragmentation

Abundant Physics Topics in Tau-Charm Energy Region

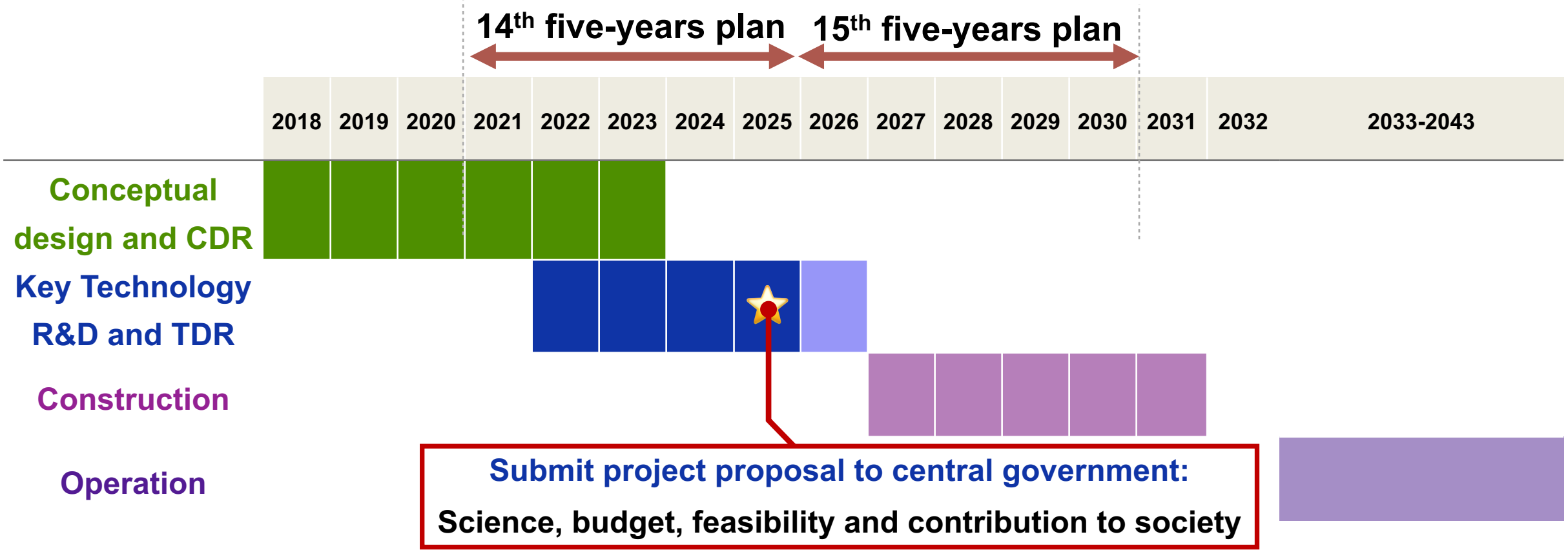
Super Tau Charm Facility (STCF)

A factory produced massive tau lepton and hadrons, to unravel the mystery of how quarks form matter and the symmetries of fundamental interactions



- Center-of-Mass energy coverage : 2-7 GeV;
- Peak Luminosity > 0.5×10^{35} cm^{-2} s^{-1} @ 4GeV,
- ~ 1 ab^{-1} integrated luminosity at STCF per year
- Potential to increase luminosity & realize beam polarization

Project Schedule



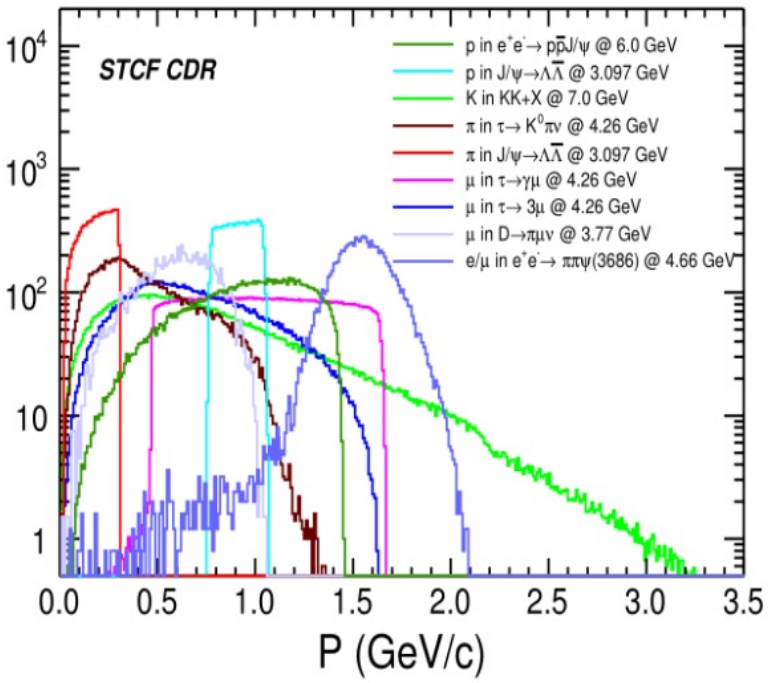
- 14th five-years plan : Conceptual design and R&D of Key technology, 364 M CNY
- 15th five-years plan : Construction 5 years, ~5 B CNY
- Operating for 10-15 years, upgrade for 3 years, operating again for ~10 years

Requirements

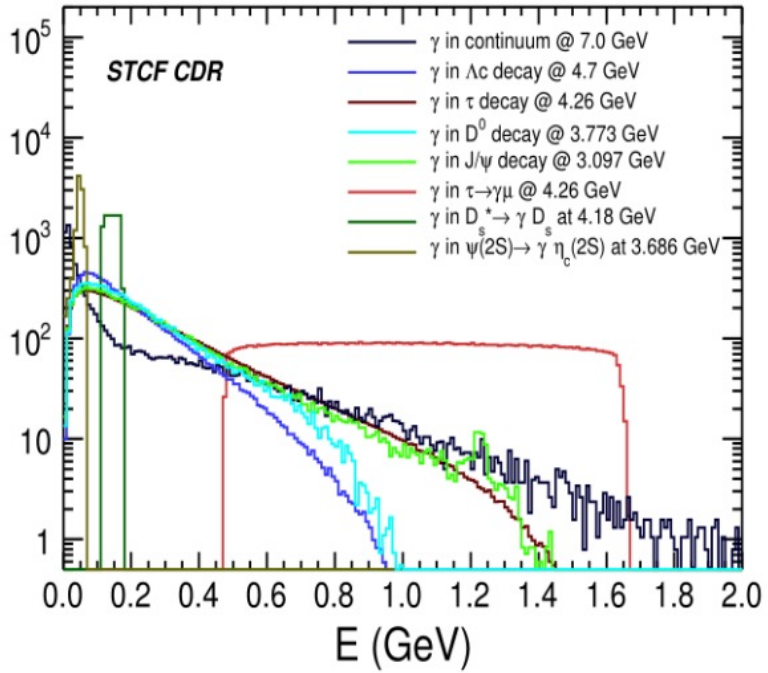
Particles at STCF

- 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 **KL, neutron** (up to 1.6 GeV)

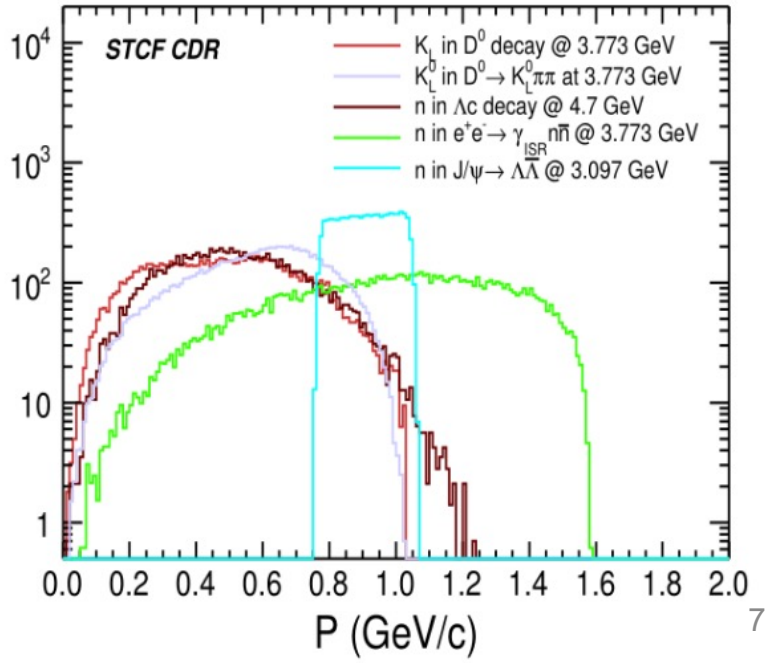
Charged particle momentum



Photon energy



K_L, n momentum



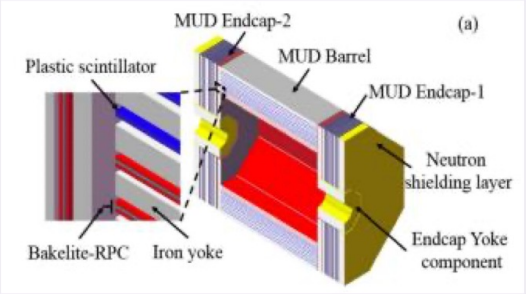
Detector Design

Core design goal:

- Efficient and precise reconstruction of exclusive final states
- Precise measurement of low-p particles (< 1 GeV)
- Outstanding PID: π/K and μ/π separation up to 2 GeV

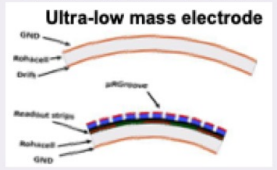
Muon detector

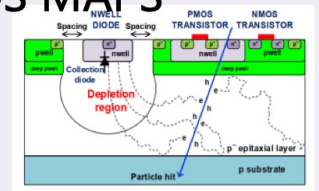
- RPC + scintillator strips

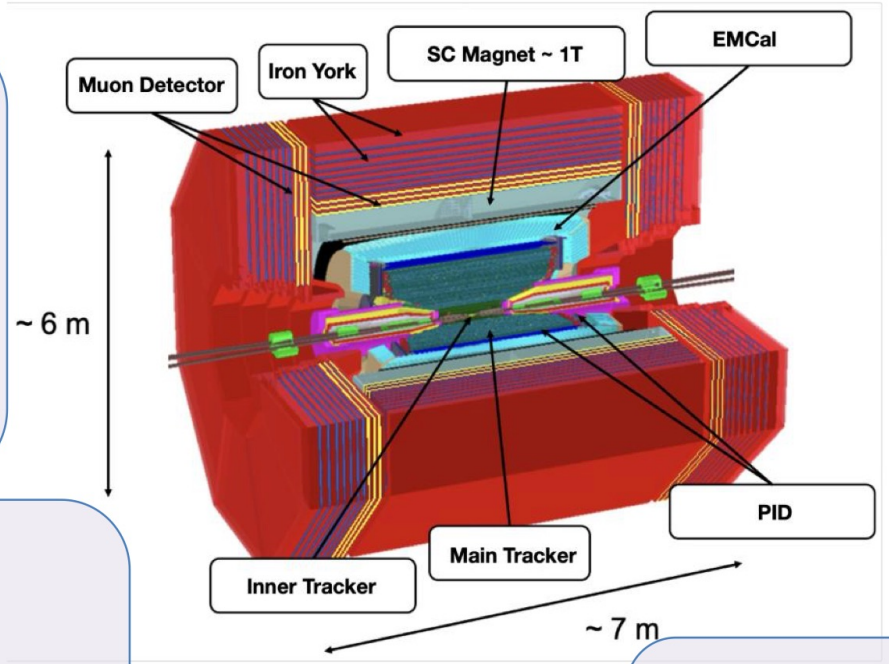


Inner tracker

- OPT1-MPGD: cylindrical MPGD (μRGroove)
- OPT2-Silicon: CMOS MAPS

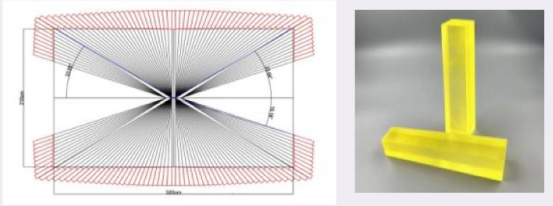






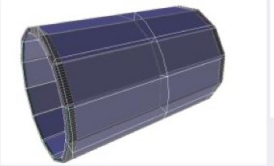
EMCal

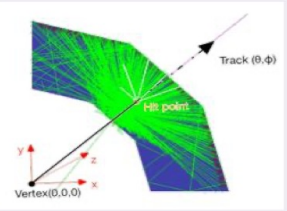
- Pure Cesium iodide (CsI) + APD



PID

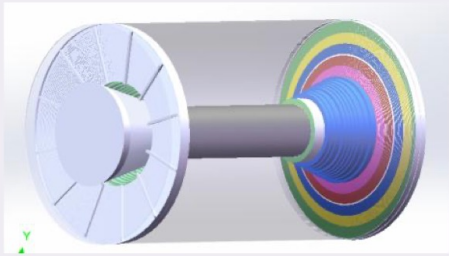
- Endcap: DIRC-like TOF (DTOF)
- Barrel:
  - OPT1-DTOF
  - OPT2-RICH





Main tracker

- Drift chamber



Requirements of offline software

- **Higher event rate, background, CPU consumption** at STCF than BESIII
  - develop **highly accurate and efficient** detector simulation and digitization 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

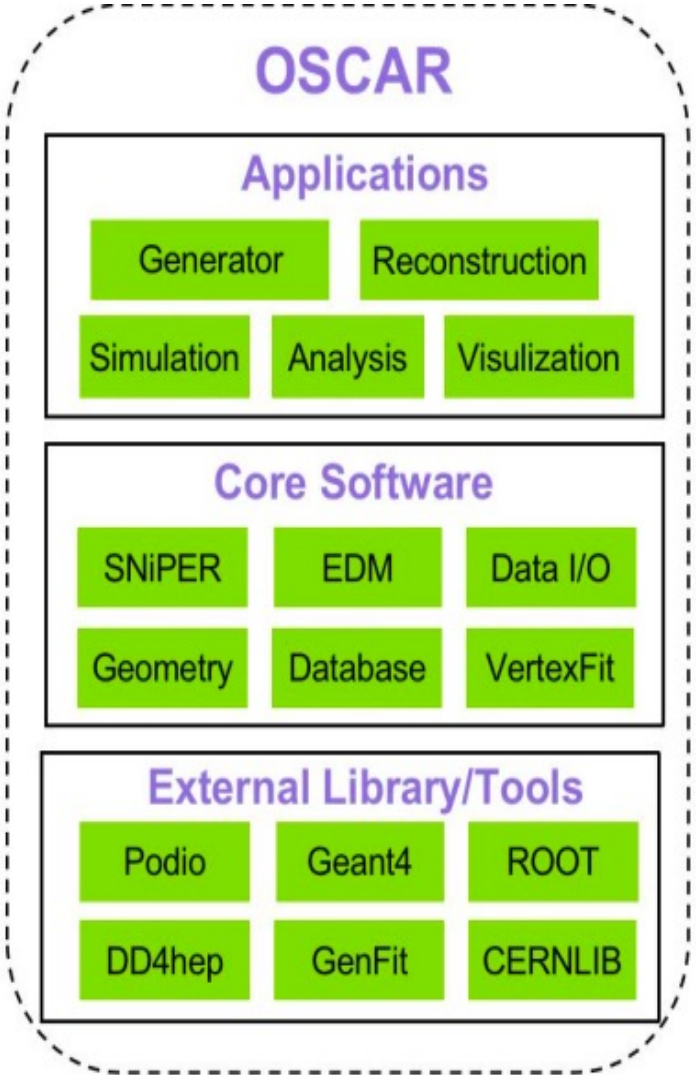
Offline Software

Core Software

Application Software

Offline Software of Super Tau-Charm Facility (OSCAR)

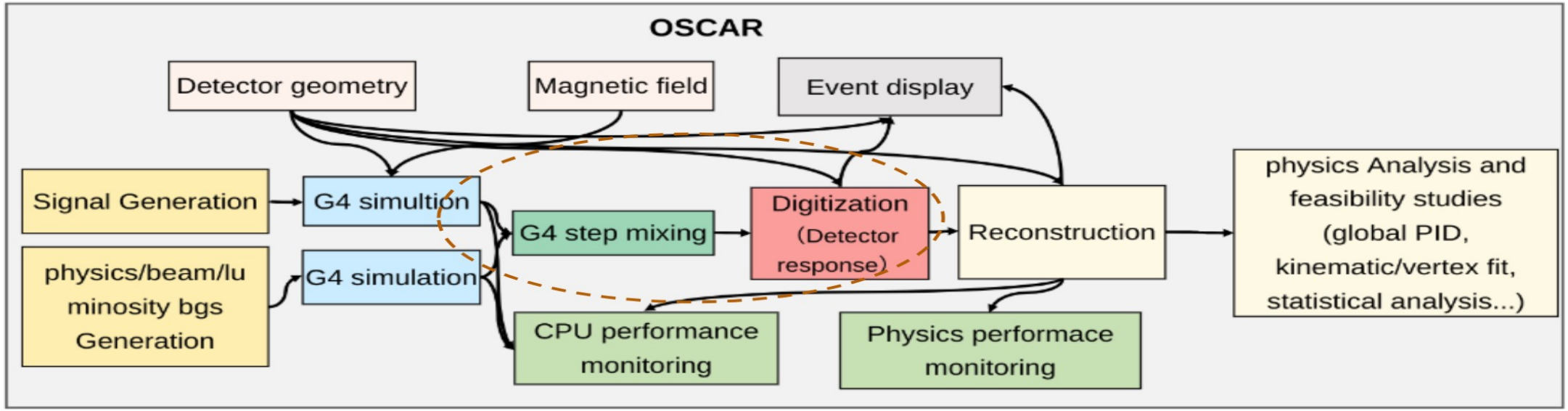
- Designed since 2018 for detector design, MC data production and physics analysis at STCF:
  - **External libraries and tools**
    - DD4hep, Podio, TBB, ONNEX,Geant4, ROOT, etc.
  - **Core software** providing common functionalities for
    - The underlying framework: SNiPER
      - Event loop control (sequentially and concurrently)
    - Detector data and event data management
    - supporting efficient parallel and heterogeneous computing
  - **Application Software:** STCF specific parts, **Most** developed from scratch, **Some** ported from BESIII
    - Generator,
    - Detector simulation
    - Digitization
    - Event Mixing
    - Reconstruction
    - Analysis Tools



W.H. Huang et al 2023 JINST 18 P03004

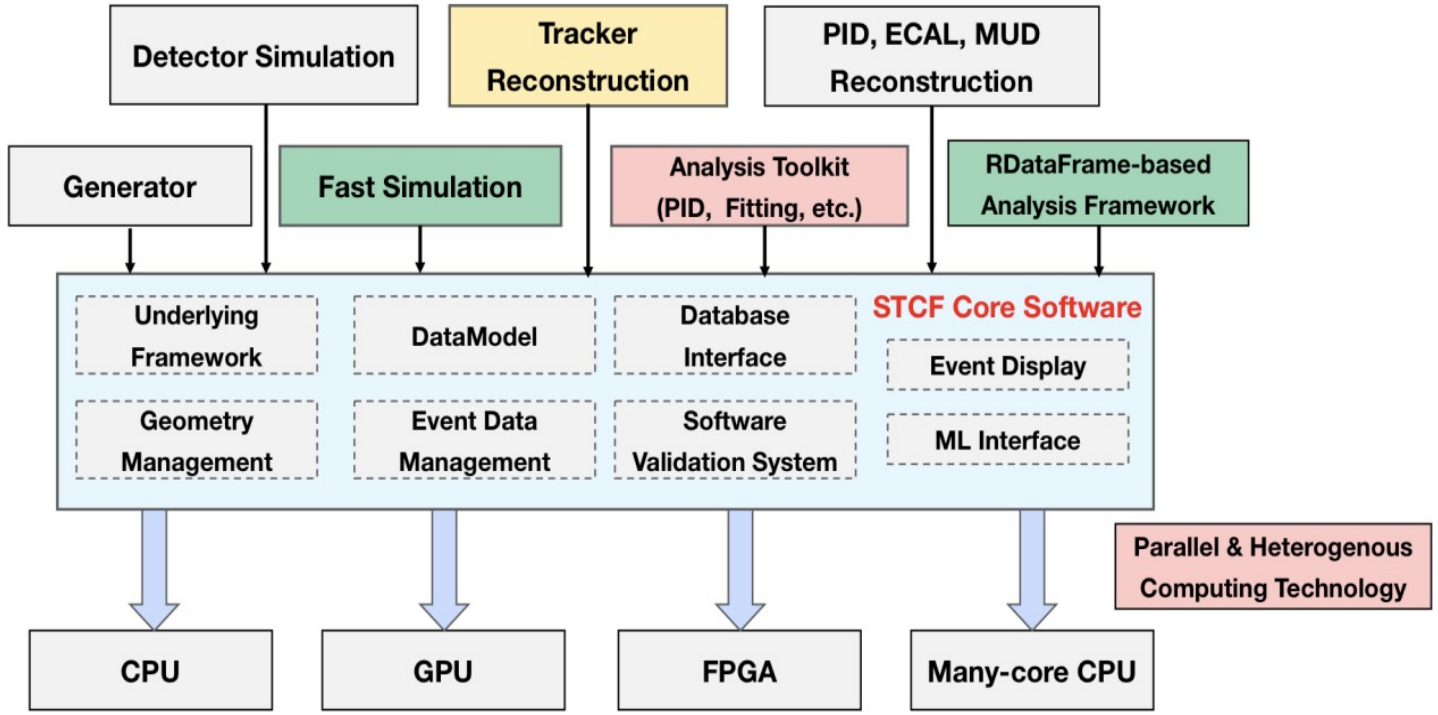
STCF offline software (OSCAR)

- Latest version, **OSCAR.2.7.0**, was released in Feb. 2026
  - Established the full chain of simulation + digitization + reconstruction + analysis
    - Speed: ~ **1 sec/evt** for Psi(2S) -> pi+pi- J/psi
    - Event size: **100 kb/evt**
  - Supported **detector simulation**, **reconstruction**, **calibration** and **physics analysis**



OSCAR Software Features

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

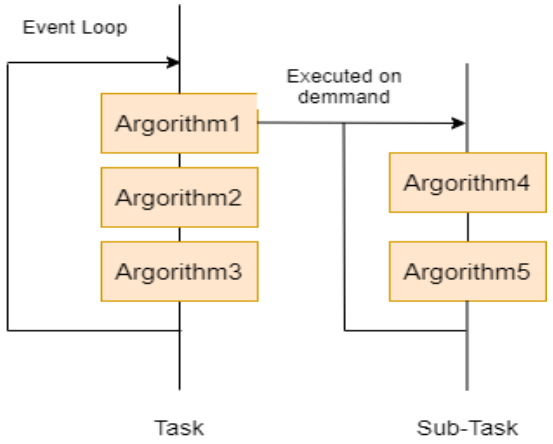


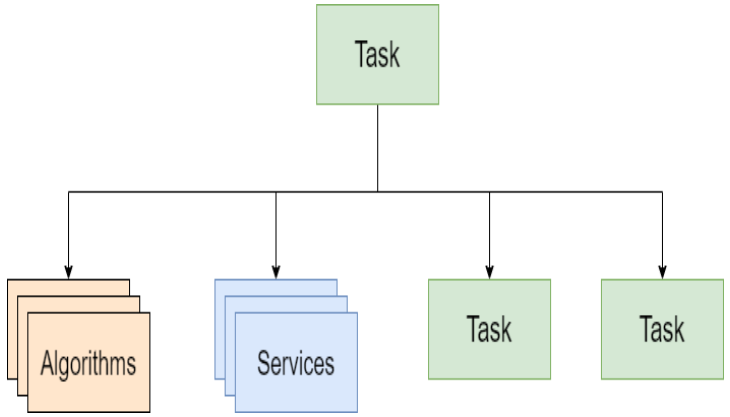
Comput.Softw.Big Sci. 9 (2025) 1, 3
Mod.Phys.Lett.A 39 (2024) 40, 2440006
JINST 18 (2023) 03, P03004
J.Phys.Conf.Ser. 2438 (2023) 1, 012054

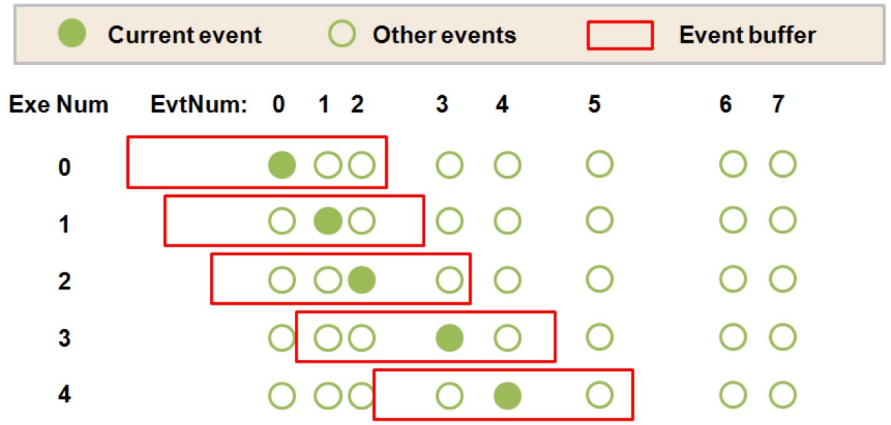
Core Software

SNiPER Framework

- SNiPER(Software for Non-Collider Physics Experiment) : a **general-purpose** software framework
  - **Light weighted**: only dependent on Boost. Python
  - **Highly modularized and extensible**
  - Originally developed for **JUNO**, now adopted by several experiments: **LHAASO, STCF, HERD, nEXO**, etc.
- Key components
  - **Algorithm:** an unit of event data proceeding
  - **Service:** an unit for common functions that can be called by users, anywhere when necessary.
  - **Task:** a lightweight application manager to assemble specific algorithms, service as well as sub-tasks.!

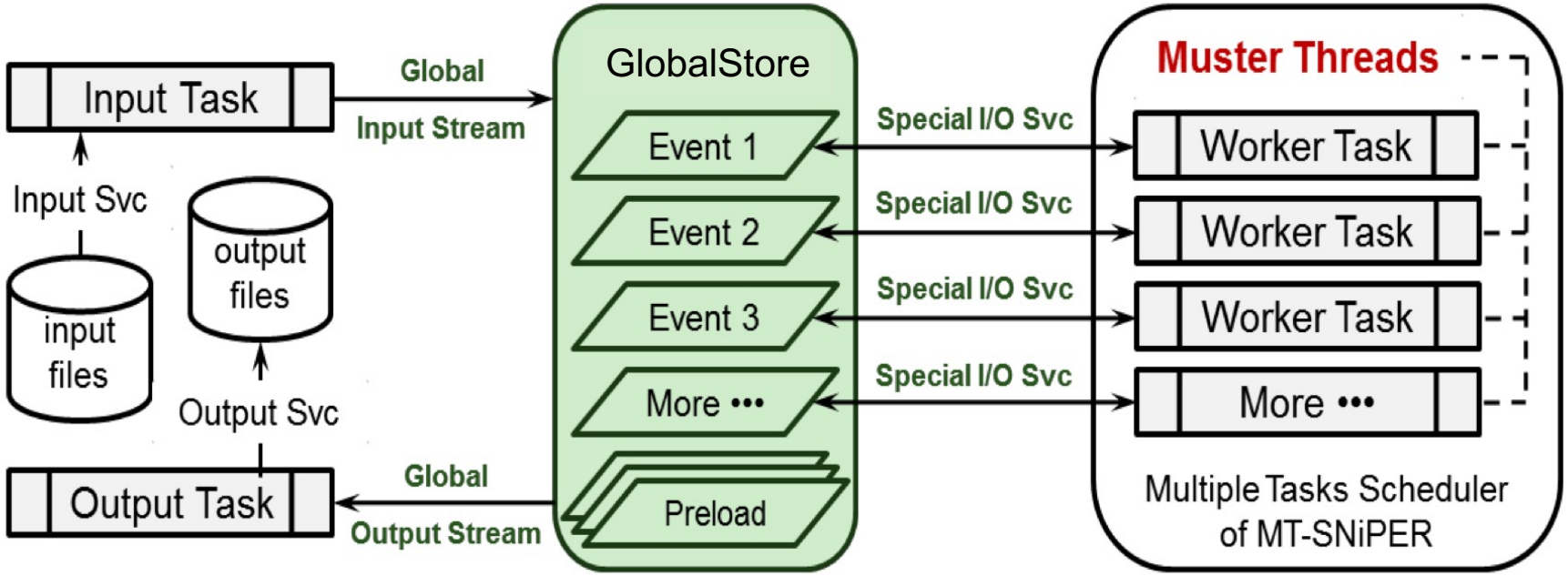






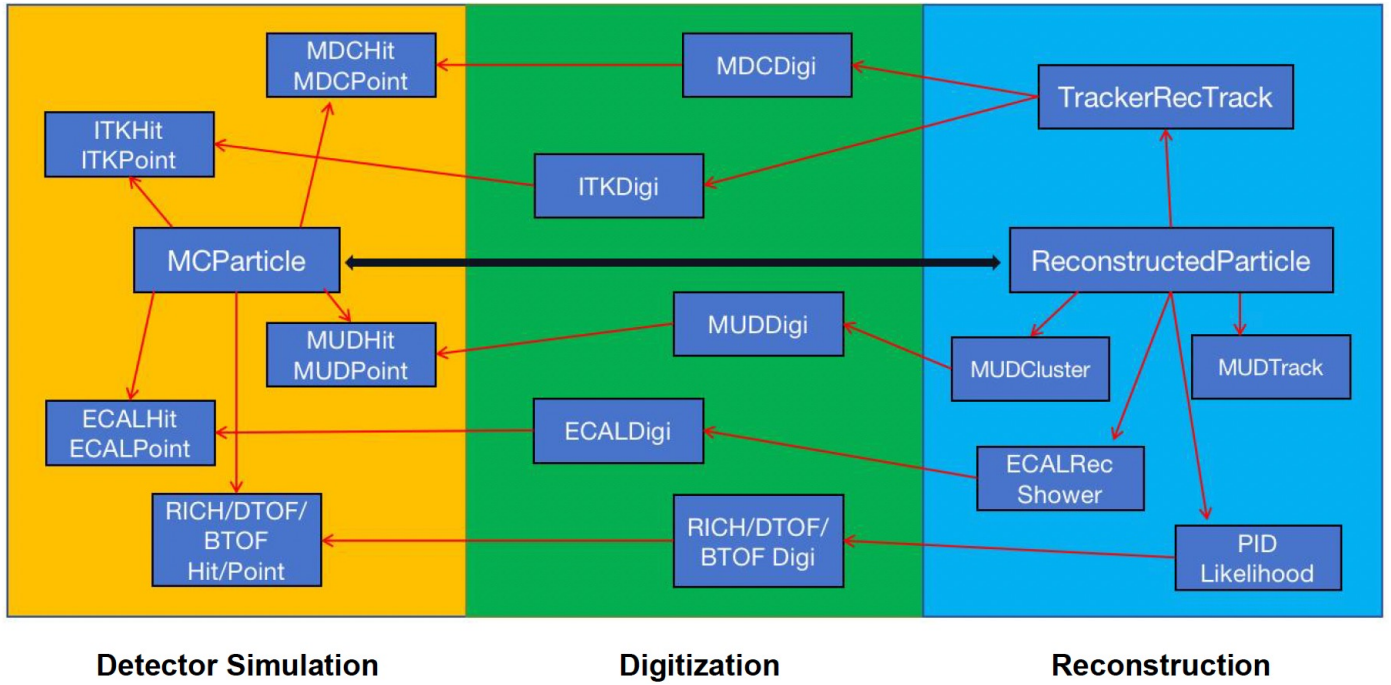
Parallelism in MT-SNiPER

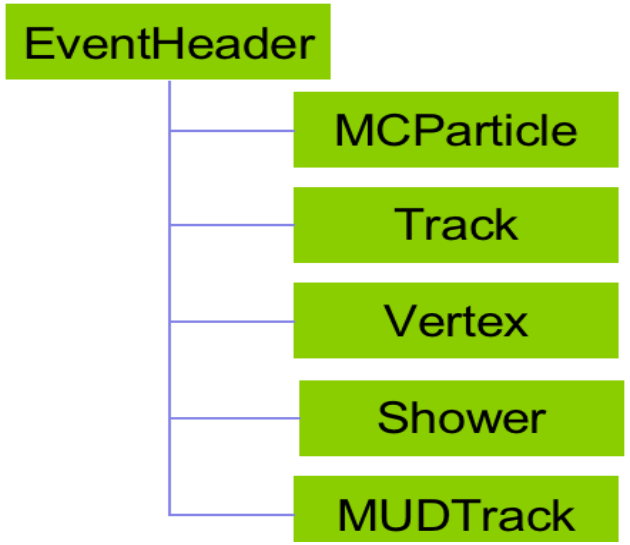
- SNiPER provides simple interfaces for building **multithreaded applications**
  - Based on Intel **TBB**
  - SNiPER **Muster** (Multiple SNiPER **Task Scheduler**) works as a thread pool/scheduler
  - A **Global Store** is developed to support **multithreaded event data management**
  - Data I/O is binded to **dedicated I/O** thread for flexibility
  - Application code is mostly consistent **for serial and parallel execution**



Serial Event Data Management

- Event Data Model (EDM) is a very crucial part and taking significant effect on performance of data 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

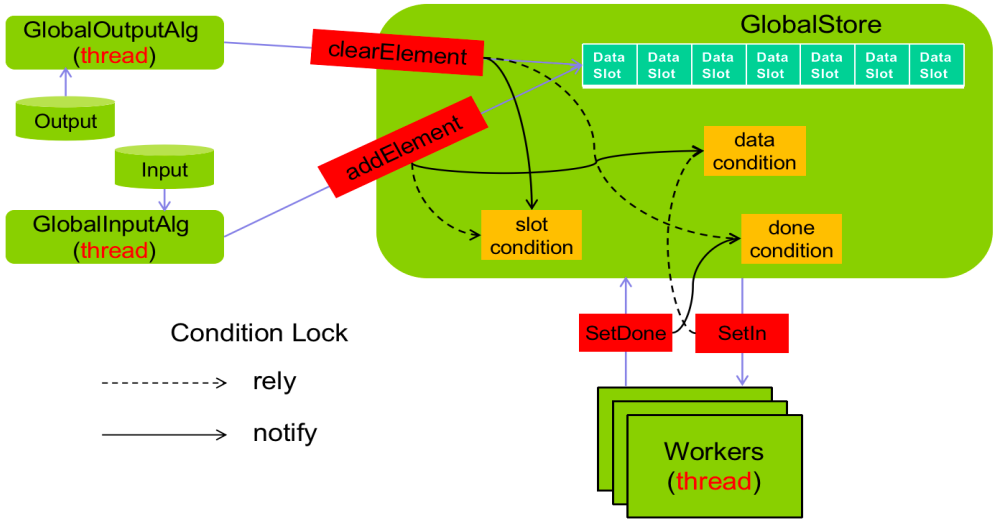


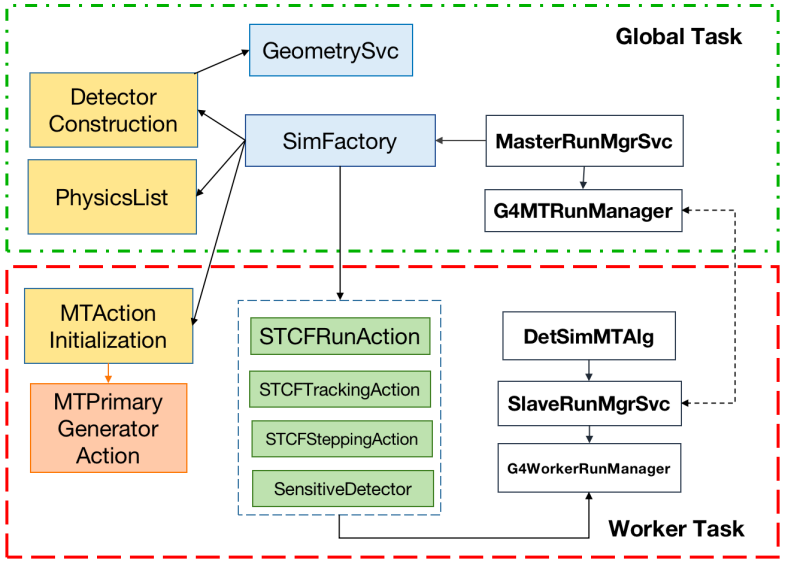


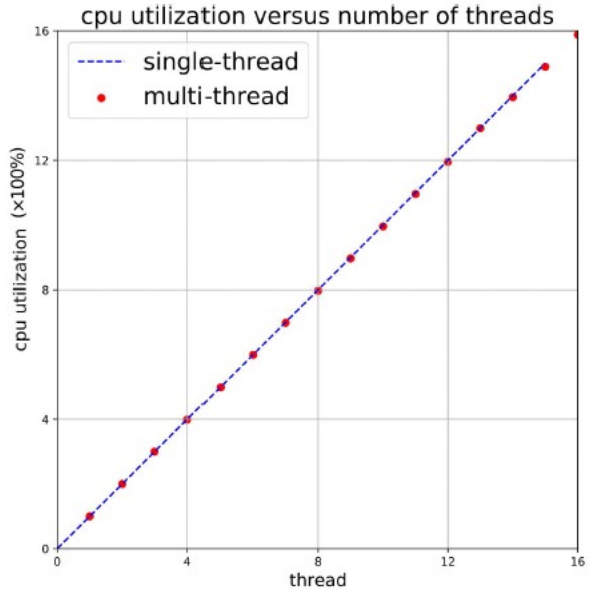
Analysis in flat structure

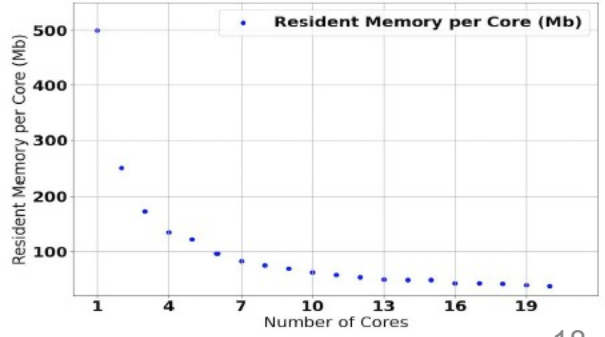
Parallelized Event Data Management

- Developed parallelized Event Data Management for parallel 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 successfully implemented
- **Track-level** and **algorithm-level** parallelism are **under R&D**



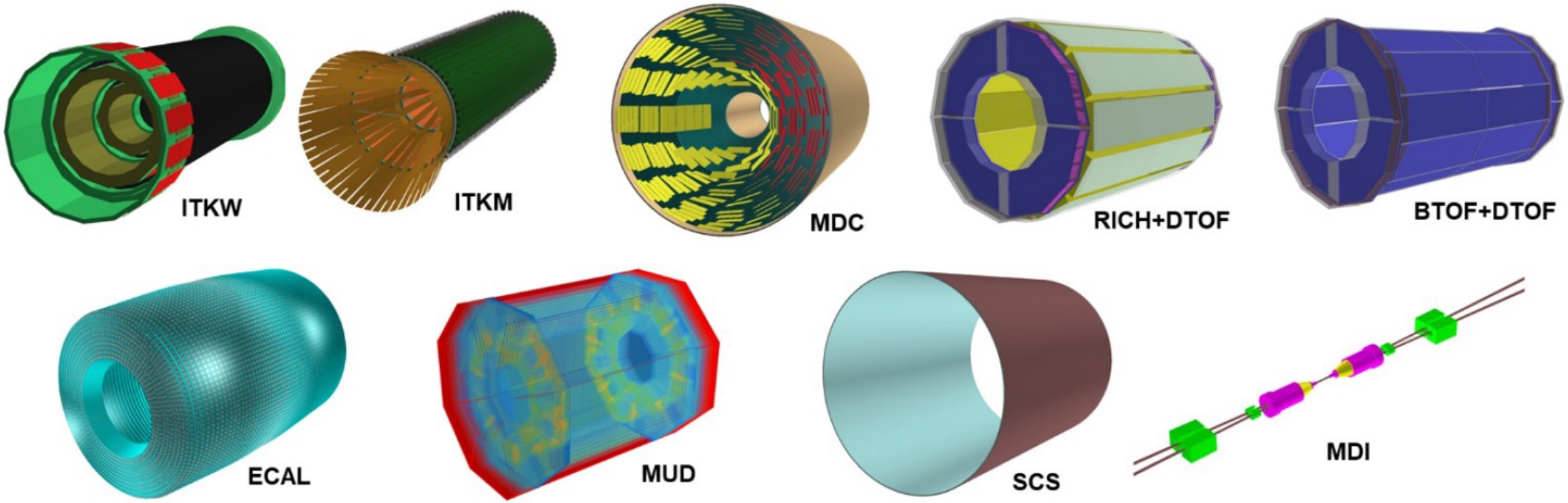






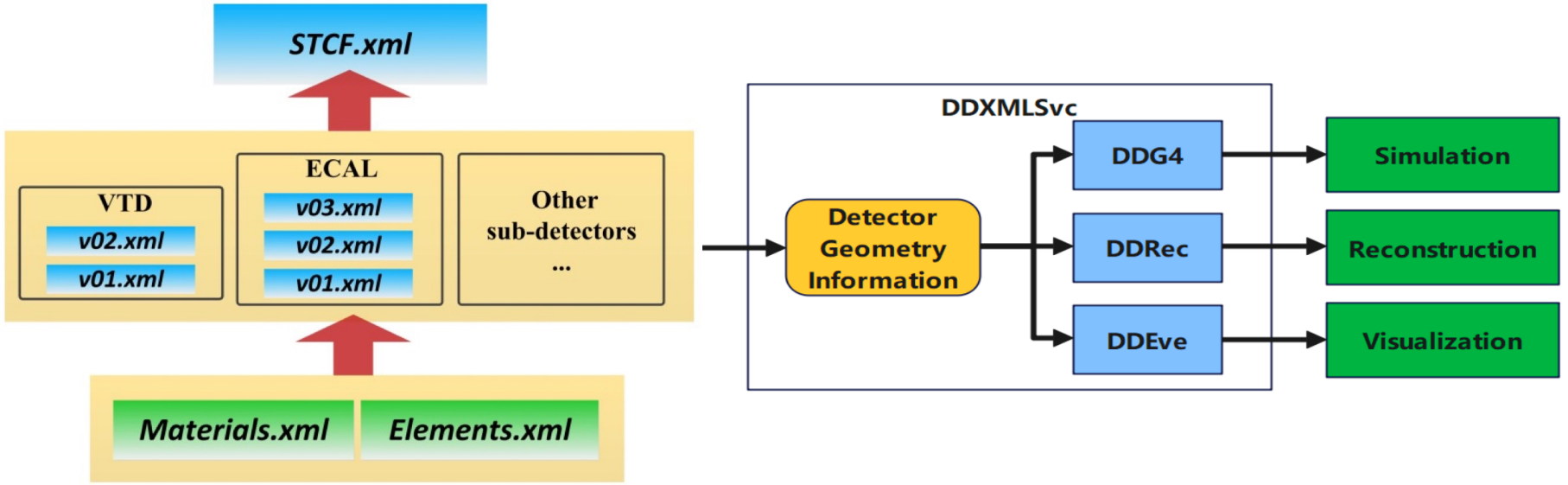
Detector Description

- The full STCF Detector is described with **DD4hep**
- Each sub-detector is implemented with **a single compact file**
  - Dimension, position, material, magnetic field , optics property



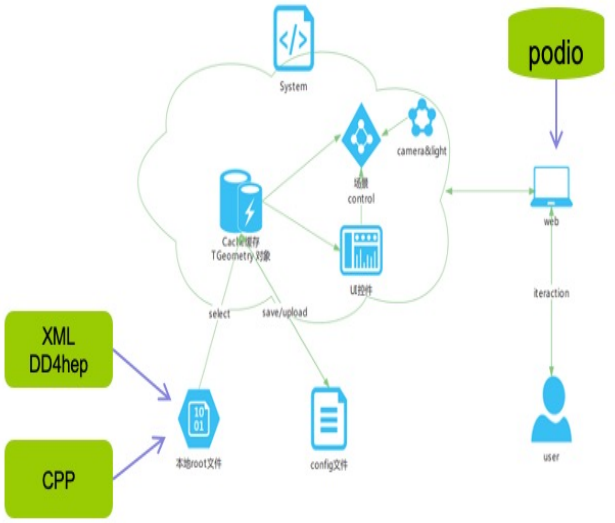
Geometry Management System

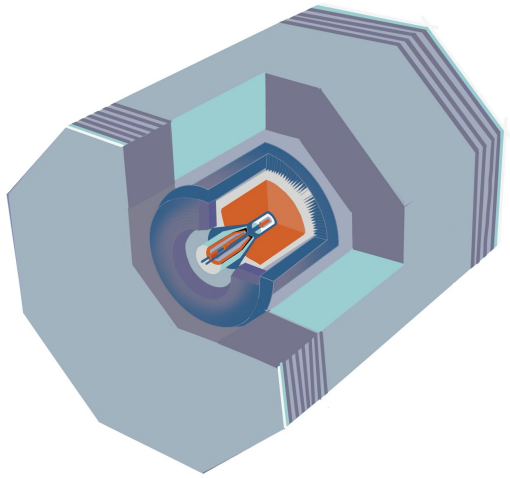
- Developed geometry management system for **all applications**
  - detector simulation, reconstruction and event display
  - The version number is used for **different design options**
- Ensure **consistent** detector information between different applications with a single source of detector description
  - **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

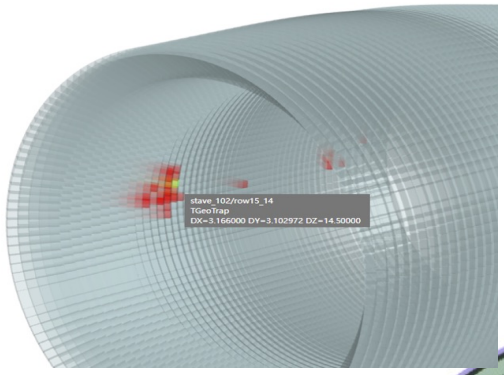


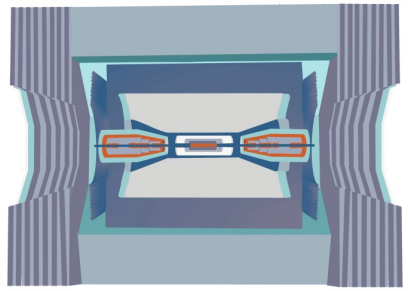
Detector/Event Visualization

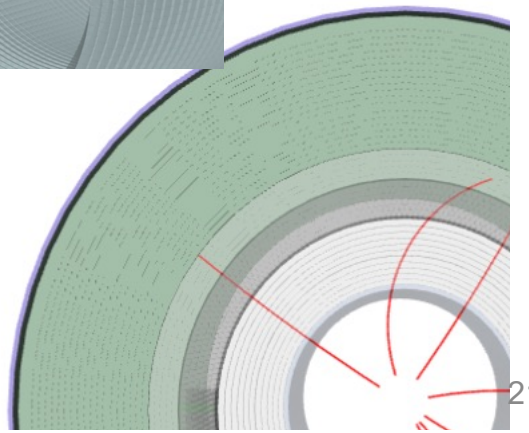
- A common geometry and event display system is **being developed** to visualize detector geometry, simulated hits and reconstructed particles
  - Based on **Web3D** technology and the open-source **JSROOT**
  - 3D engine and graphic library based on **Three.JS**
  - Using the Vue.js HTML5 development framework to implement the **Web interface**





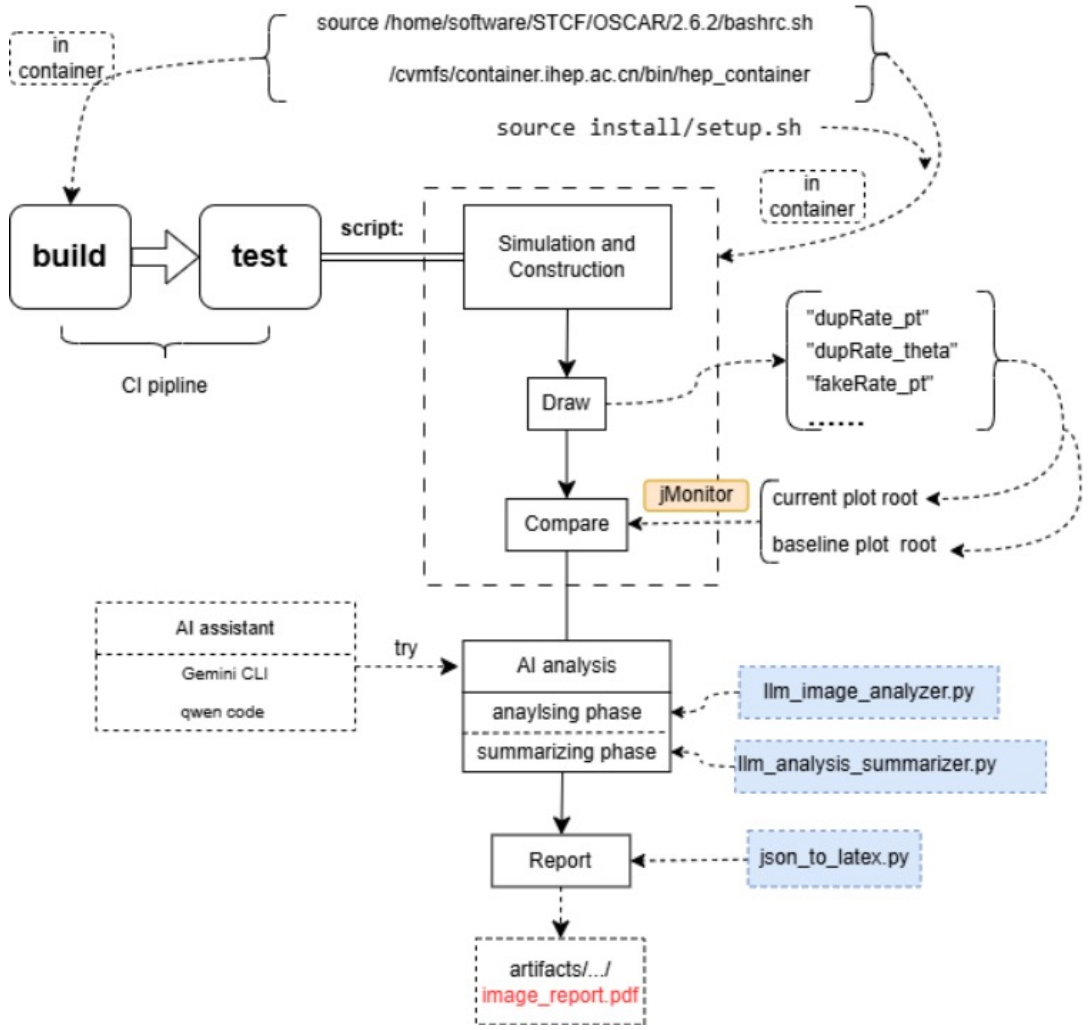






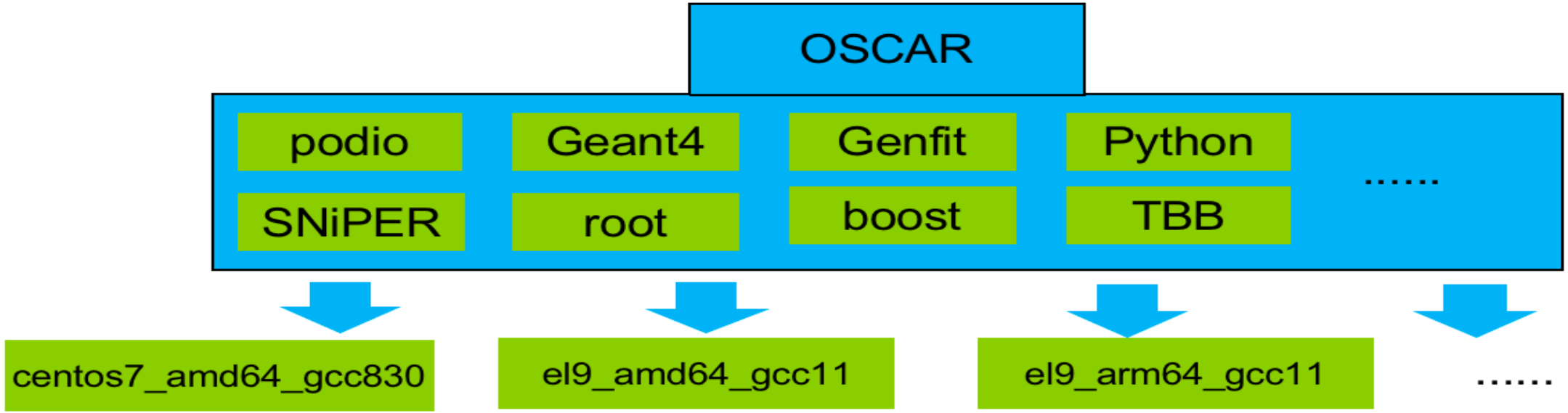
Software Validation powered by LLM

- A modern software validation system powered by **LLM** is developed
  - **Full-scale software validation** included:
    - Unit test
    - Integration test
    - Physics validation
    - Performance profiling
  - Integrated with **Gitlab CI for** automation
    - Support both **local machine** and **cluster** (HTCondor)
    - Integrated with **LLM backend** to analyze validation results



Software Deployment by Spack

- **Spack** is an open-source, cross-platform package manager designed for **scientific computing and HPC software deployment.**
  - Well supports multiple software versions and configurations on different platforms and compilers
  - Automatically handles versioning, depending, patching as well as platform problems
- It was adopted to deploy STCF offline software and achieve "**one-click installation**" regardless of platforms, processor architectures, versions, etc.



Application Software

Simulation

Digitization

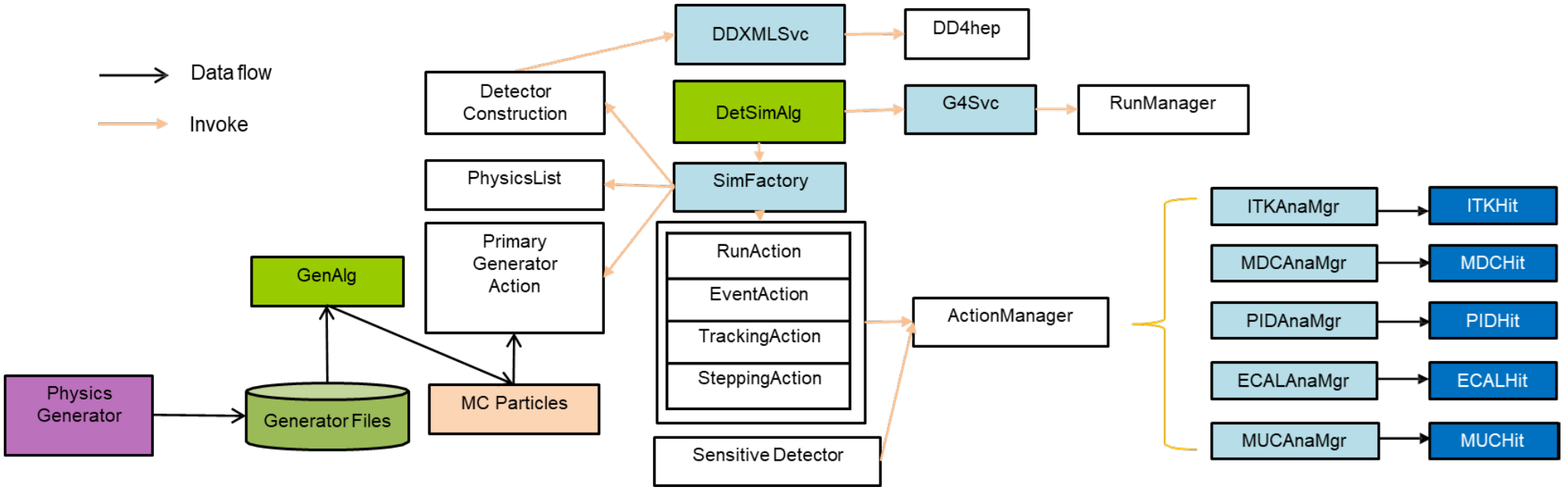
Event mixing

Reconstruction

Analysis

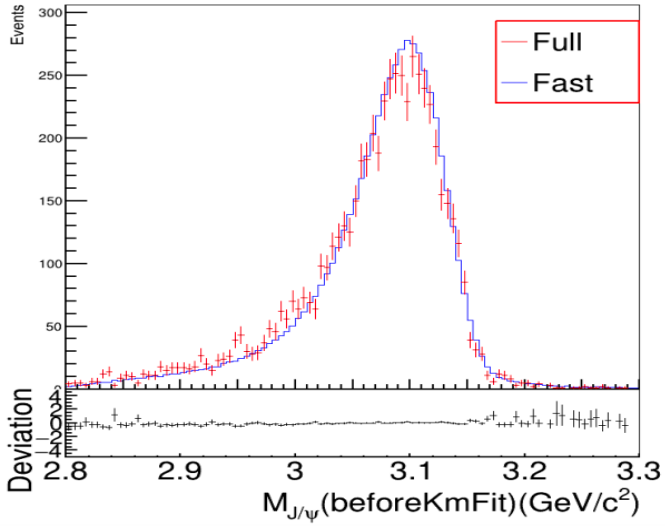
Full Simulation

- **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, G4Step level



Fast simulation

- General scheme based on **parameterization and sampling**
  - efficiency sampling
  - resolution realized based on CDF
  - detector responses extracted for different $(p, \cos(\theta))$ regions
  - feasible configuration for detector optimization and sensitivity study



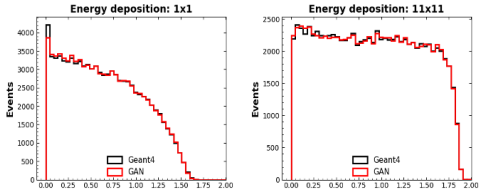
- Resource **consumption reduction**
  - time: about O(10^2) faster
  - storage: more than 10 times smaller

More details in Xiaoshuai's talk:
Fast simulation framework in STCF

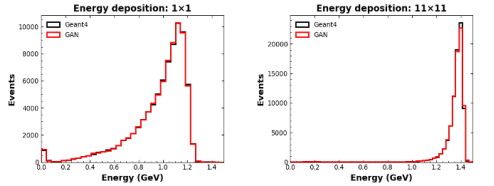
$\pi^+\pi^-\mu^+\mu^-$ (4 tracks)	Full Sim+Rec	Fast Sim
CPU time	~300ms(sim) + 800ms(rec) / event	~2ms / event
Event size	~77kb(sim) + 26kb(rec) / event	~2kb / event

for photons:

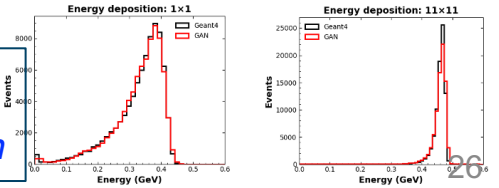
P = 0-2 GeV



P = 1.5GeV



P = 0.5GeV

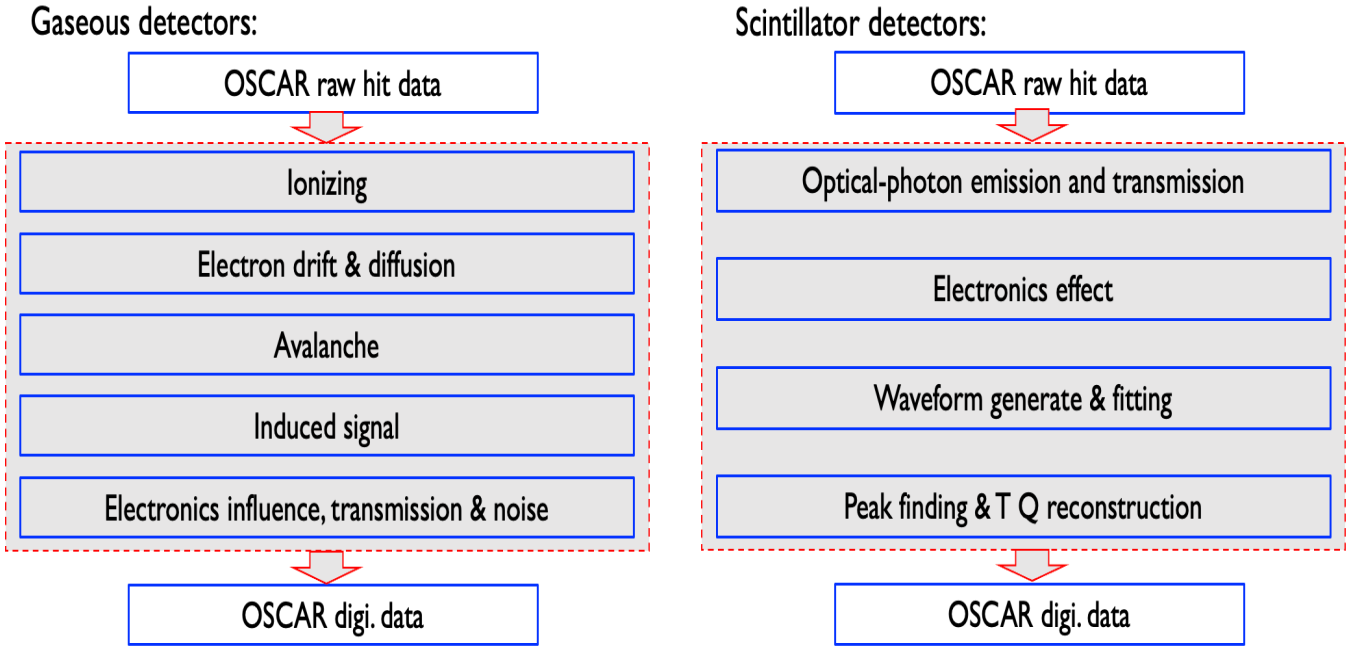


- An **EMC fast simulation** package is also developed based on **GAN** and achived good consistent results with Geant4.

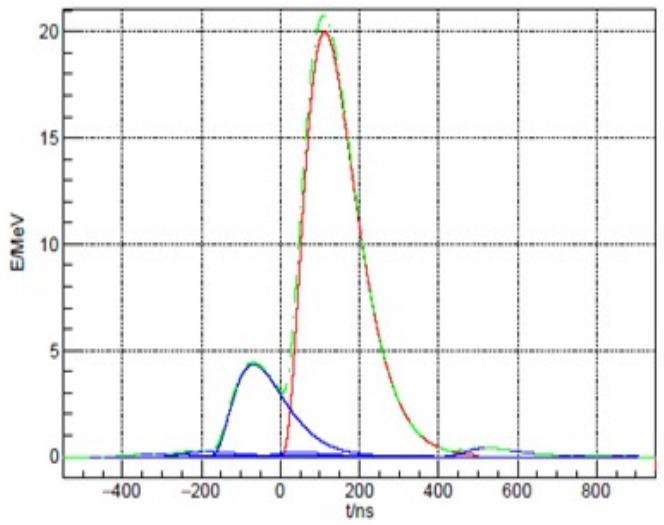
More details in Xiao's talk:
STCF Electromagnetic Calorimeter Fast Simulation

Digitization

- Developed an **unified digitization framework** for all sub-detectors on OSCAR
- Each sub-detector implemented its digi. algorithms and produced the digi. Information
  - The **sampling method** (Fast Digitization/Smearing) implemented for all sub-detectors
  - **Full digitization** are under optimizing with considering real processes
  - The **reconstruction algorithms are under tuning** with the digi. information as their inputs

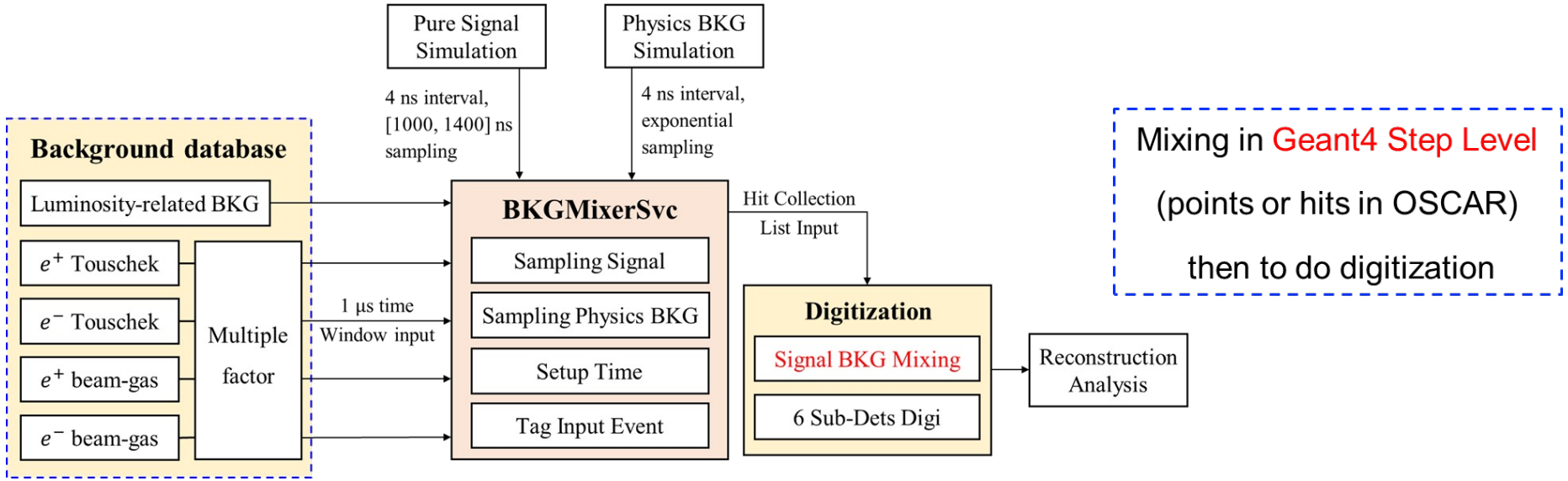


ECAL pipeline fitting



Background mixing

- Event mixing with background at the **Geant4 step level**
  - Simulated background particles as input, a unified algorithm applied to each sub-detector.
- MC hit data digitized to be **same as the detector measurement**
  - Considering **electronic response, noise and other effects,** as input for reconstruction and analysis

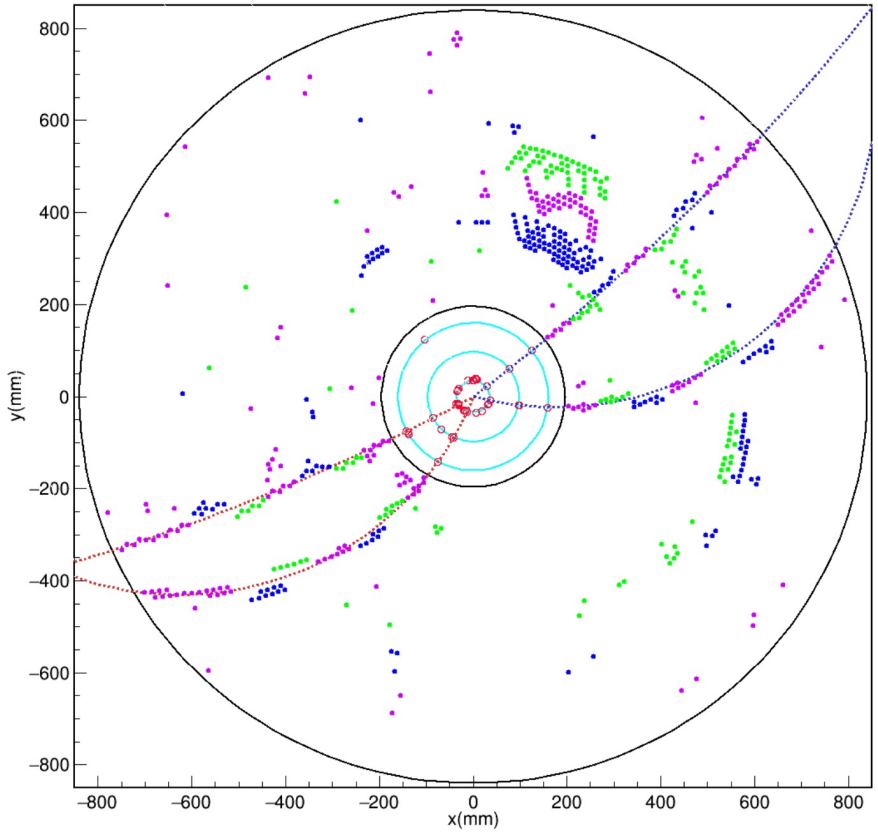


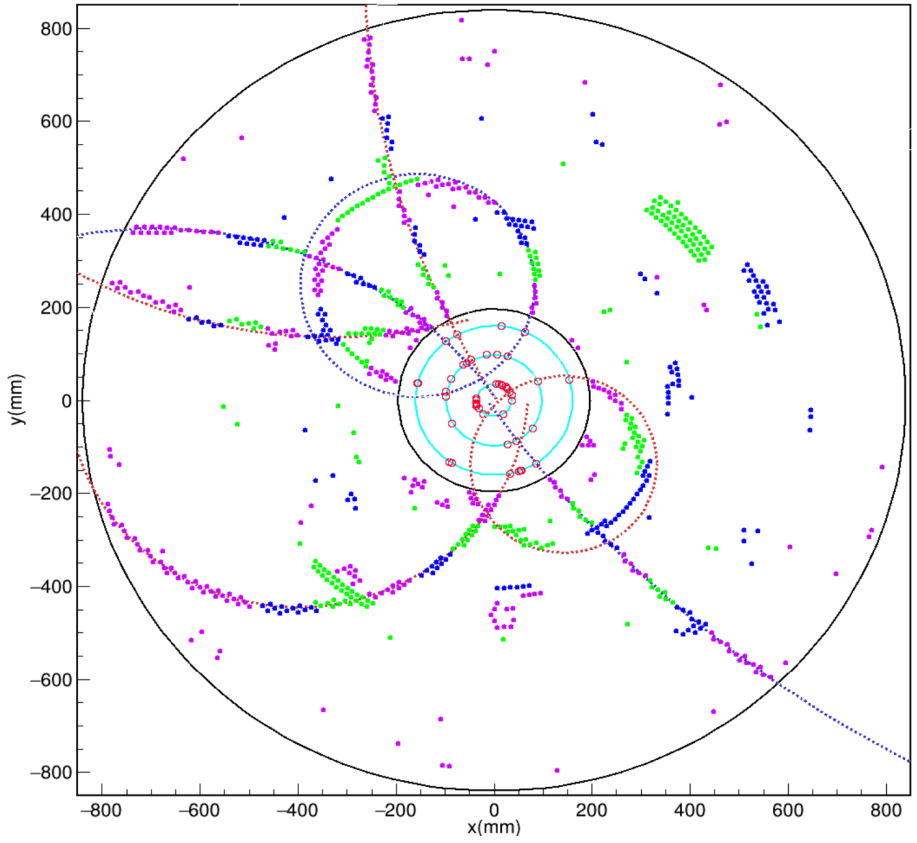
- Another Mixing Framework
  - Some or all hits **mixing in digi. hit level**, pre-digitize the Bkg MC hit to produce **a Bkg digi. Database**
  - **One order of magnitude time savings**, but the digi merge methods **is developing**

Tracking

STCF tracking environment

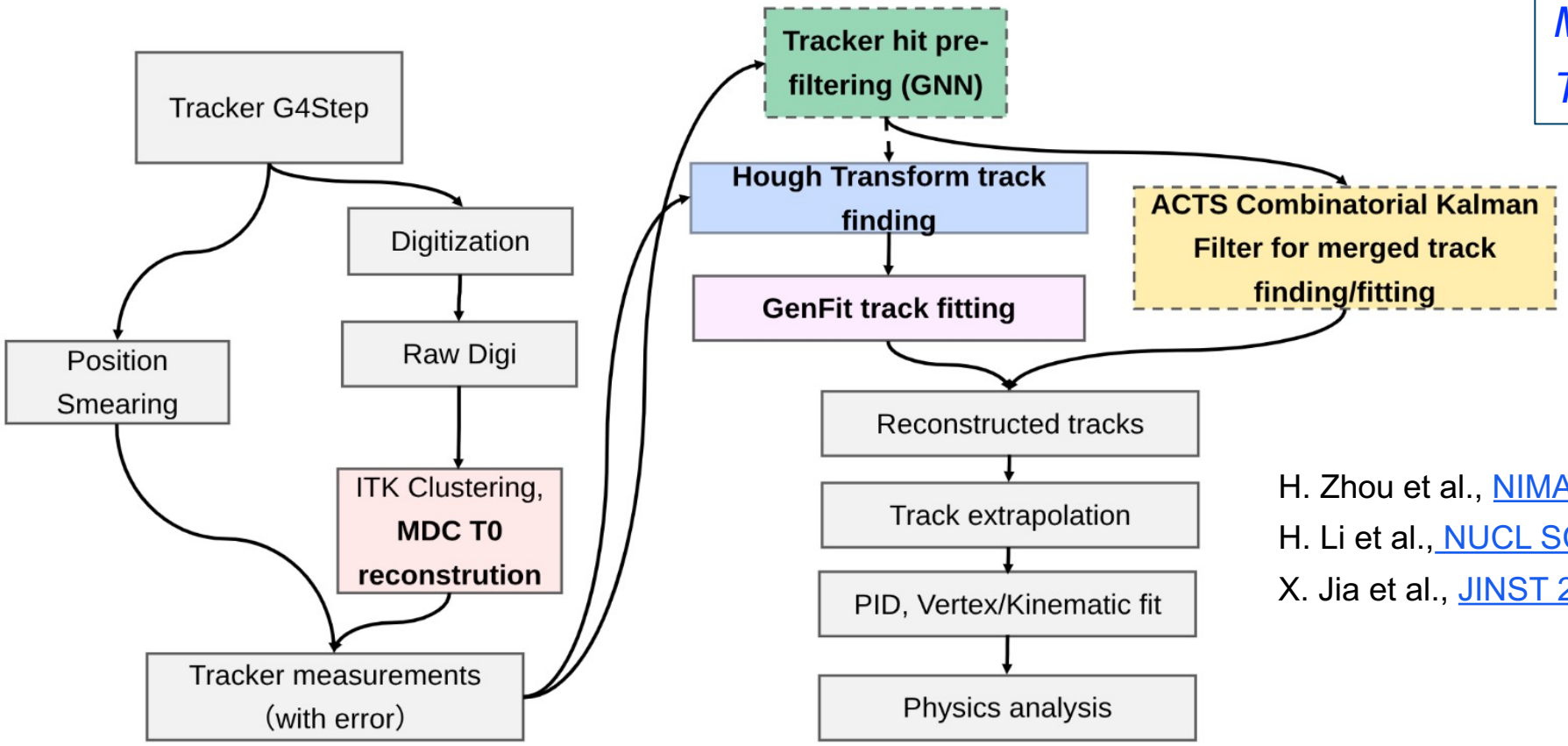
- Both **prompt** and **displaced** tracks with **high beam backgrounds**.
- Most tracks have p_T < **500 MeV**.





Tracking strategies

- **Hough transform** + **GenFit** shows decent performance for both prompt and long-lived particles
- **ACTS** is robust against beam backgrounds and superior in terms of CPU consumptions
- **GNN** is very useful in terms of noise removal



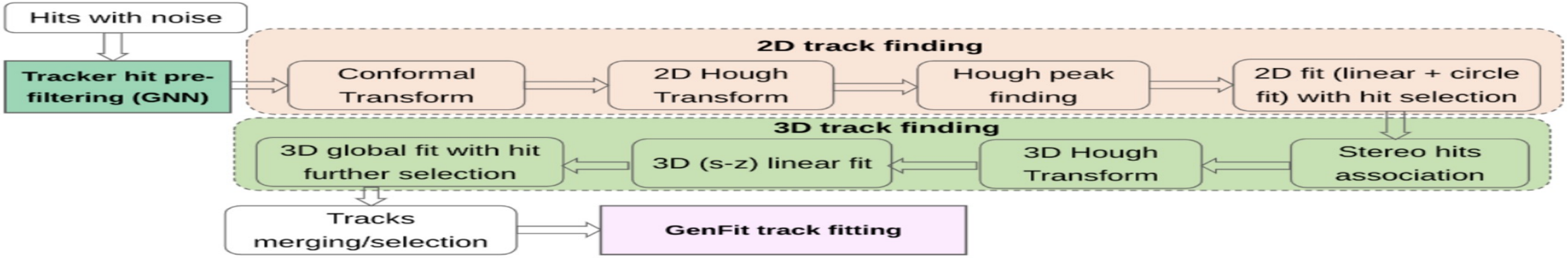
More details in Jin Zhang's talk: Track reconstruction in STCF

H. Zhou et al., NIMA 1075, 170357 (2025)

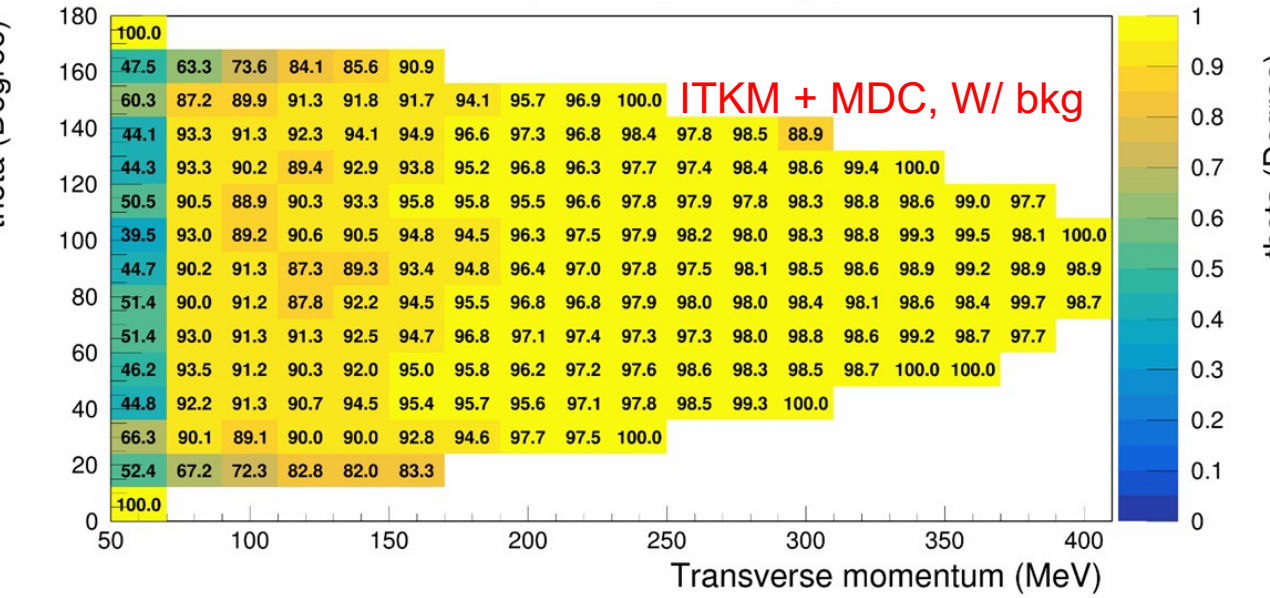
H. Li et al., NUCL SCI TECH 36, 171 (2025)

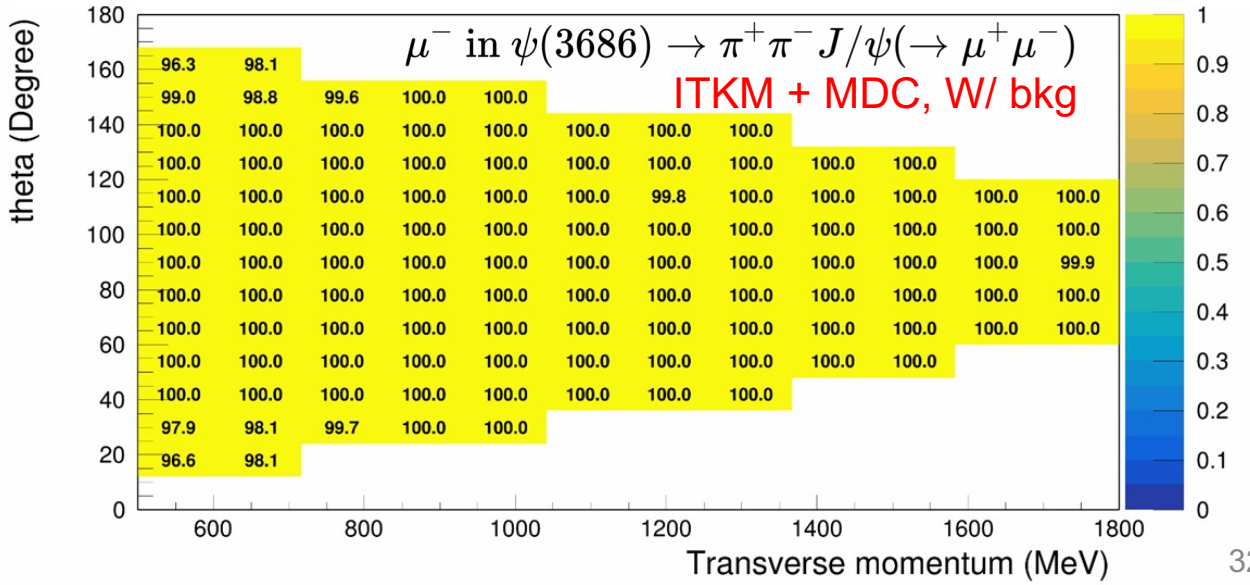
X. Jia et al., JINST 21 P03009 (2026)

Tracking efficiency for prompt tracks

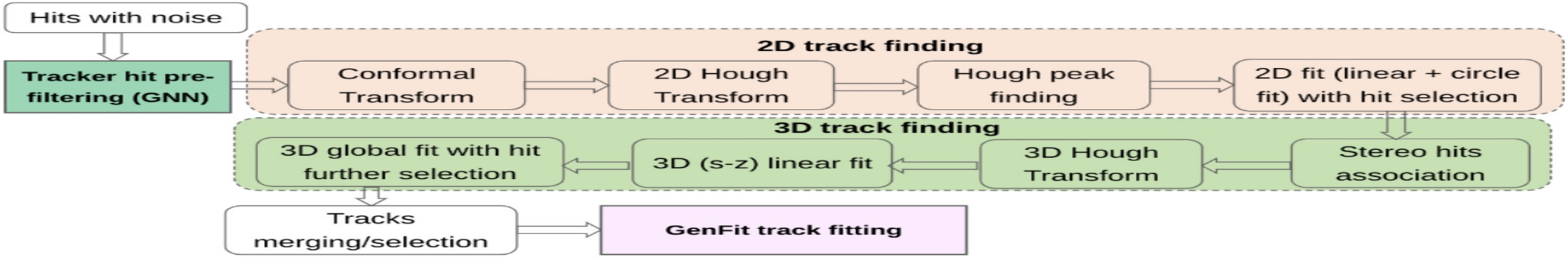


- Tracking performance for **prompt** tracks has been well consolidated
  - Except detector edge, **>90%** tracking efficiency for **p_T > 75 MeV**

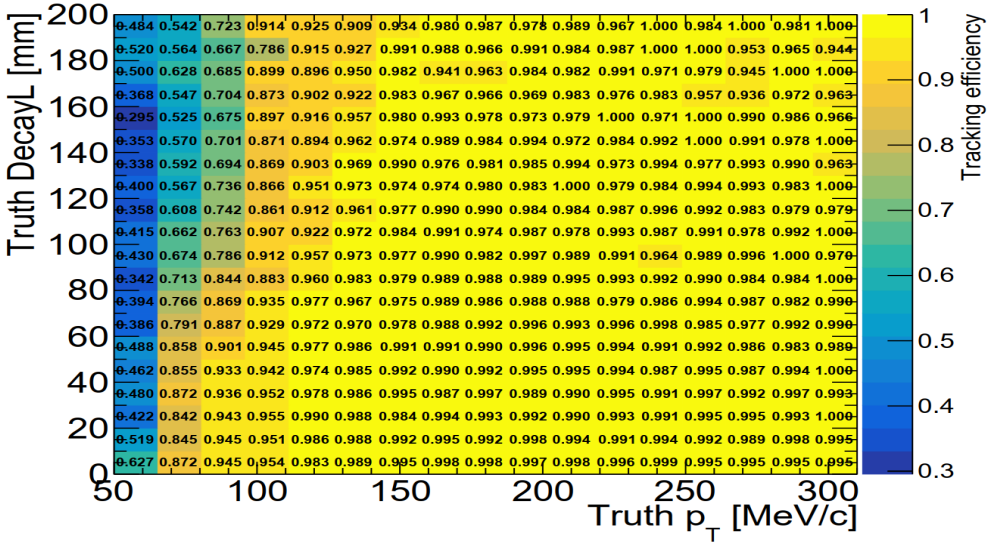


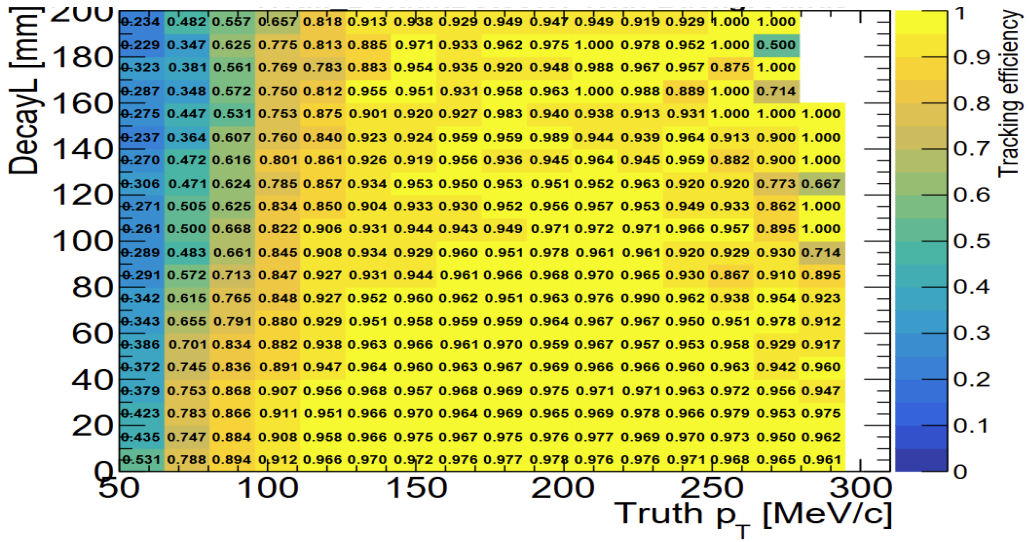


Tracking efficiency for displaced tracks



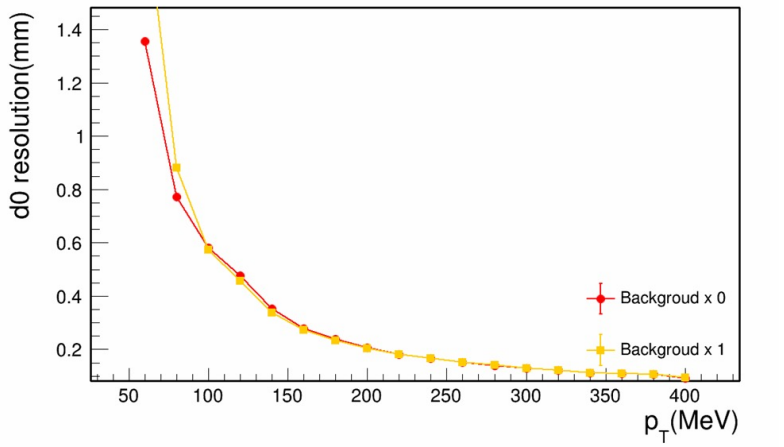
- Tracking performance for **displaced tracks** also achieved by by using additional Hough transform with displaced segment position as reference point
  - **>80%** track efficiency with decay length up to **200 mm**

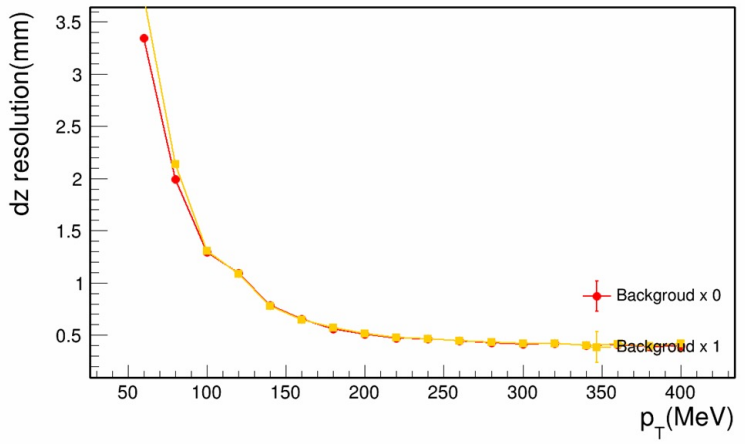


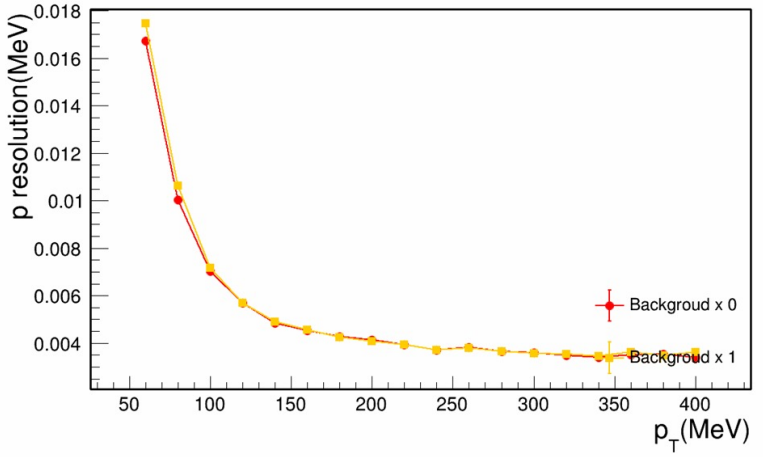


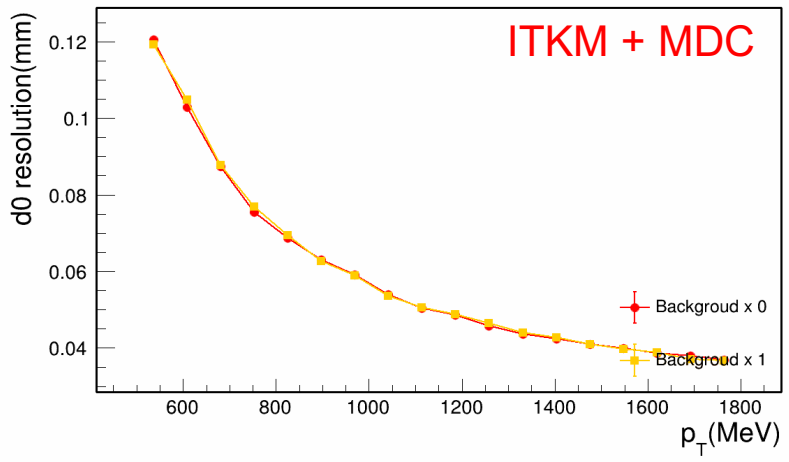
Tracking parameters resolution

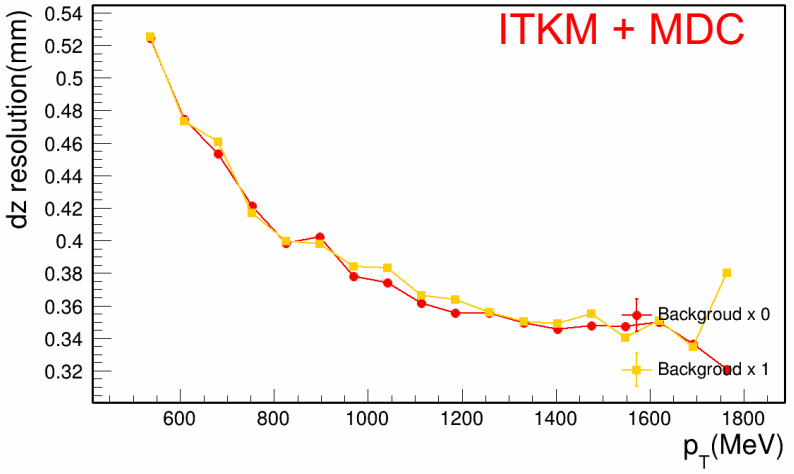
- p relative resolution is ~**0.5%** at p_T = **1 GeV**
- d0 (z0) has resolution down to **40 (350) um** with p_T > **1.5 GeV**

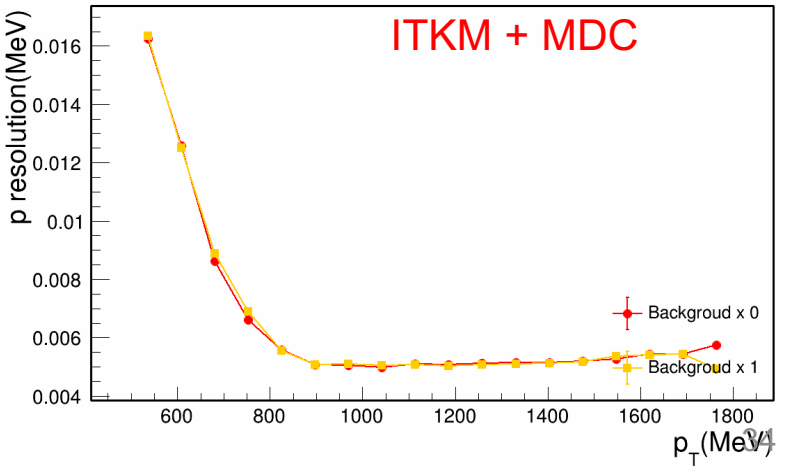








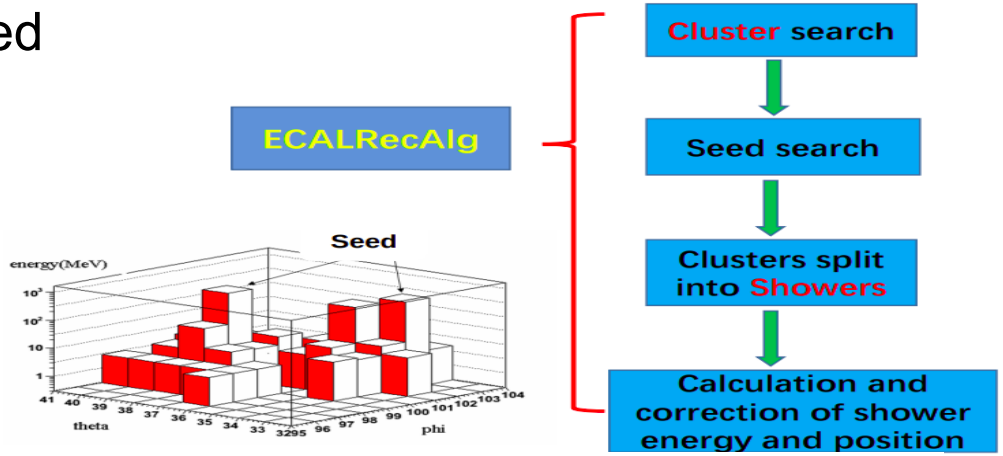


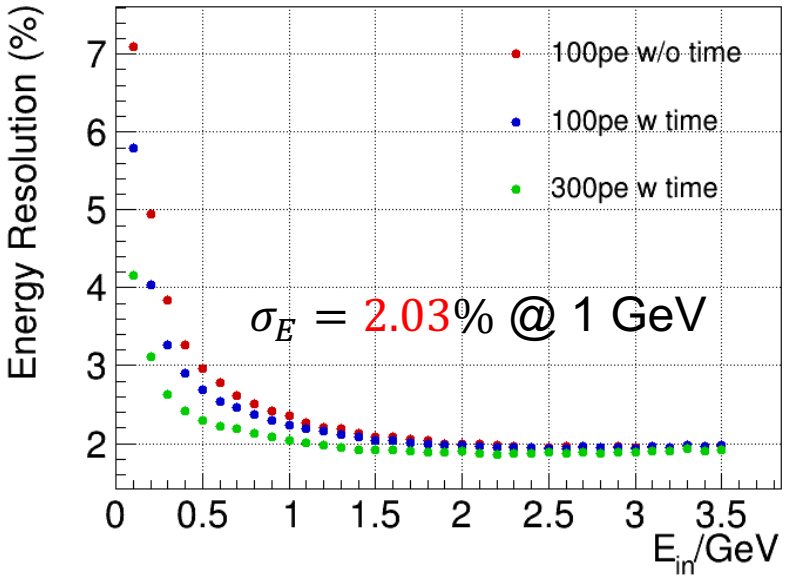


EMC performance

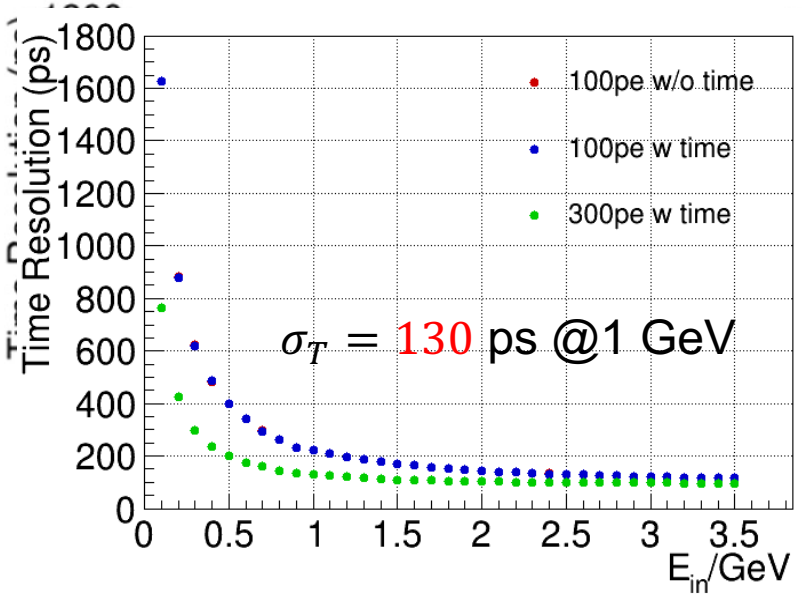
EMC reconstruction performance

- The complete reconstruction algorithms of ECAL are developed
  - Using realistic digitization parameters from beam test
- The **energy**, **time and position resolution** with backgrounds **meet physics requirements**

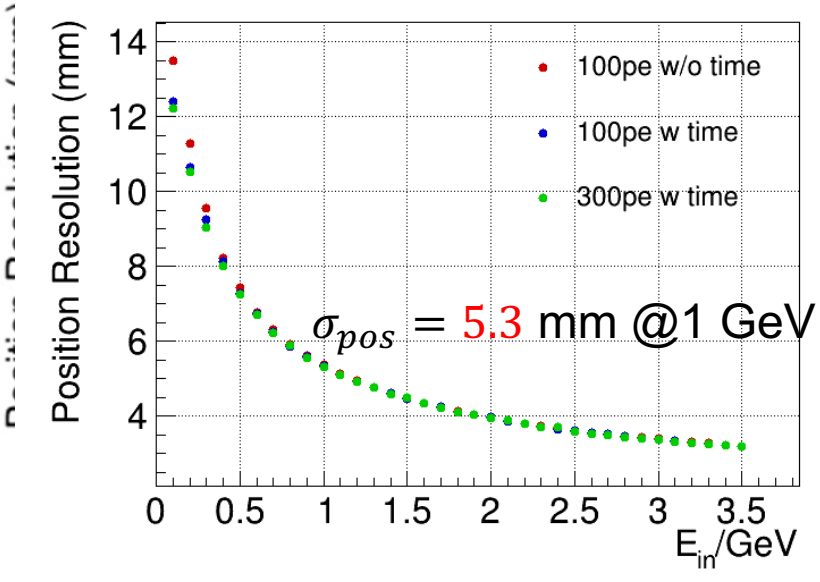




Energy



Time

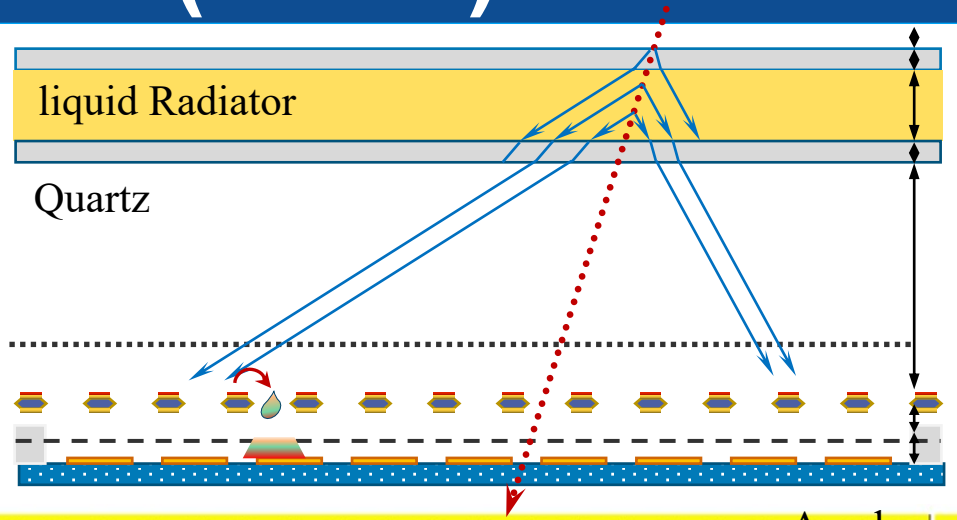


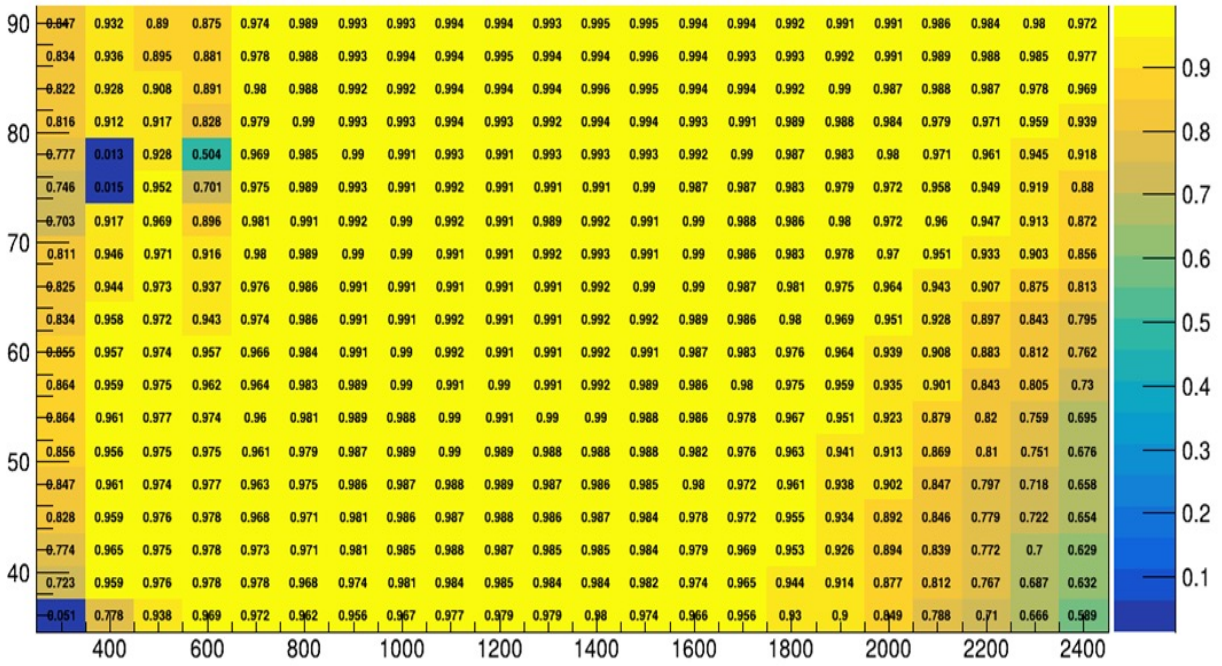
Position

PID performance

π/K PID performance (RICH)

- Two **RICH** reconstruction algorithms are developed
  - **Θc/ PDF reconstruction algorithms**
  - Both achieved **>97% π eff @ K mis-ID=2%**





π/K PID efficiency (θc algorithm costing low CPU)



π/K PID efficiency (PDF alog. better performance)

π/K PID performance (RICH)

- Two **RICH** reconstruction algorithms are developed
  - **PDF/ θc reconstruction algorithms**
  - Both achieved **>97% π eff @ K mis-ID=2%**

- 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:

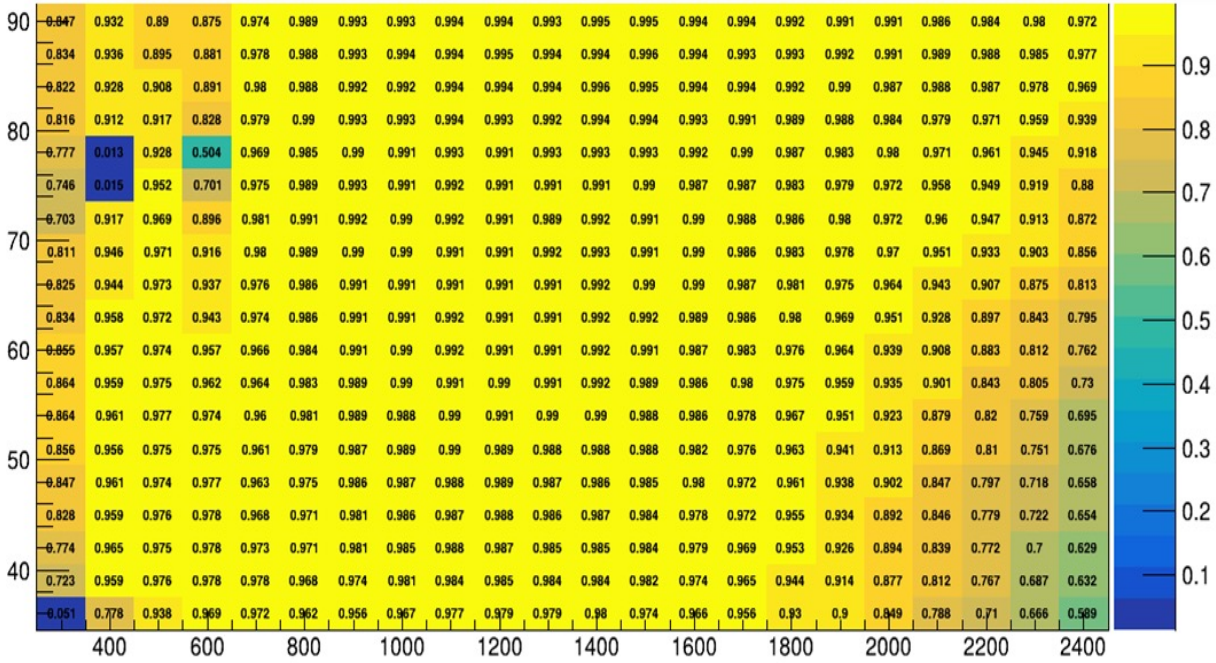
$$\ln \mathcal{L}_h = \sum_i^{npads} \ln pdf_{i,h}$$

- π, K separation:

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



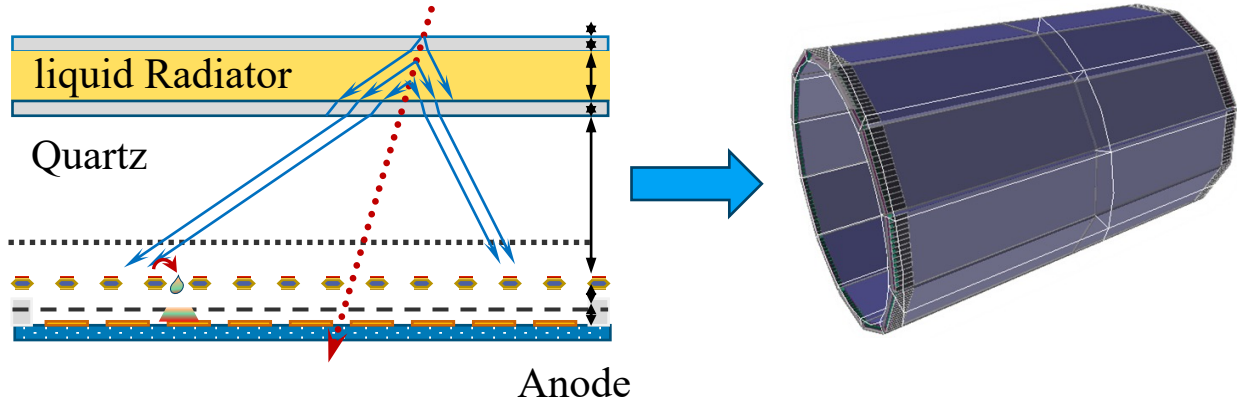
π/K PID efficiency (PDF alog. better performance)

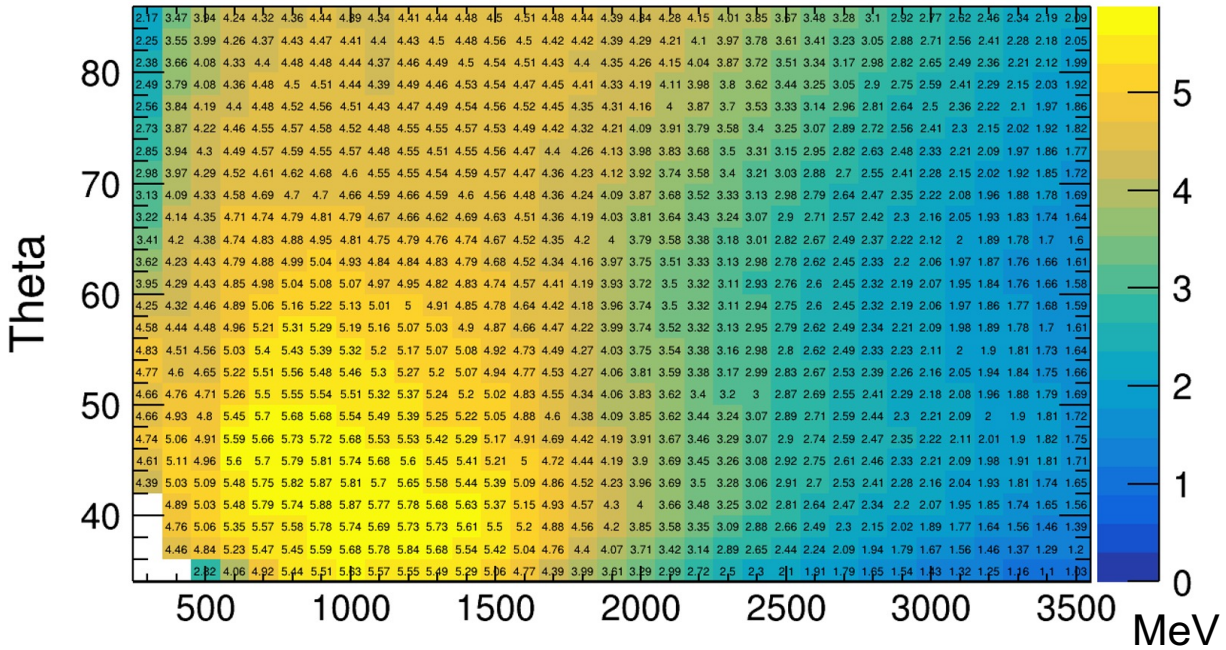


π/K PID efficiency (θc algorithm costing low CPU)

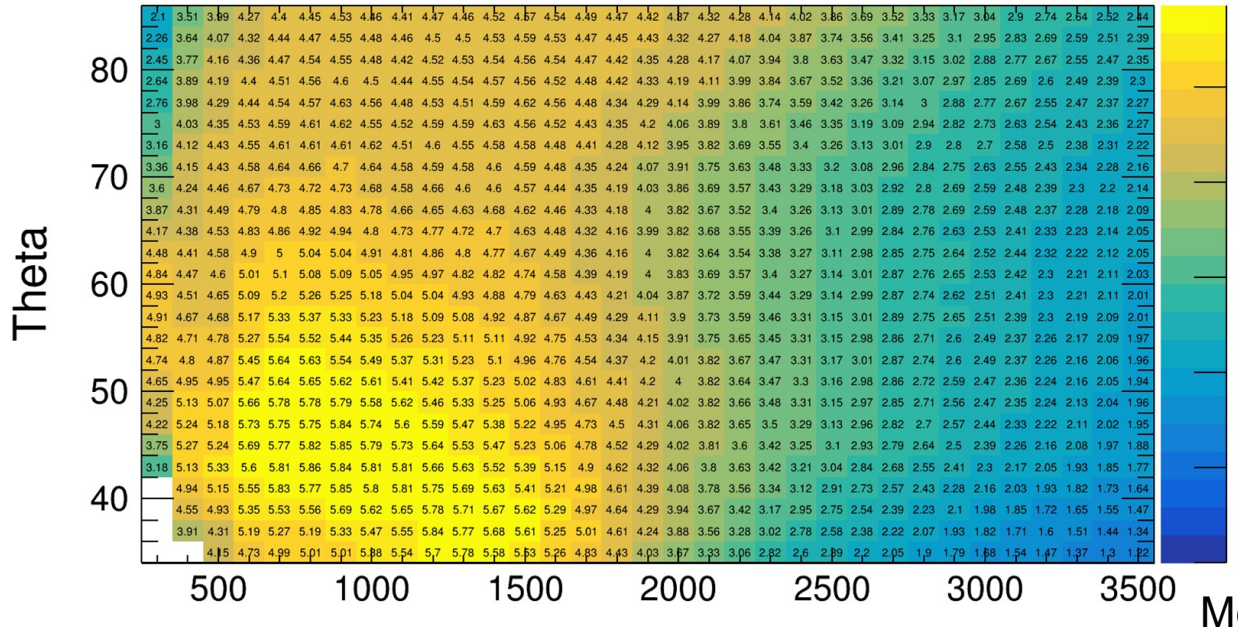
π/K PID performance (BTOF)

- One new option: **RICH -> BTOF** to have more photon electron and better performance @ θ=90°
- Developed T**iming** and **Imaging** methods
  - achieves **~ 4σ π/K** separation at 2 GeV/c
  - **consistent** performance from both methods
  - meeting the STCF requirement





Separation power of the Timing method

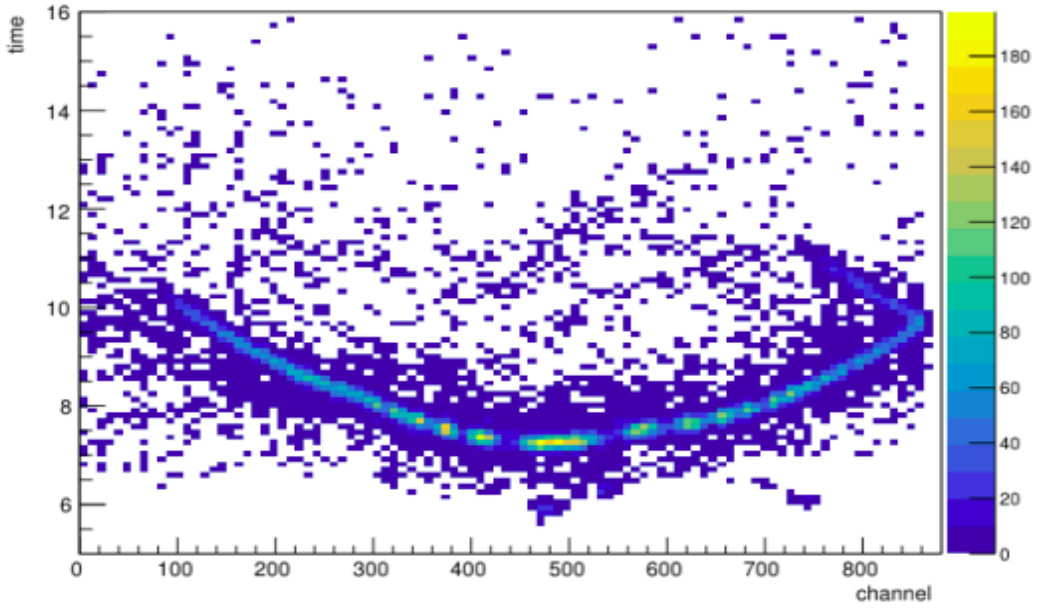


Separation power of Imaging method

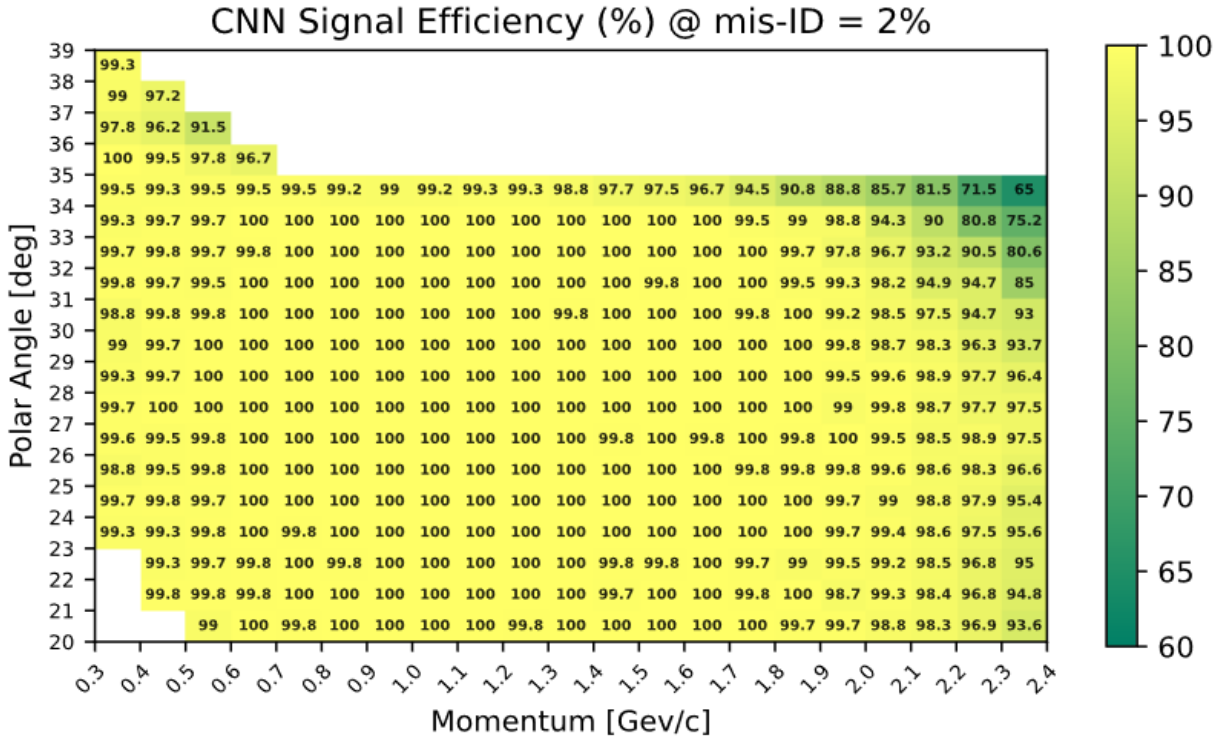
π/K PID performance (DTOF/RICH/BTOF)

- A **CNN (based on EfficientNetV2S) algorithm** is developed first **DTOF** and then extended to both **RICH** and **BTOF**
  - utilizing both timing and spatial information of the DTOF hits
  - achieved >97% π eff @ K mis-ID=2% for P< 2 GeV/c

More details in wanlin's talk: CNN for PID systems in STCF

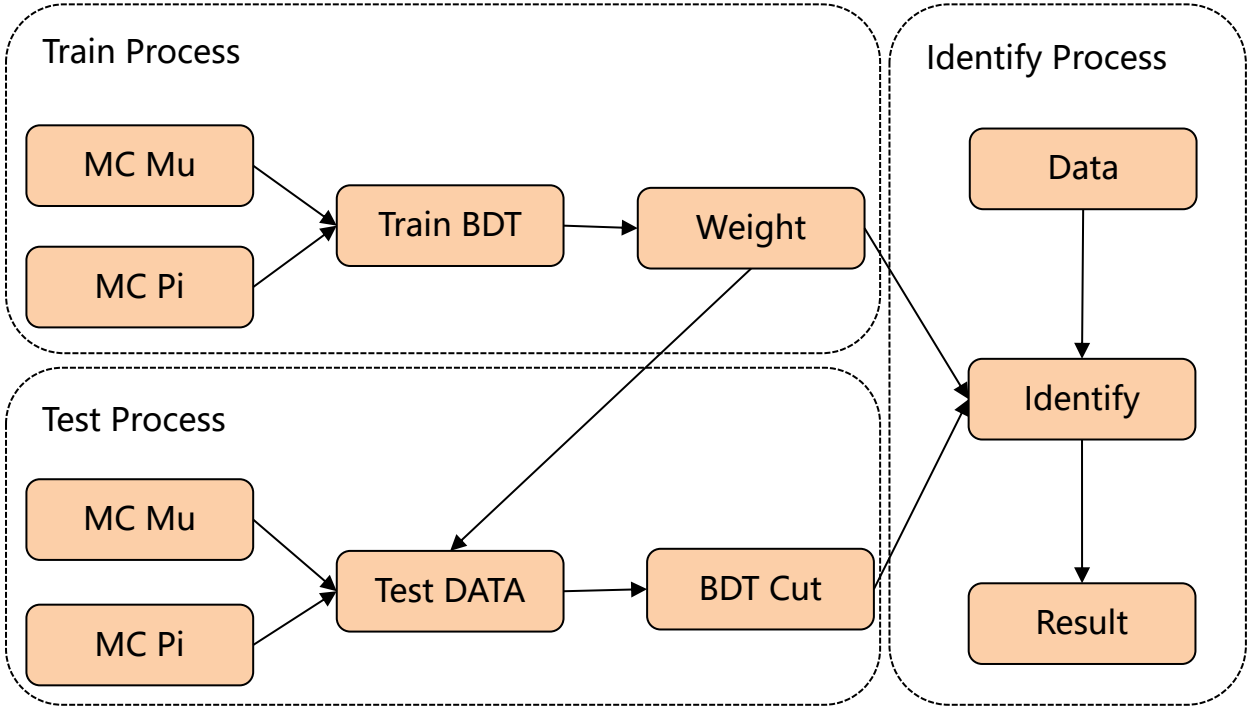


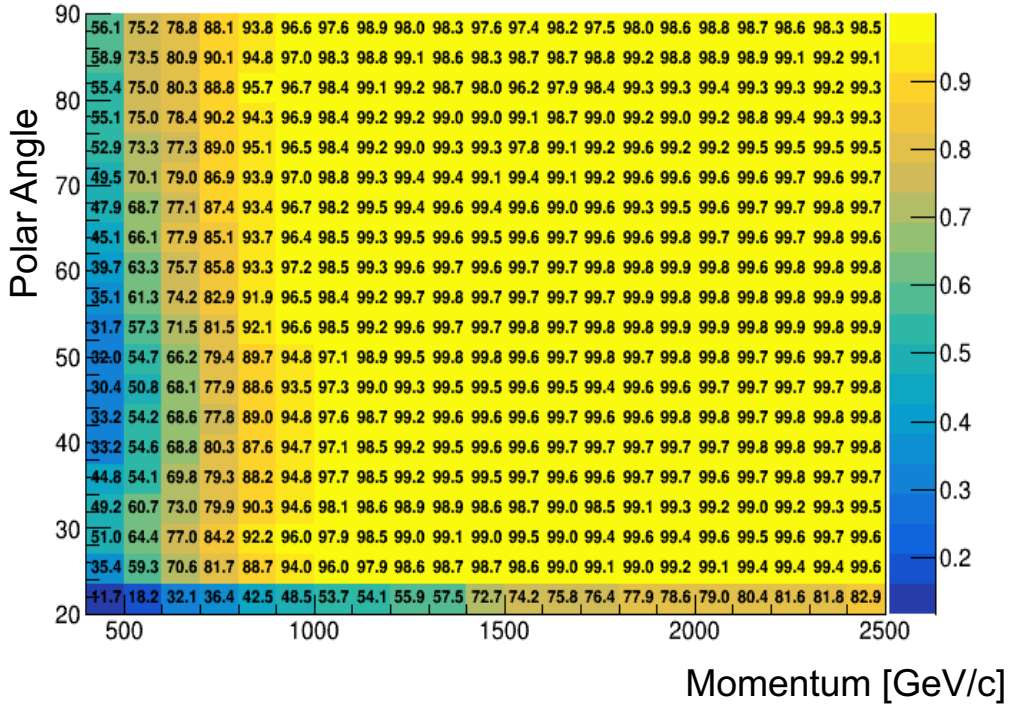
X-label: the hit position of Cherenkov photon by PMT
Y-label: the arrival time of Cherenkov photon by PMT
Value: the number of photons within this bin



μ/π PID performance (MUD)

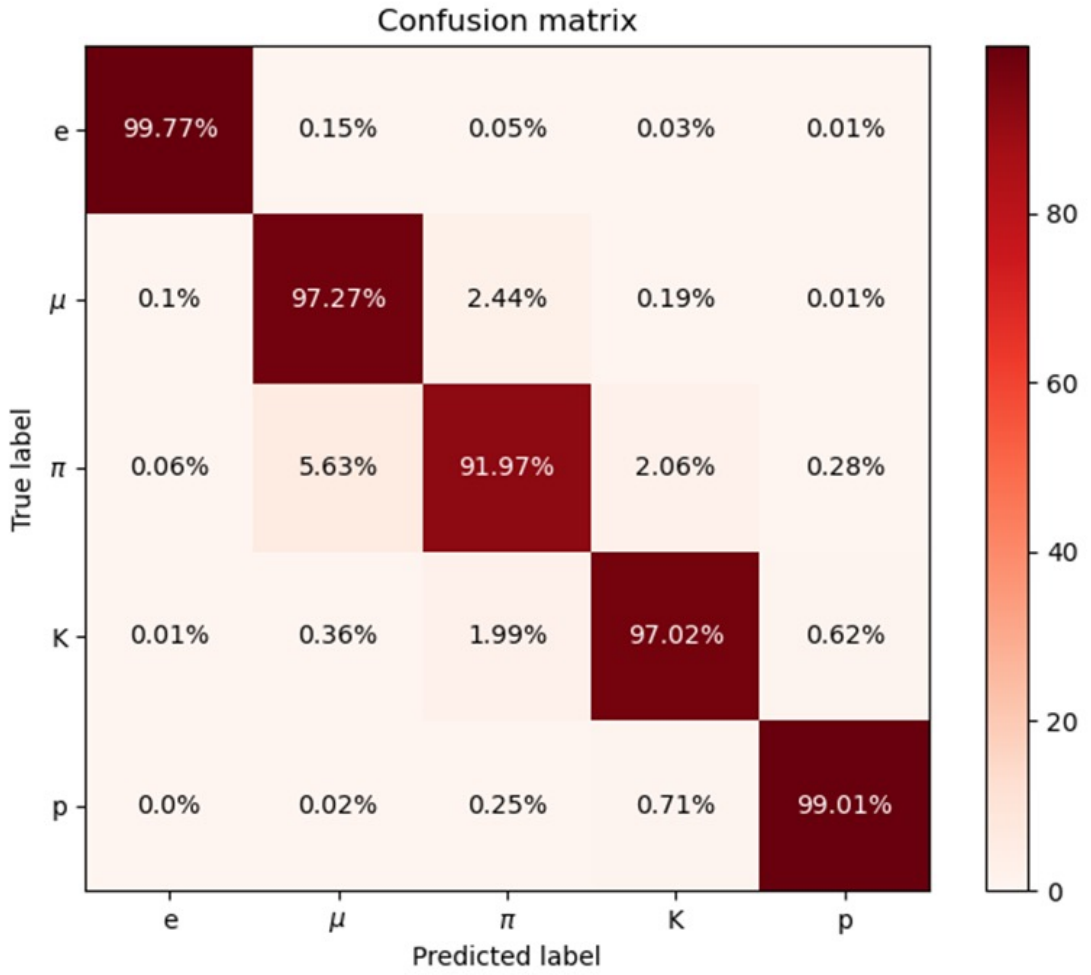
- Regarding μ/π identification, a **BDT-based algorithm** is developed on top of realistic simulation and digitization
  - μ efficiency > **95%** @suppression = 30 with P > **1 GeV**





BDT-based global PID performance

- **A BDT-based global PID algorithm is developed** to take the measurements from each sub-detectors
  - **~ 100%** efficiency for e and proton
  - **>95%** efficiency for K
  - **> 90%** efficiency for μ (>95% with p >1.5 GeV)
  - **>90%** efficiency for π in most regions

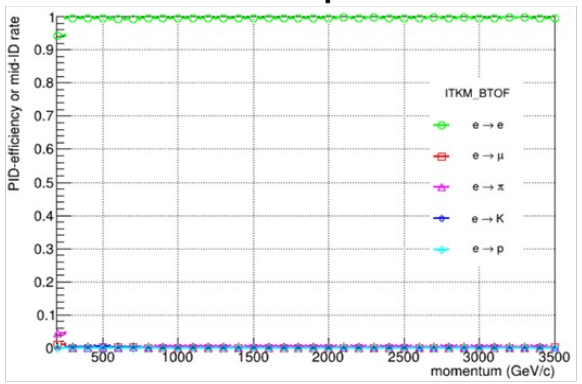


Traditional global PID performance

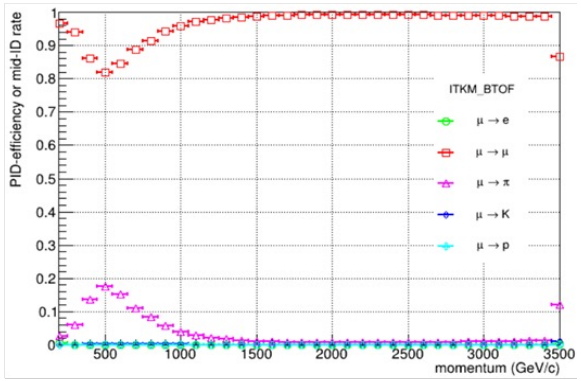
- A **combined likelihood with weight** was applied to integrate all PID information.
  - different weights for the same likelihood values across particle comparisons.
  - achieves better performance

- $\mathcal{L}_h = \prod_{det} \mathcal{L}_{det,h}^{w_{det}}$
- $w \propto \frac{\overline{\mu}_{\Delta logL}}{\overline{\sigma}^2_{\Delta logL}} \propto \frac{N_\sigma}{\overline{\sigma}_{\Delta logL}}$

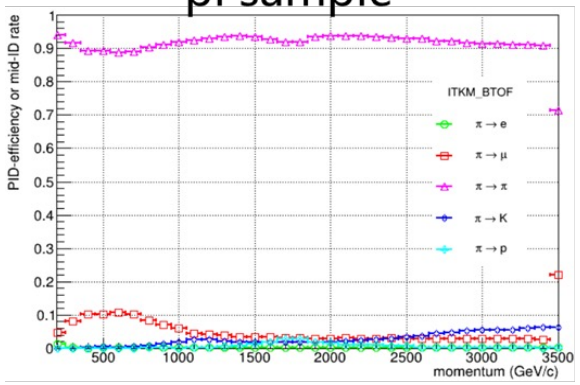
e sample



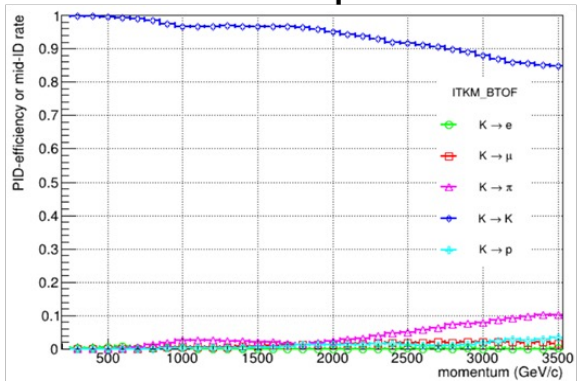
mu sample



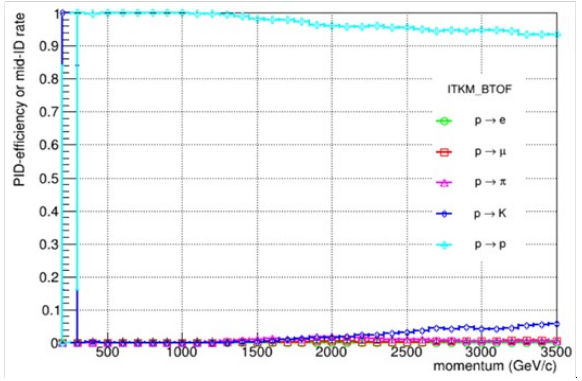
pi sample



K sample



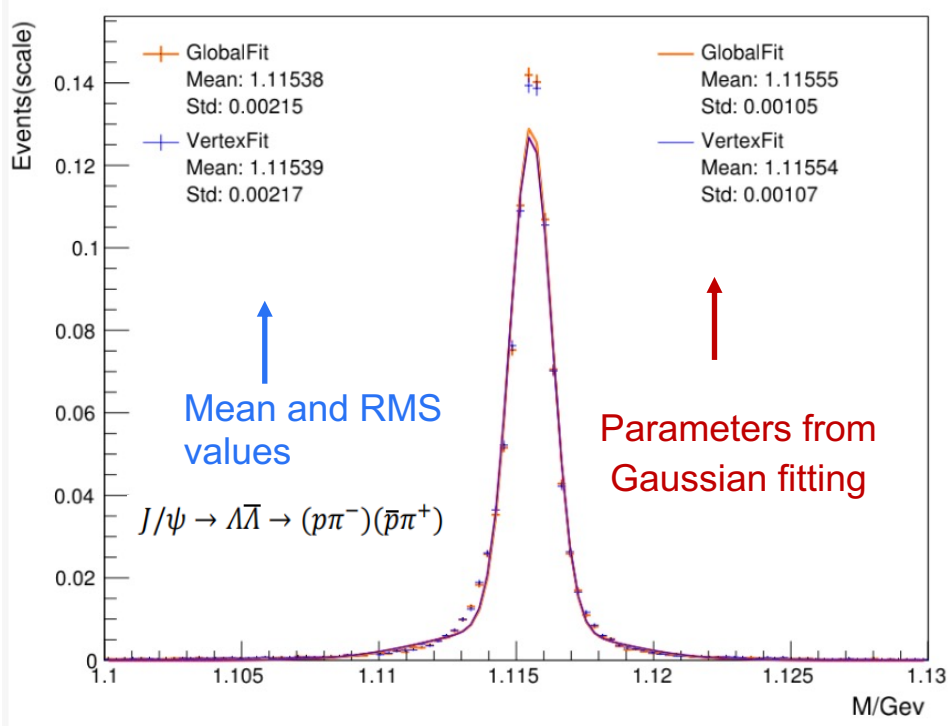
p sample



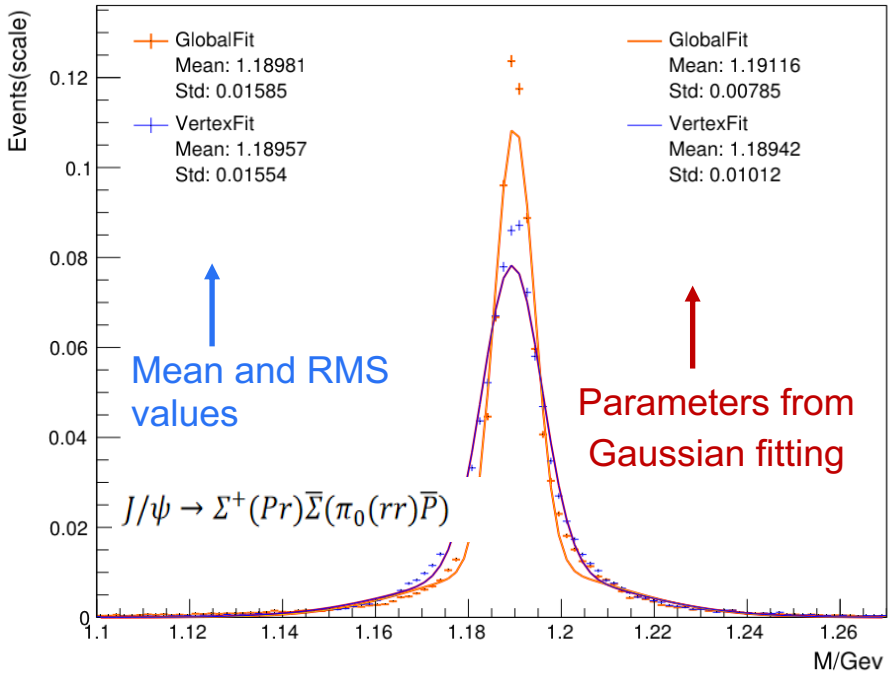
Analysis tool

Kinematic and Vertex fitting tools

- Developed the **kinematic and vertex fit** package for STCF
  - The **vertex fitting algorithm** (VertexFit) imported from **BESIII**
  - The **global vertex fitting** algorithm imported from **BelleII,** which **simultaneously performs kinetic and vertex fit** and gives **better performance for long-lived particles**



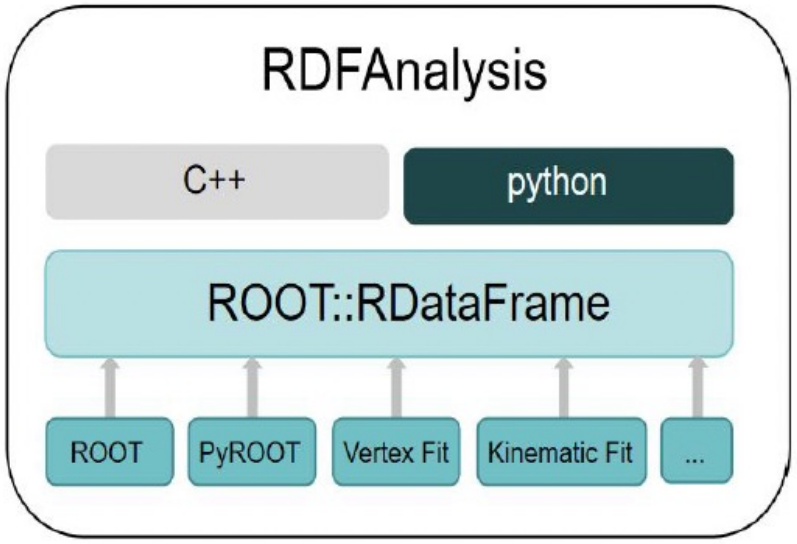
Invariant mass of Λ

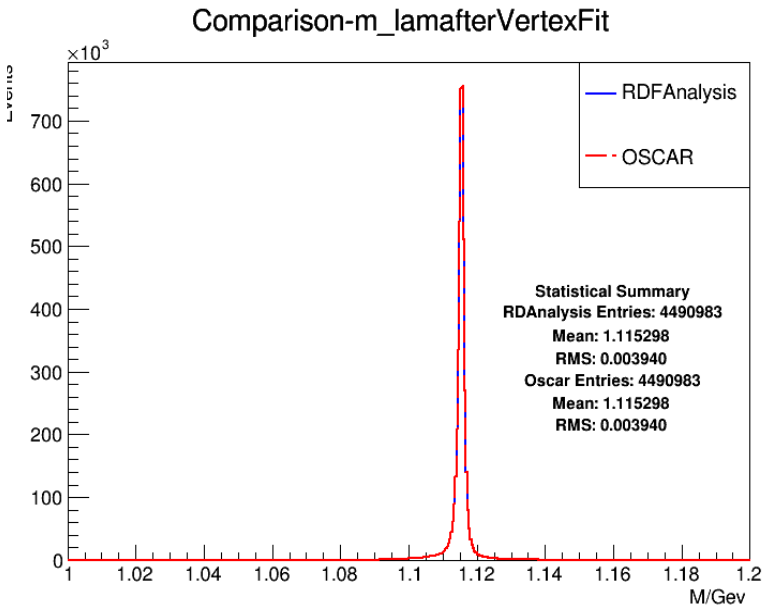


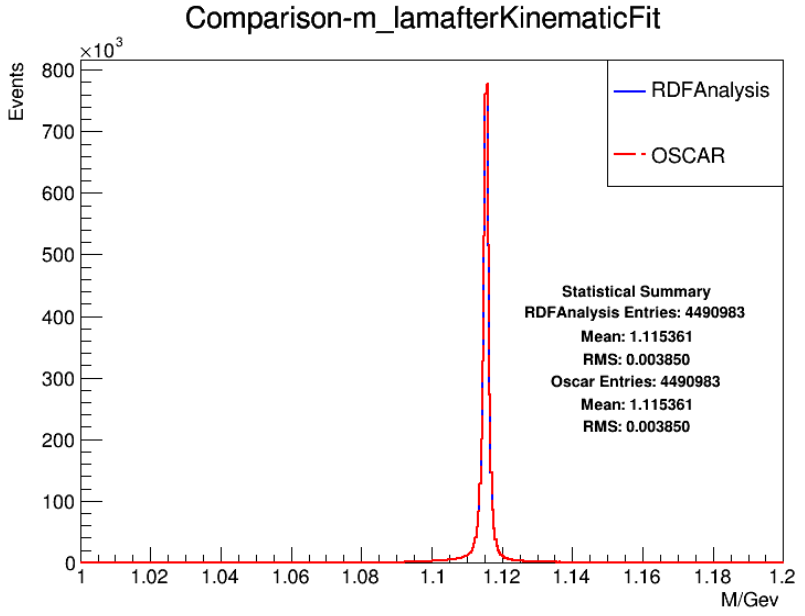
Invariant mass of Σ

RDataFrame analysis framework

- A **Parallelized Analysis Framework** is developed to take advantage of multi-threading during physics analysis
  - Based on **ROOT RDataFrame** supporting declarative programing and concurrent processing
  - **Analysis tools** (GlobalPID, VertexFit) imported into the framework
- The physics analysis results of process $J/\psi \to \Lambda\bar{\Lambda}$ using 10 million events are **consistent** with the traditional methods.







Physics Study under OSCAR

Fundamental Symmetries

Topic	Simulated Process
Baryon CPV and EDM	$J/\psi \to \Lambda\bar{\Lambda}$
	$J/\psi \to \Xi^+\bar{\Xi}^-$
τ physics (QE, EDM, rare decay)	$\tau \to \pi(\pi\pi^0)\nu$
	$\tau \to 3\mu$
	Radiative decay
	τ EDM
Charm mixing, strong phase	C-even correlated $D\bar{D}$
	$D^*_{(s)} \to \gamma/\pi^0 + D^0_{(s)}$
CPV in Λ_c^+ decays	$\Lambda K^+, \Sigma^0 K^+, \Sigma^+ K^0_S$
CP and CPT in $K^0_S \to \pi^+\pi^-$ decay	$J/\psi \to K^0_S K\pi$
CLFV decay	$J/\psi \to \gamma ll'$
	$D^0 \to \bar{\Lambda} + ALP$

Quark Confinement

Topic	Simulated Process
Charmonium-like, tetra- and penta-quarks	$X(3872)$
	$e^+e^- \to \pi Z_c(3900)$
	$e^+e^- \to P_c\bar{p} \to J/\psi p\bar{p}$
	$e^+e^- \to J/\psi\eta_c$
Collins FFs	$e^+e^- \to KK, \pi\pi, K\pi$
Hadron EMFFs	$e^+e^- \to p\bar{p}$
From factors in hyperon decay	$J/\psi \to e^+e^-\pi^0/\eta$
	$\Sigma^+ \to \gamma p$
	$\Lambda \to pe\nu$
	$\Xi^- \to \Lambda e\nu$

Physical Fund. Parameters

Topic	Simulated Process
R scan	$e^+e^- \to hadrons$
Two-photon physics	$\gamma\gamma \to hadrons$
	$\eta_c \to \gamma\gamma$
Tau mass	$e^+e^- \to \tau^+\tau^-$ near threshold
CKM matrix	$D \to \pi\mu\nu$
	$D \to \mu\nu$
	$D \to inclusive$
	$\tau \to K\nu$
Strong phase	$D^0 \to K^-\pi^+\pi^0$
α_s	$e^+e^- \to \tau^+\tau^-, \tau \to l\nu\nu, \tau \to \pi\pi^0\nu_\tau$

Summary

- The STCF offline software system (**OSCAR**) is being developed based on **SNiPER** and 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 detector simulation+ background mixing + digitization+ reconstruction +analysis is built
  - Software performance mostly **meets the CDR requirements** and it is **used for physics analysis**
  - Oscar **Tutorial & Hackathon** took place from Feb. 22-23, 2025 (https://indico.pnp.ustc.edu.cn/event/3624/)

- **Future plan**
  - Optimization of recon. algorithms, especially with digitization information after event mixing
  - Improving computing performance to meet high luminosity requirements
  - Collaboration with JINR on the development of SPD offline software

Detectors	Simulation	Digitization	Reconstruction	Analysis tools		
				Global PID	Traditional combined PID	Kinematic& Vertex Fit
ITK	✓	✓	✓			
MDC	✓	✓	✓			
RICH	✓	✓	✓	Charged tracks: ✓	✓	✓
DTOF	✓	✓	✓			
BTOF	✓	✓	✓	Neutral tracks: ✓		
EMC	✓	✓	✓			
MUD	✓	✓	✓			

✓ Good state

✓ Under optimization

✓ Under development

Thanks for your attention !