STCF全模拟框架及更新计划

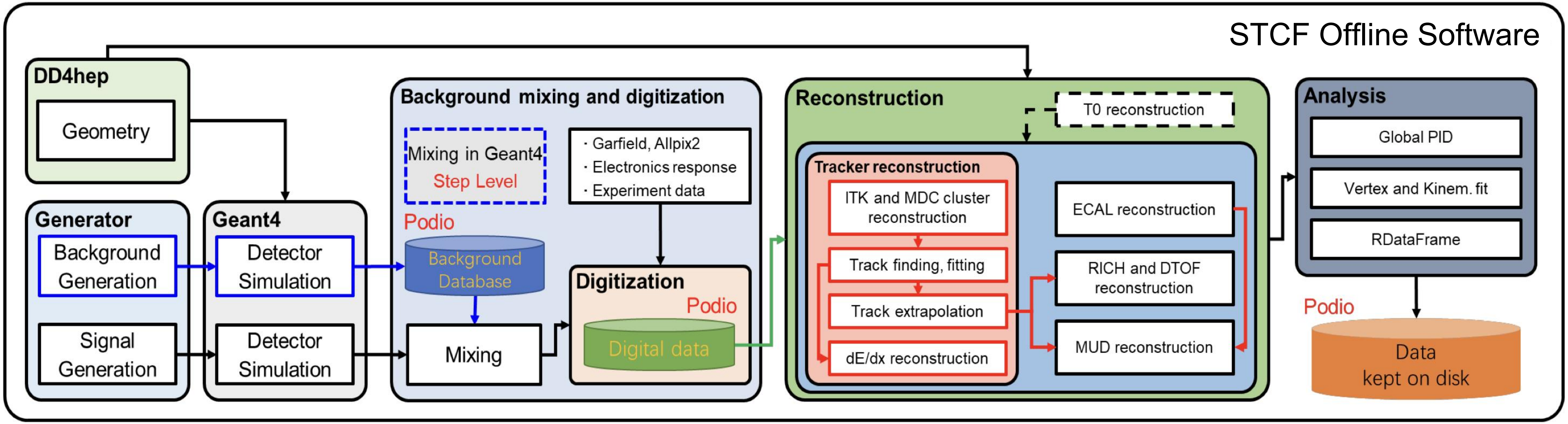
李腾　(tengli@sdu.edu.cn)

代表离线软件组

2025年9月20日

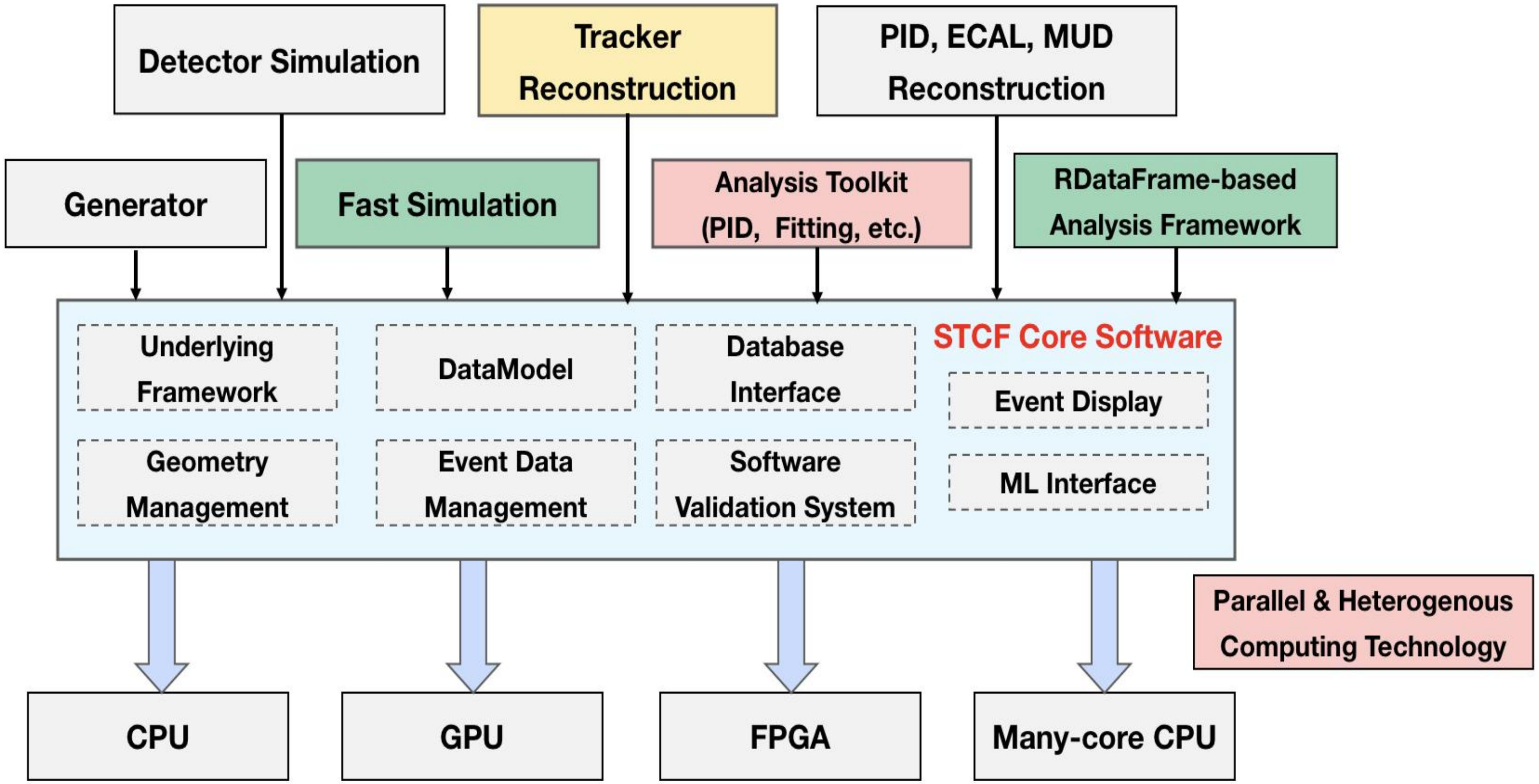
STCF Offline Software System （OSCAR）

- OSCAR is develeped to provide common functionalities for **detector simulation**, **reconstruction**, **calibration** and **physics analysis** at STCF, implement raw data processing chain and MC data production tasks



Overview of OSCAR System

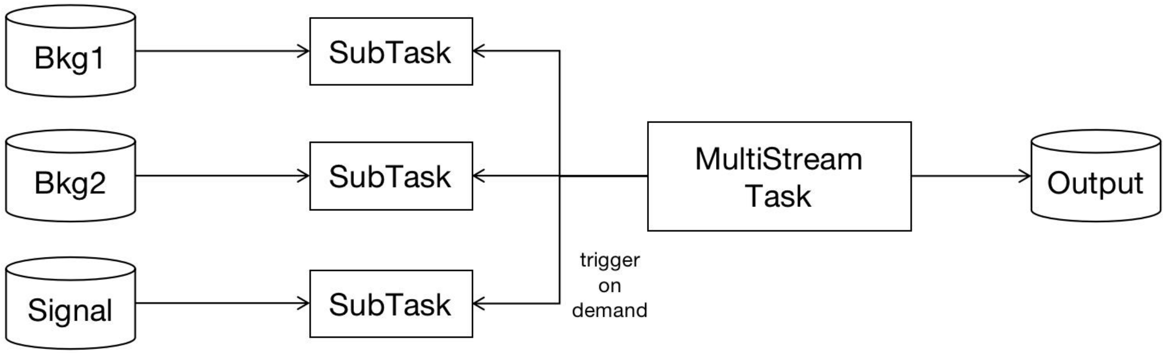
- Core software (common functionalities for data processing)
  - Underlying framework
  - Event data management
  - Geometry/conditions data management
  - Event display
  - Software distribution/validatioı
- STCF applied software
  - Full/fast simulation
  - Tracking software
  - ECAL software
  - Particle identification
  - Analysis toolkits

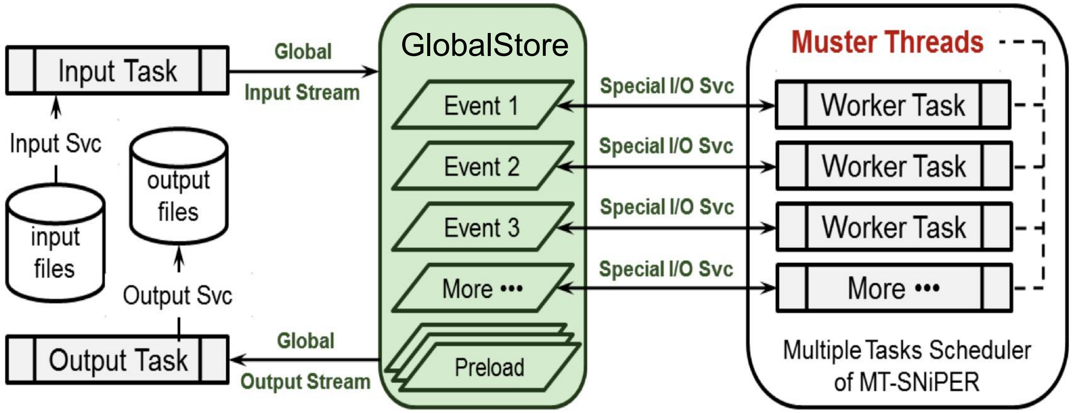


Framework

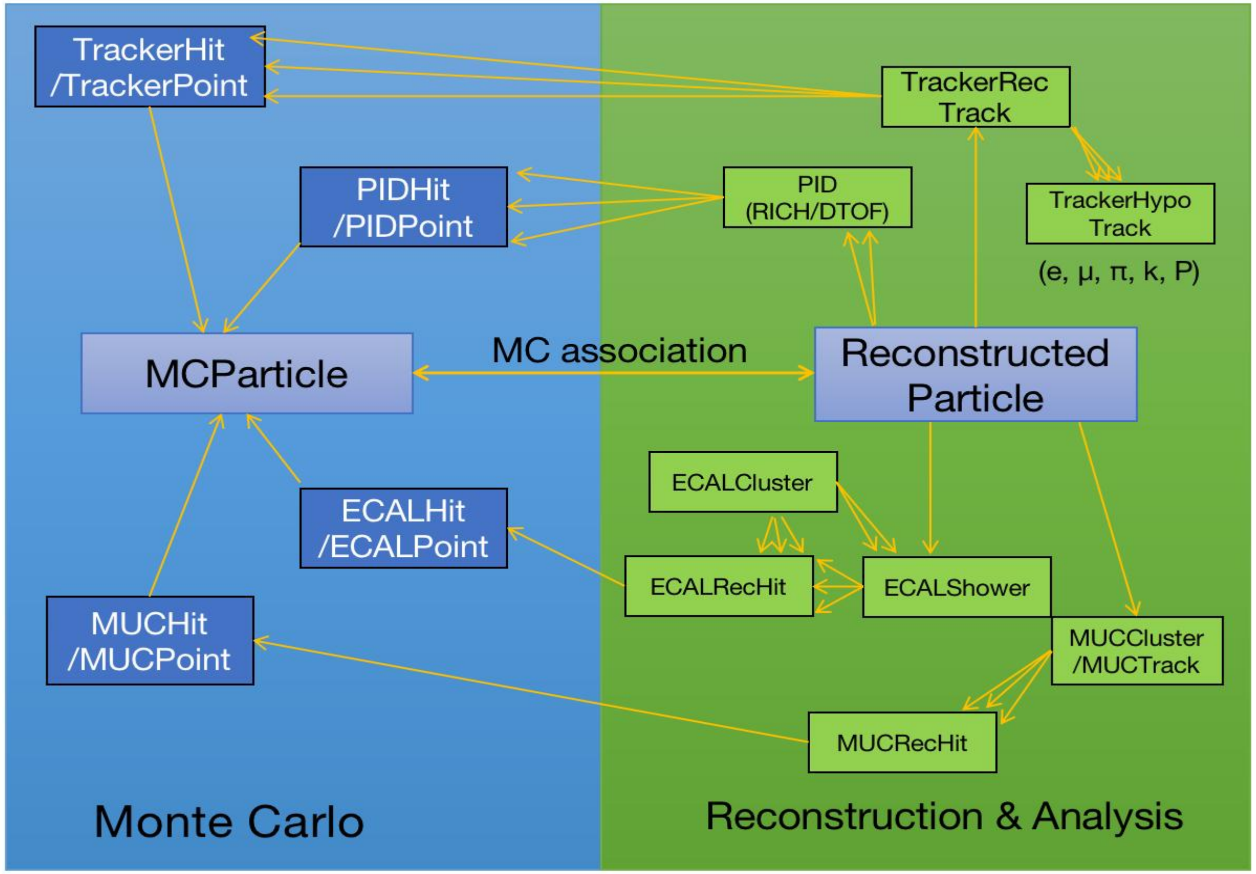
Underlying Framework: SNiPER

- The underlying framework builds the skeleton of OSCAR
  - Provide basic functionalities of **event loop control, algorithm scheduling, thread management, user interface, job configuration, logging** etc.
- OSCAR adopts SNiPER as the underlying framework
  - Developed since 2012, maintained by **10+ developers from IHEP, SDU, etc.**
  - Adopted by JUNO (neutrino), LHAASO (cosmic ray), nEXO (neutrinoless double beta decay) and HERD (dark matter)
- Advantages of SNiPER
  - Lightweighted, efficient and highly extendable. Support of flexible data processing chain.
  - Efficient multithreading. C++/Python hybrid programing.





Event Data Model



All EDM for simulation, digitization and reconstruction defined and preliminary optimization

Further optimization is needed

YAML

```
MCParticle:
  Description : "Data class for storing Monte Carlo particles"
  Author : "SDU"
  Members :
    - int trackID                 // track ID
    - int PDG                     //PDG code of the particle
    - int generatorStatus         //status of the particle as defined by the generator
    - int simulatorStatus         //status of the particle from the simulation program - use BIT constants
    - int type                    //particle type. only generatorStatus== 1 or 2 has the type
    - float charge                //particle charge
    - float time                  //creation time of the particle in [ns] wrt. the event, e.g. for preassig
    - double mass                 //mass of the particle in [MeV]
    - Vector3d vertex             //production vertex of the particle in [mm].
    - Vector3d endpoint           //endpoint of the particle in [mm]
    - Vector3f momentum           //particle 3-momentum at the production vertex in [GeV]
    - Vector3f momentumAtEndpoint //particle 3-momentum at the endpoint in [GeV]
    - Vector3f spin               //spin (helicity) vector of the particle.
```

code generator

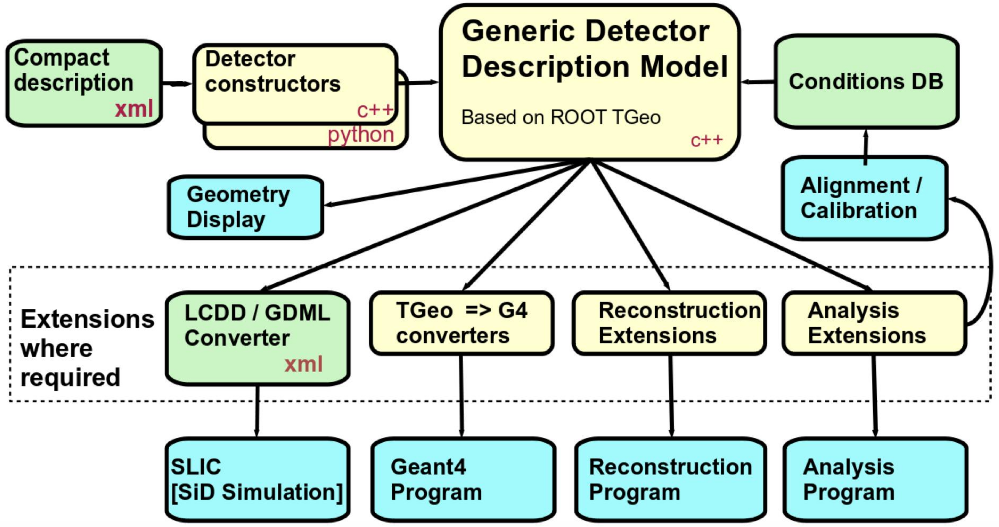
C++

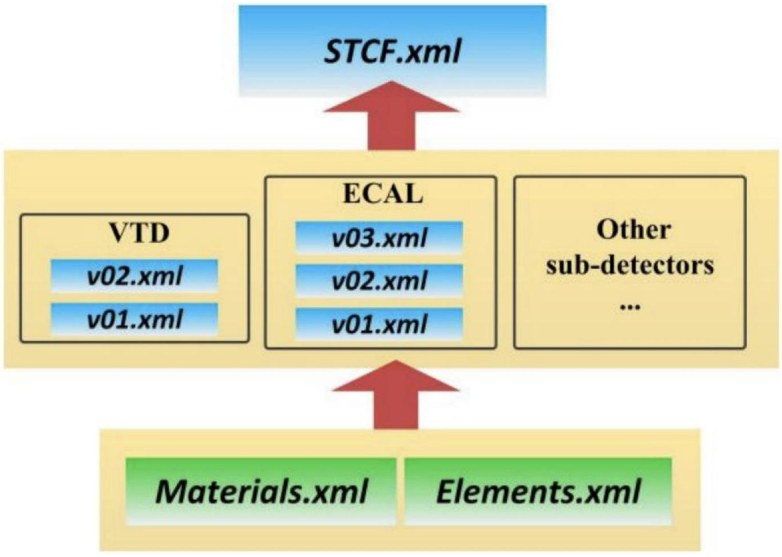
MCParticle.h	MCParticle.cpp
MutableMCParticle.h	MutableMCParticle.cpp
MCParticleObj.h	MCParticleObj.cpp
MCParticleData.h	MCParticleData.cpp
MCParticleCollection.h	MCParticleCollection.cpp
MCParticleCollectionData.h	MCParticleCollectionData.cpp

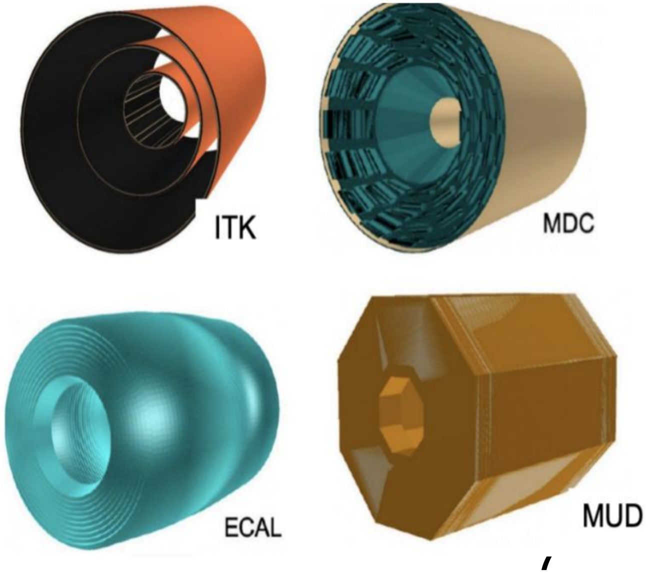
Based on YAML definition, generate EDM C++ code accordingly

Geometry Management System

- Geometry Management System (GMS) in OSCAR is based on DD4hep
- Single source of detector information for detector description, simulation reconstruction and event display
  - Complete geometry (and geometry used for beam tests) defined with XML files and C++ parser
  - Further detector optimization will be based on this



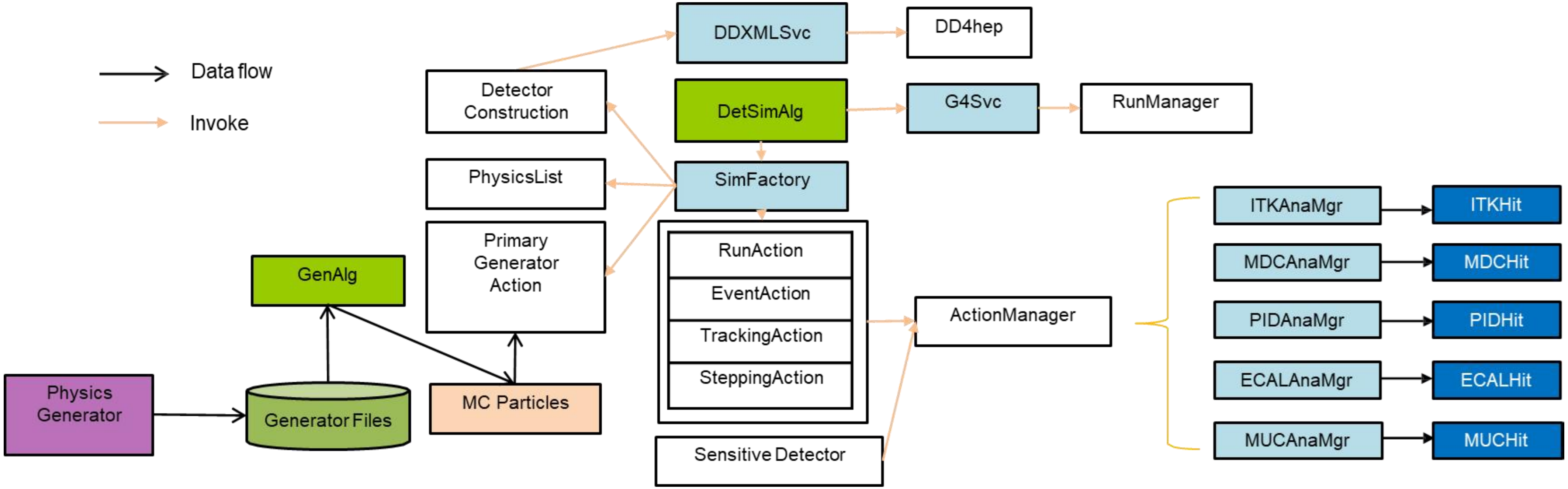




MC Simulation

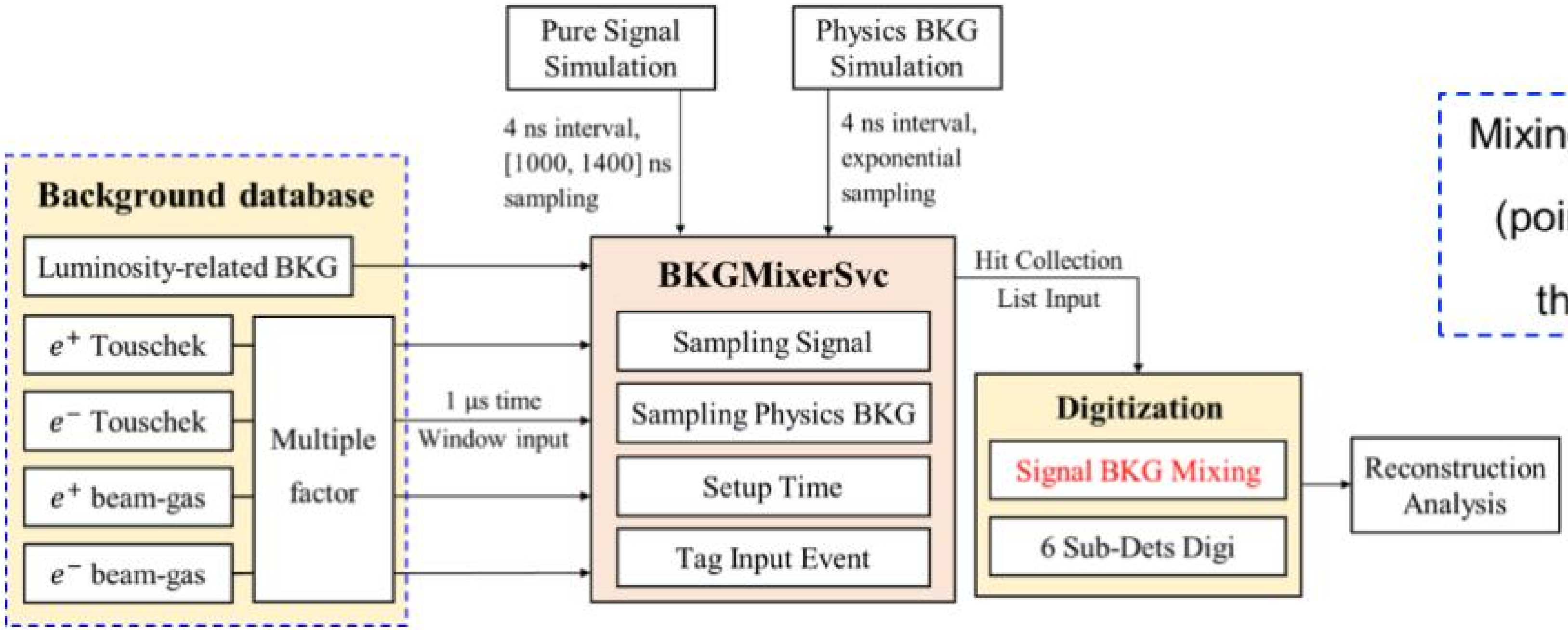
Full Simulation Based on Geant4

- **Full chain** of detector simulation from generator to simulated information is built
- Providing **flexible configuration** for different purposes of detector simulation
  - **Generator:** EvtGen, Babayaga, KKMC, Phokhara, DIAG36, BBBrem, **HepEvtReader**
  - **Geometry** for different detector design options from DD4hep
  - **User actions** for recording MC truth information, G4Step level



Digitization and Mixing

- Developed a **unified digitization framework** for all sub-detectors within OSCAR
  - All sub-detector moved to **sampling method,** including **MDC dE/dx** and **hit position**
- Event mixing with background at the **Geant4 step level**
  - Simulated background particles as input, a unified algorithm applied to each sub-detector
  - Considering **electronic response, noise and other effects,** as input for reconstruction and analysis



Mixing in Geant4 Step Level

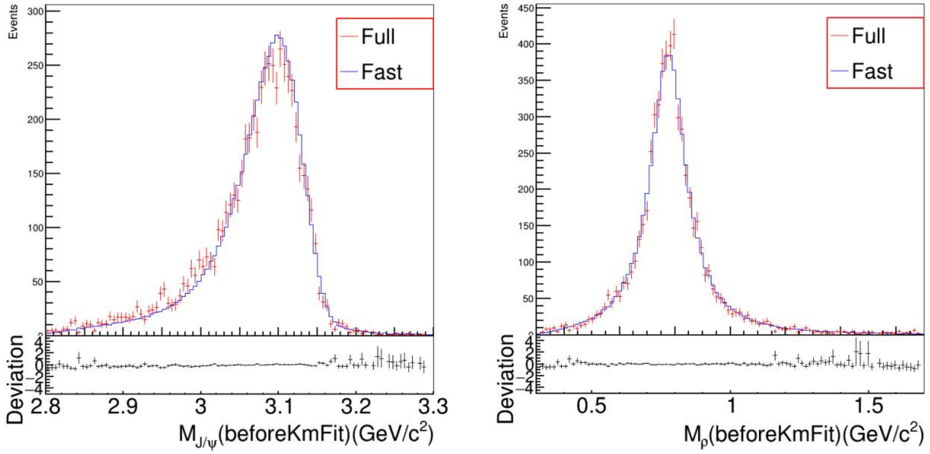
(points or hits in OSCAR)

then to do digitization

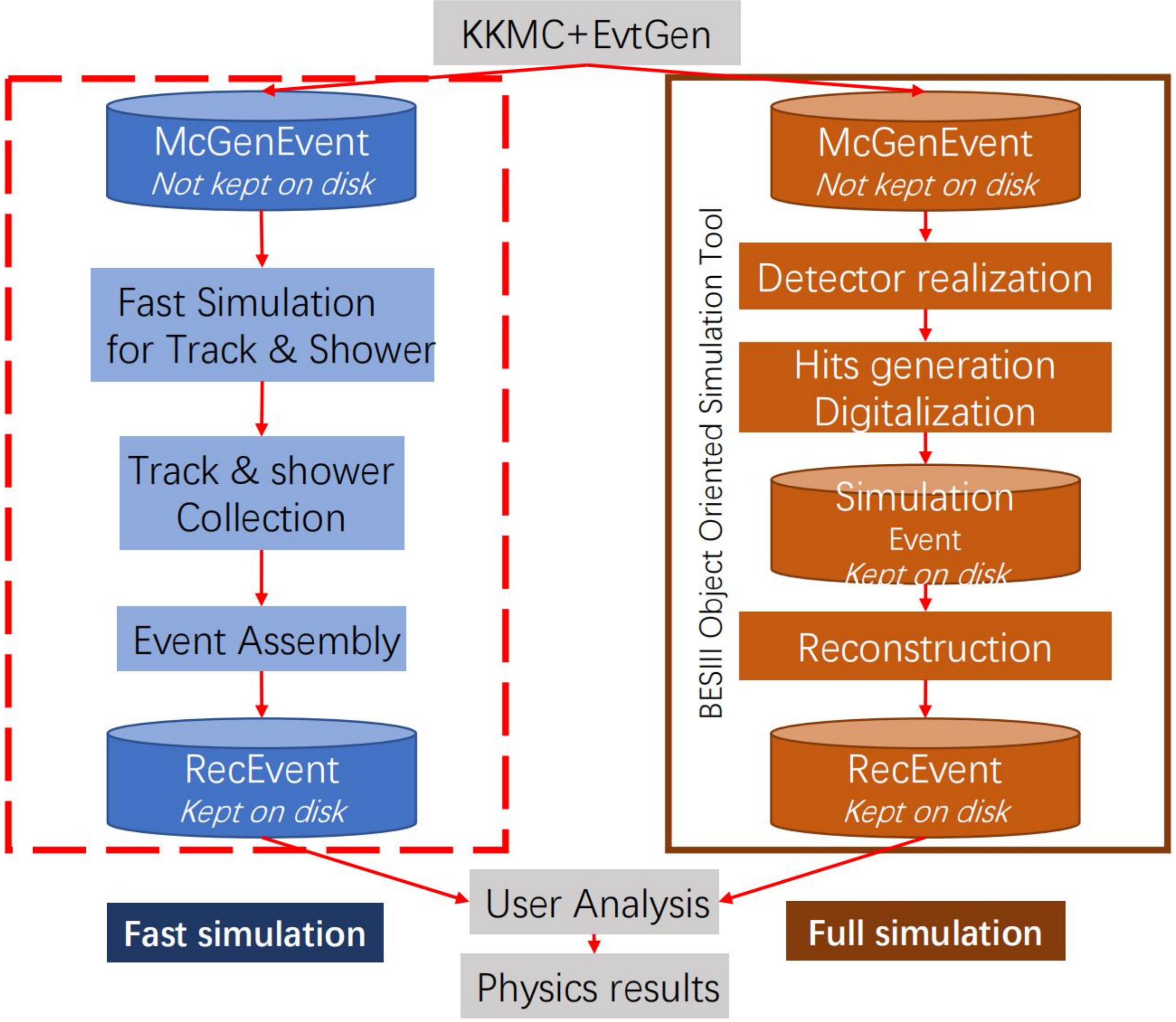
基于参数化的快速模拟方法

Details in Xiaoshuai's talk

- 通过全重建样本得到探测器性能：
  - 能动量分辨、位置分辨、重建效率、粒子鉴别效率等
  - 细分到特定 P, cos(θ), φ区间， 以便准确描述
  - 获得探测器响应的CDF曲线、效率直方图
- 利用探测器响应的CDF曲线、效率直方图进行抽样



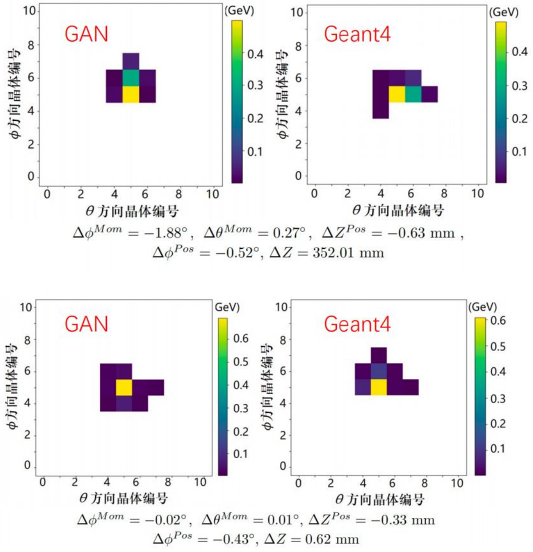
相对全模拟获得百余倍速度提升，十余倍磁盘空间节省



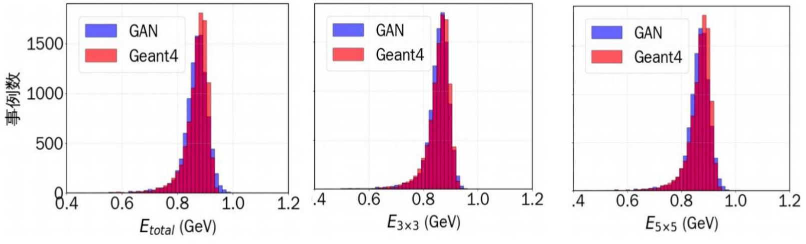
基于机器学习方法的ECAL快速模拟

- 将粒子在电磁量能器中的能量沉积视为图像，开发深度卷积生成对抗神经网络(DCGAN), 实现电磁量能器的快速模拟

■ 生成器：
输入随机噪声与指定粒子参数，输出生成数据
■ 鉴别器：
输入（真实与生成）数据及动量、入射方向等信息，输出评分
■ 回归器：
输入生成数据图像，输出动量、入射方向等信息



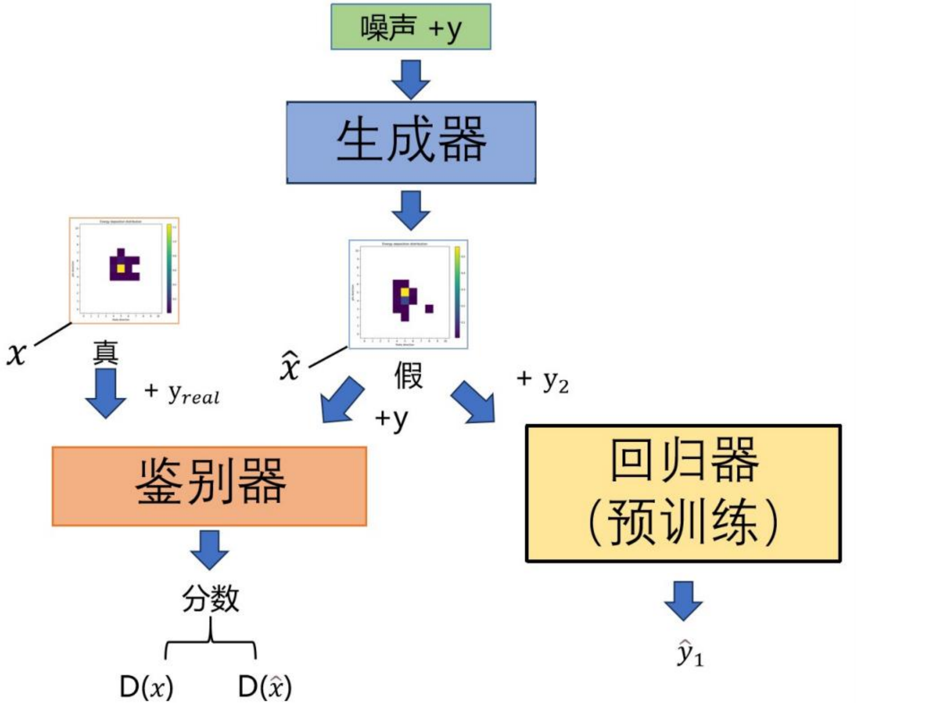
视觉效果对比



关键特征对比

模拟方法	平均时间 (ms)
Geant4 模拟	117.835
CaloGAN 模拟	5.54

模拟速度对比



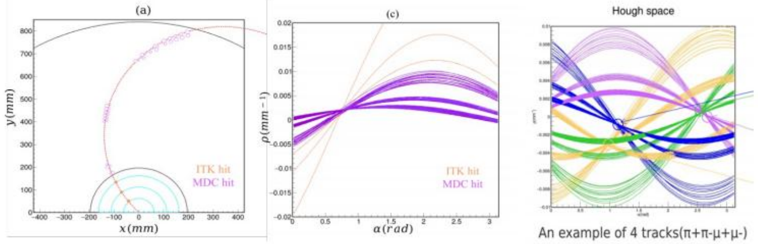
STCF-Calo-GAN模型架构图

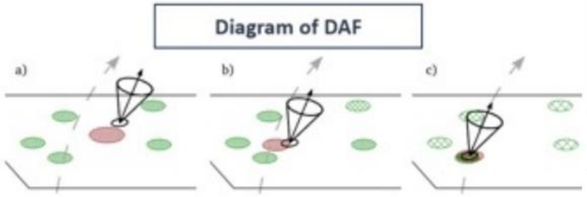
生成器损失函数：$\min_G E_{\hat{x}\sim p(fake)} \log\left(1-D(\hat{x})\right) + \|y_1-\hat{y}_1\|_1$

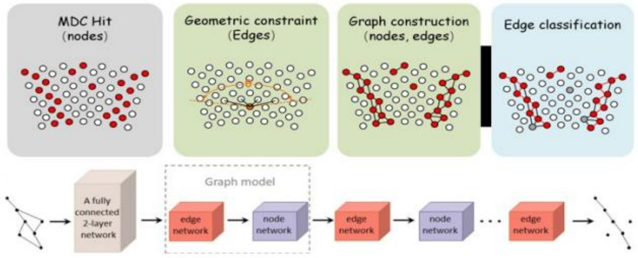
鉴别器损失函数：$\min_D E_{\hat{x}\sim p(\text{data})} \log\left(D(\hat{x})\right) + E_{\hat{x}\sim p(fake)} \log\left(1-D(\hat{x})\right)$

Tracking

基于霍夫变换的基准算法







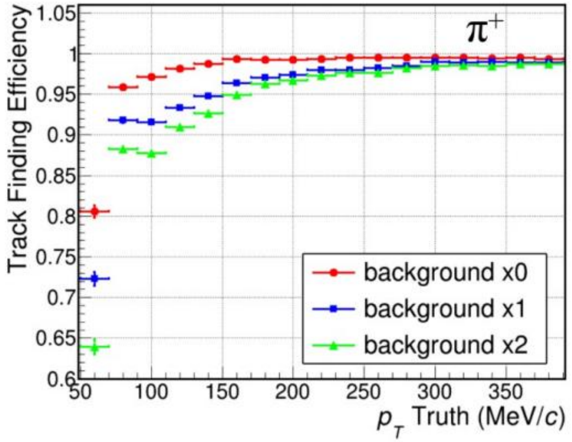
多种类型探测器测量 → GNN去噪 → Hough Transform 全局寻迹算法 → Genfit2的DAF径迹拟合算法 → 径迹

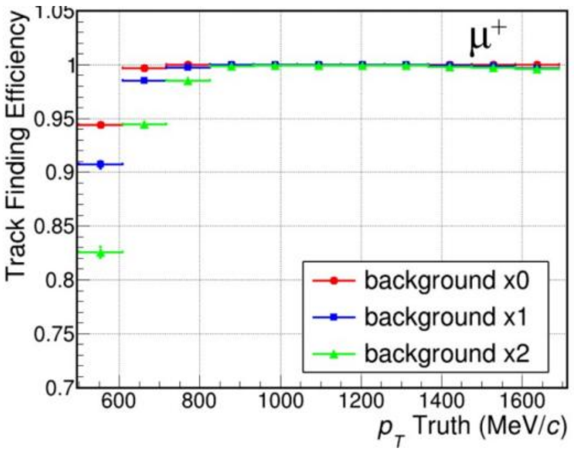
基于霍夫变换的全局寻迹(Hough)+ 自适应的卡曼滤波(DAF in Genfit2) + 图神经网络去噪(GNN)

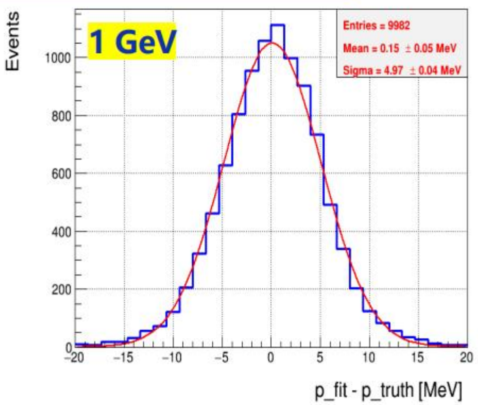
特点

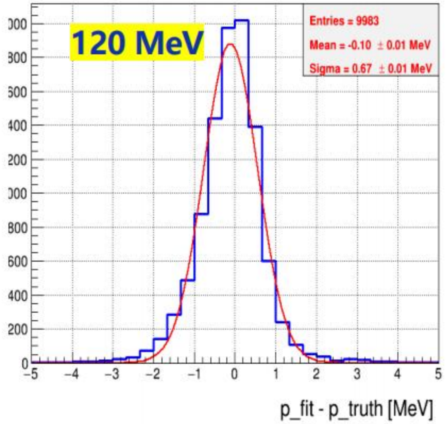
- 覆盖全动量区间，多径迹事例
- 长寿命粒子：突破传统霍夫变换的局限
- 优化低动量径迹的重建效率
- 排除噪声击中、提升径迹分辨

Tracking performance from ψ(3686) → π+π- J/ψ, J/ψ → μ+μ- ，OSCAR 模拟+重建，ITKW + MDC，smear







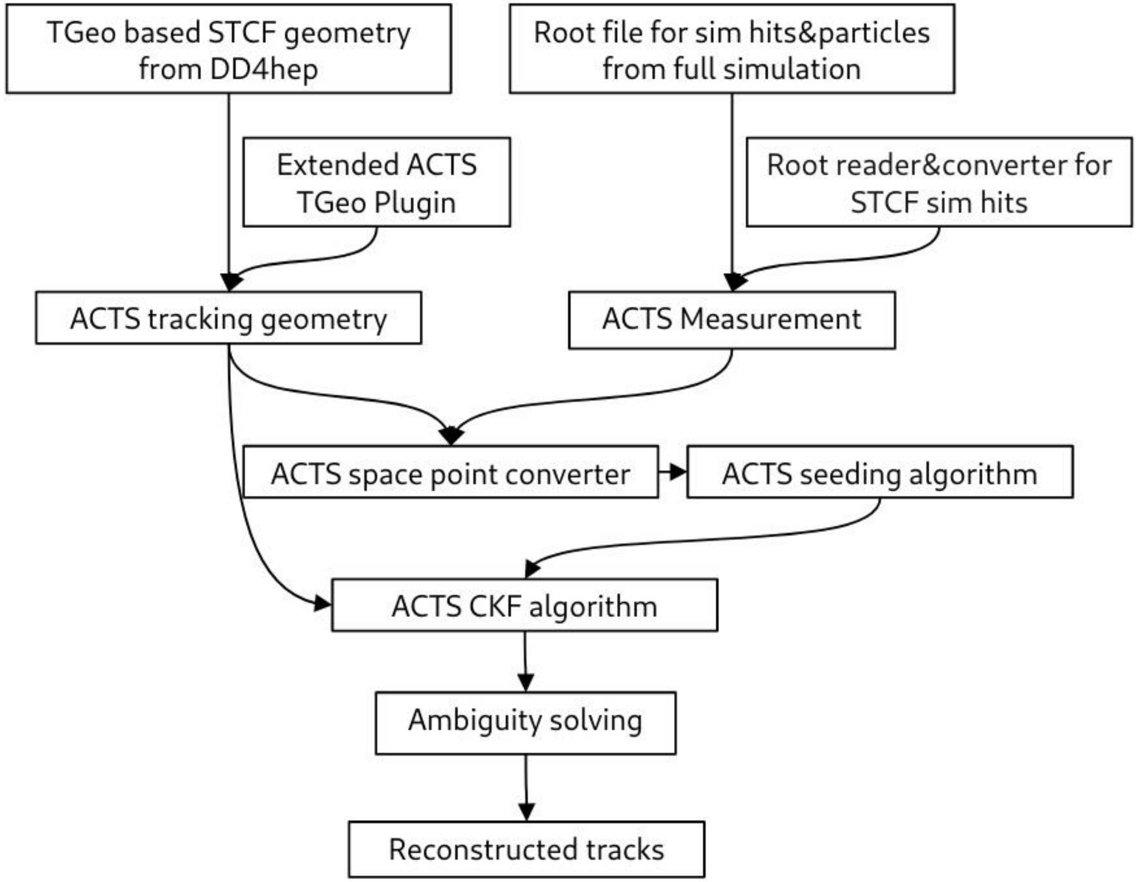


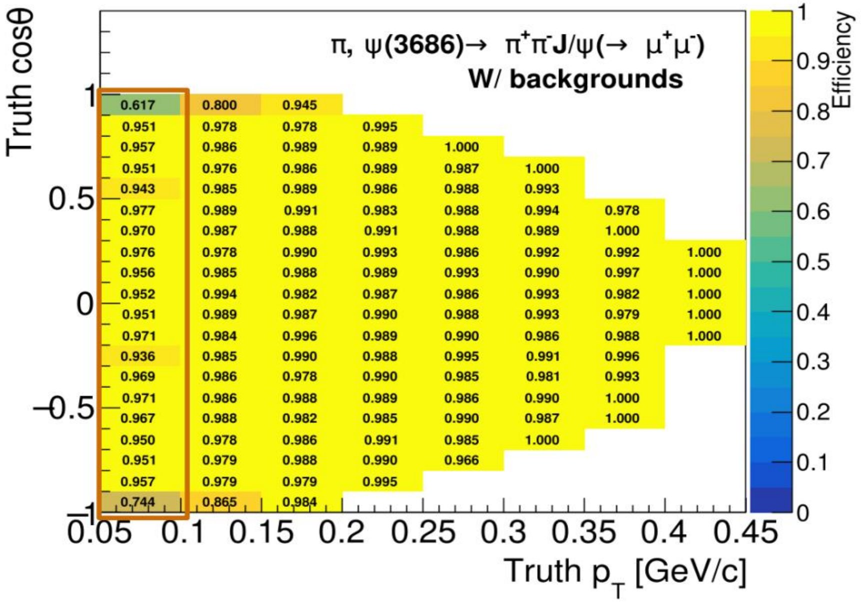
π寻迹效率达到 95%/90% without/with 1X background @100MeV;
接近99% without/with 1X background @300MeV;

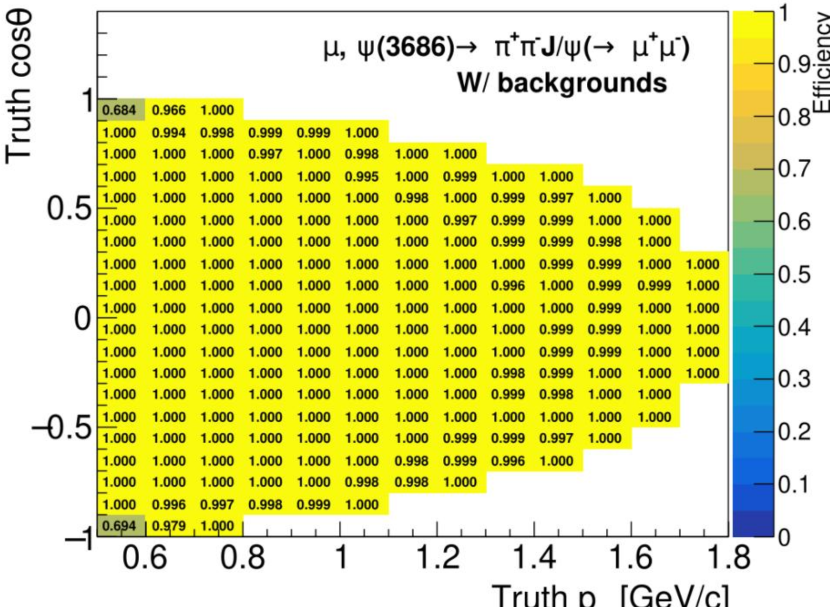
高动量径迹/低动量径迹均达到合理的分辨水平
(1GeV muon 好于0.5% w/ w/o background)

基于ACTS的径迹重建

ACTS application strategy in STCF







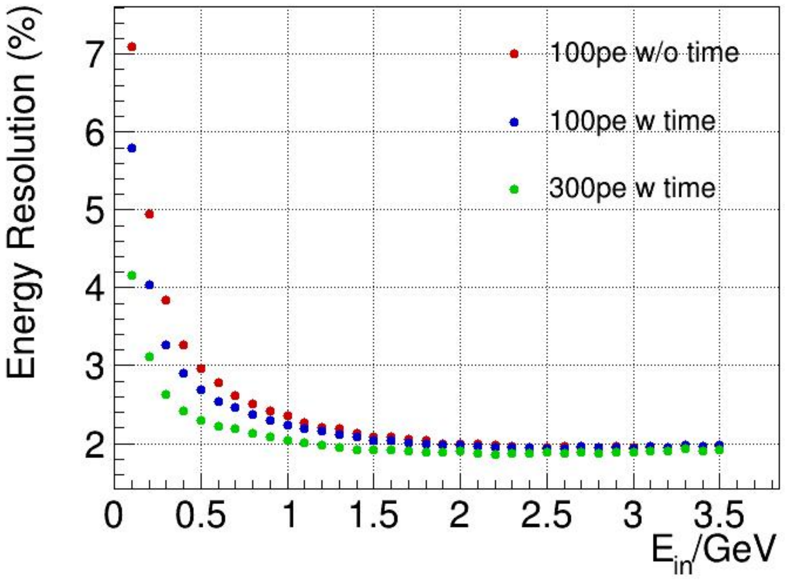
- Above 99% efficiency for p_T > 400 MeV
- ~94% efficiency for pion with p_T < 100 MeV

with background

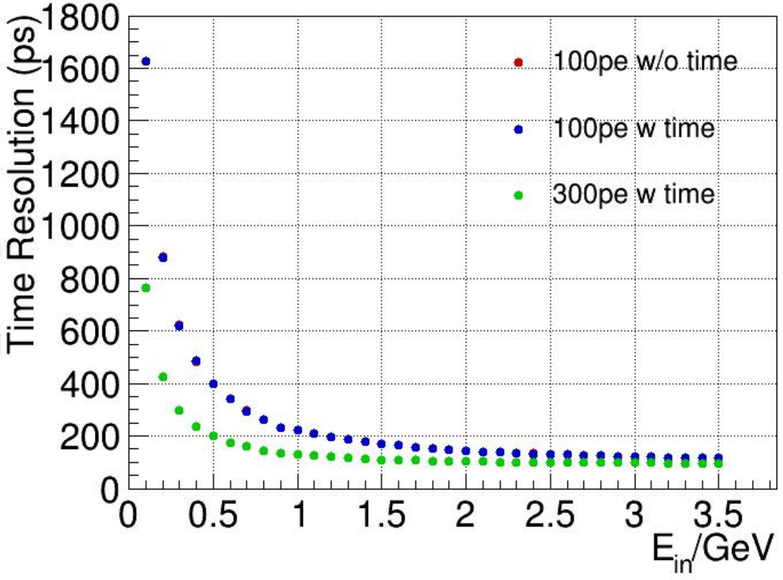
ECAL

Photon reconstruction performance

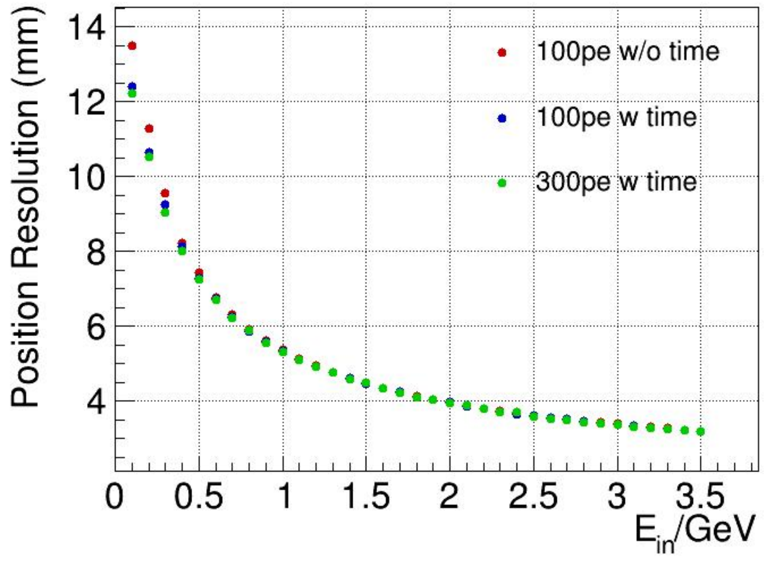
- EMC digitization algorithm has been optimized with **50% reduction** of CPU consumption
- Using realistic digitization parameters from beam test results
- Single photon energy, time and position resoluton with backgrounds can meet physics requirements



Energy



Time

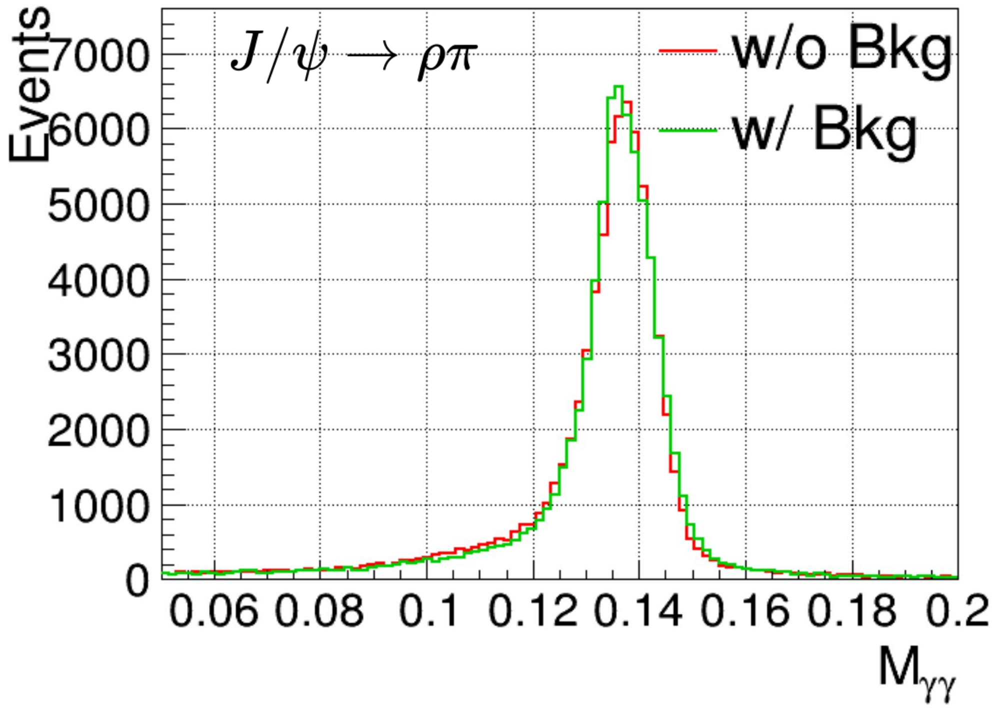


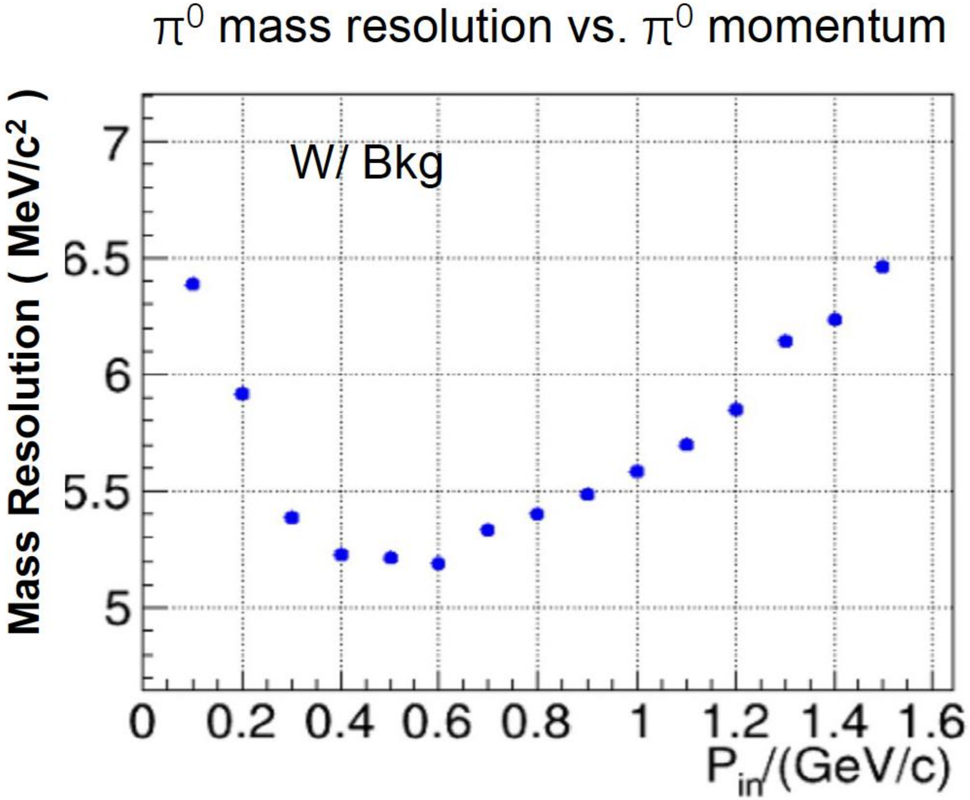
Position

Photon reconstruction performance

- Make use of time information to improve photon resolution against beam backgrounds
  - No degration of π^0 resolution with beam backgrounds

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

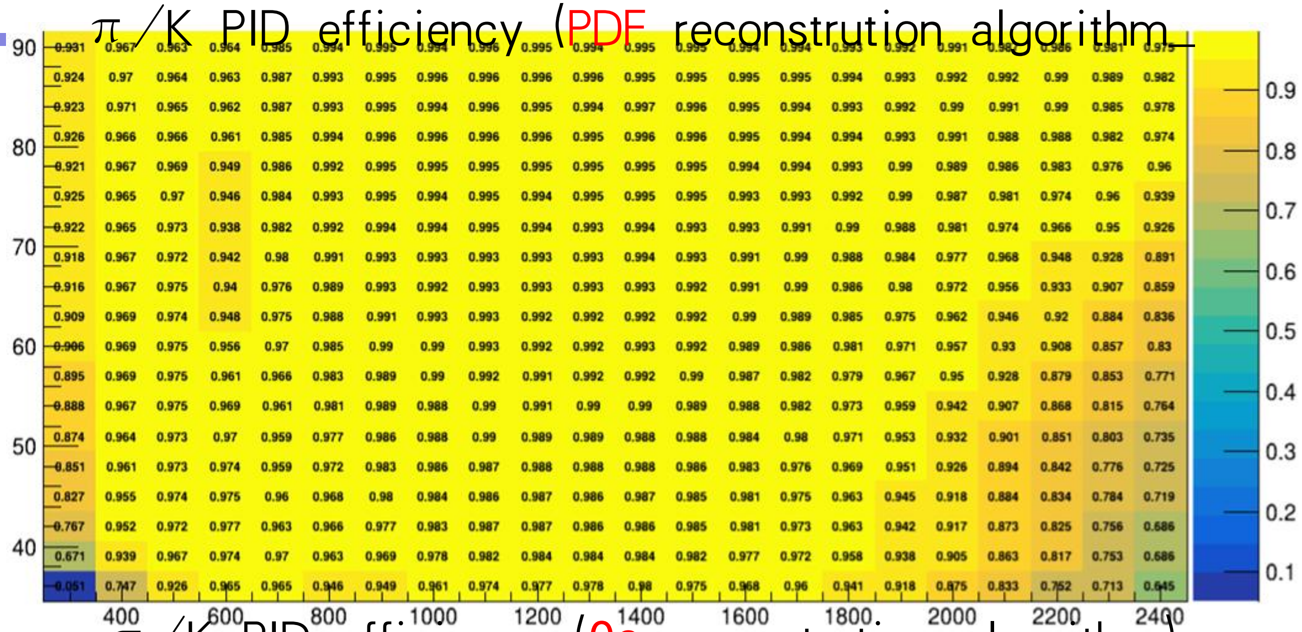


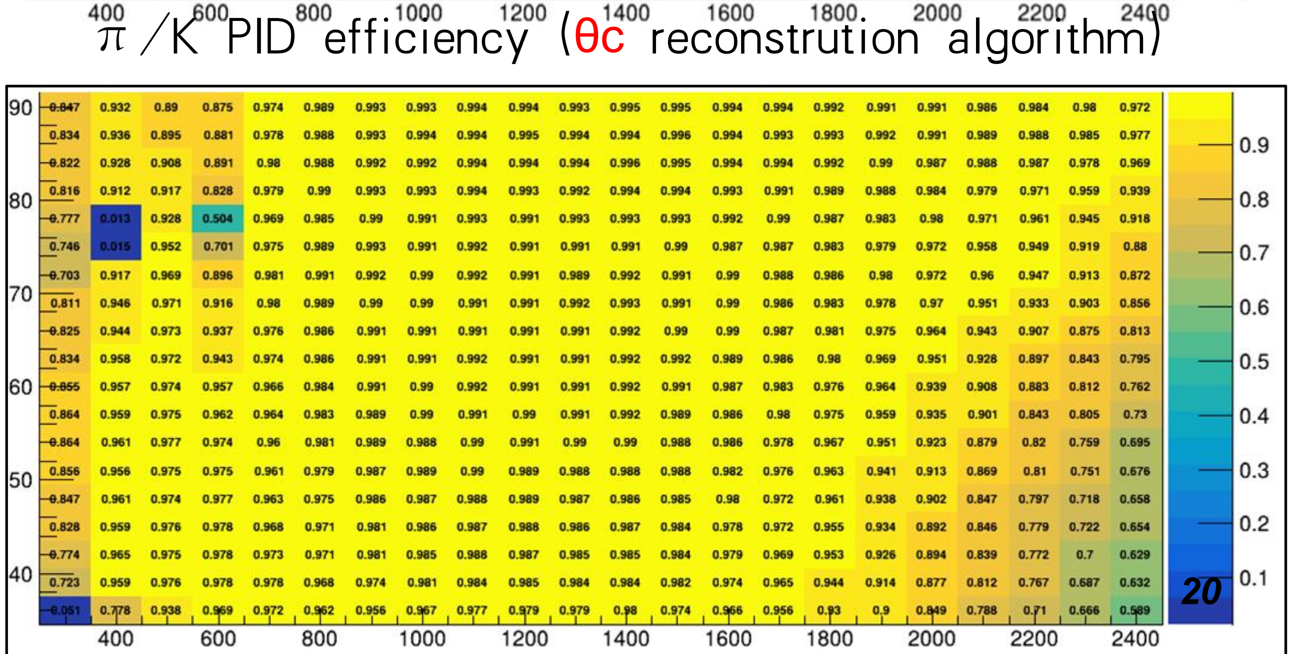


Particle Identification

π/K PID performance (RICH)

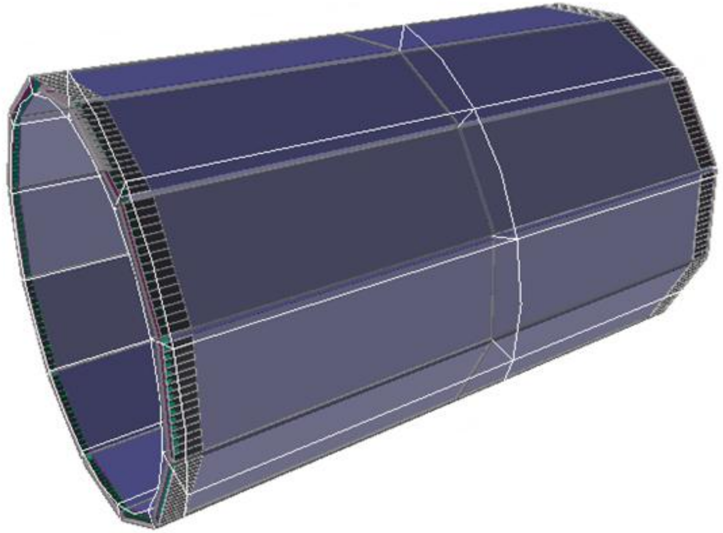
- Two RICH reconstruction algorithms (PDF/thetaC reconstruction algorithms)
- Achieving >97% π eff @ K mis-ID=2%
  - PDF: better performance
  - θc reconstruction: low CPU cost



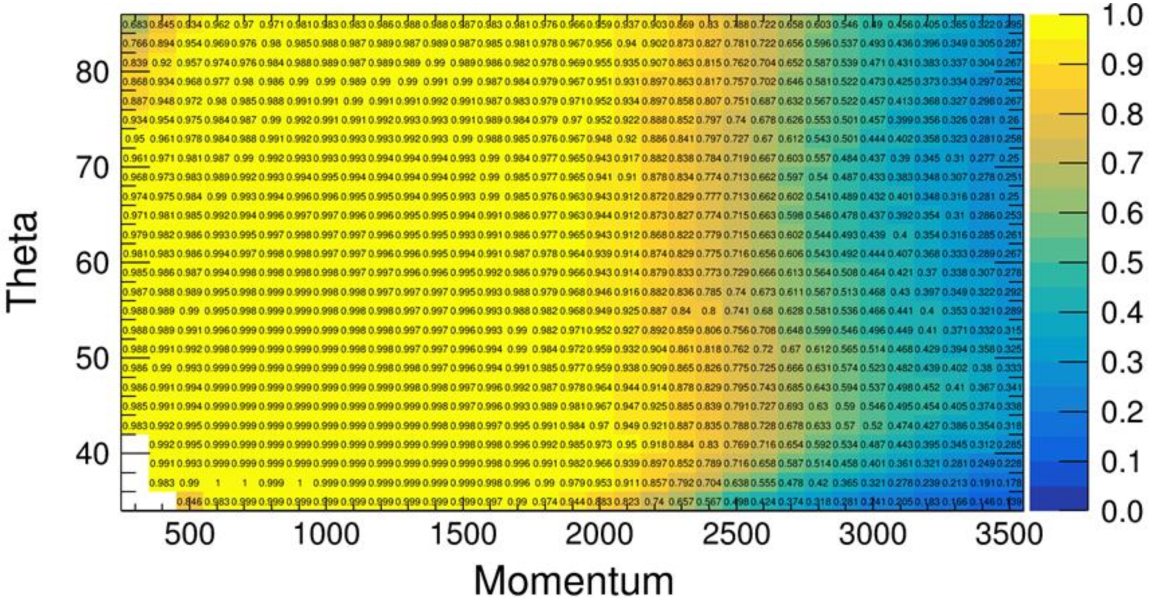


π/K PID performance (BTOF)

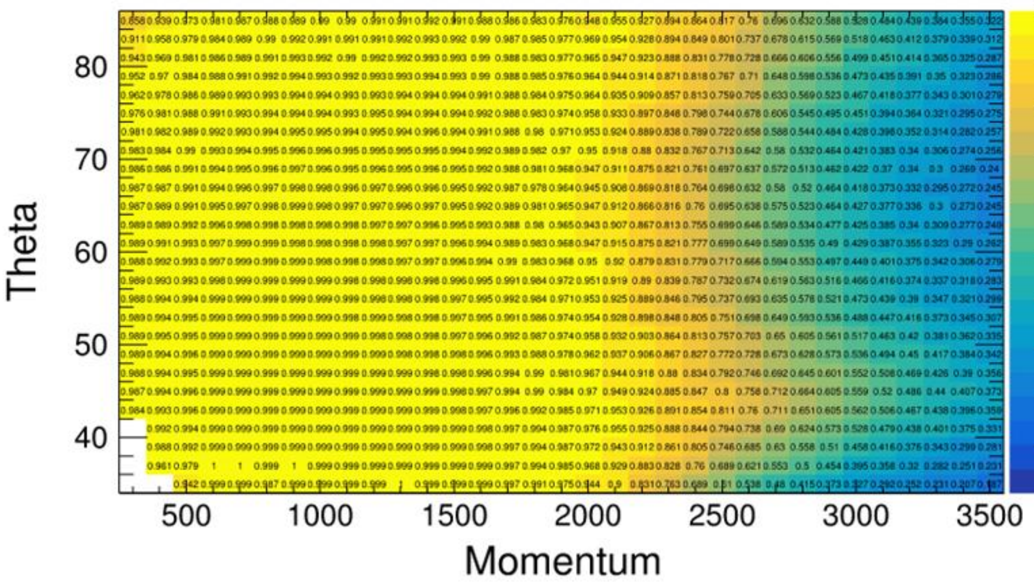
- Developed the full simulation, digi and reco chain for BTOF (geometry parameters is under optimization), the performance fufills the STCF requirement (>97% π eff @ K mis-ID=2%)
- TODO: optimize design to improve performance



Effi_pi(1 Bkg), timing method

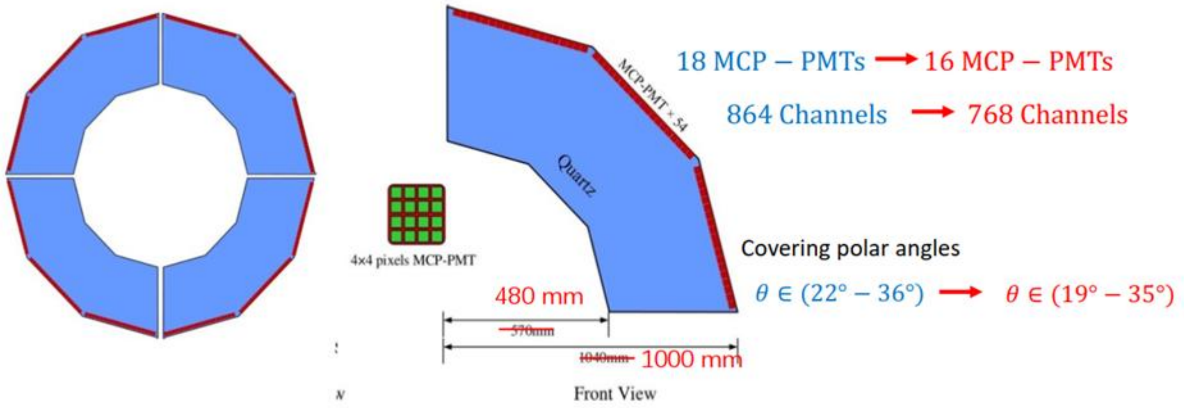


Imaging method

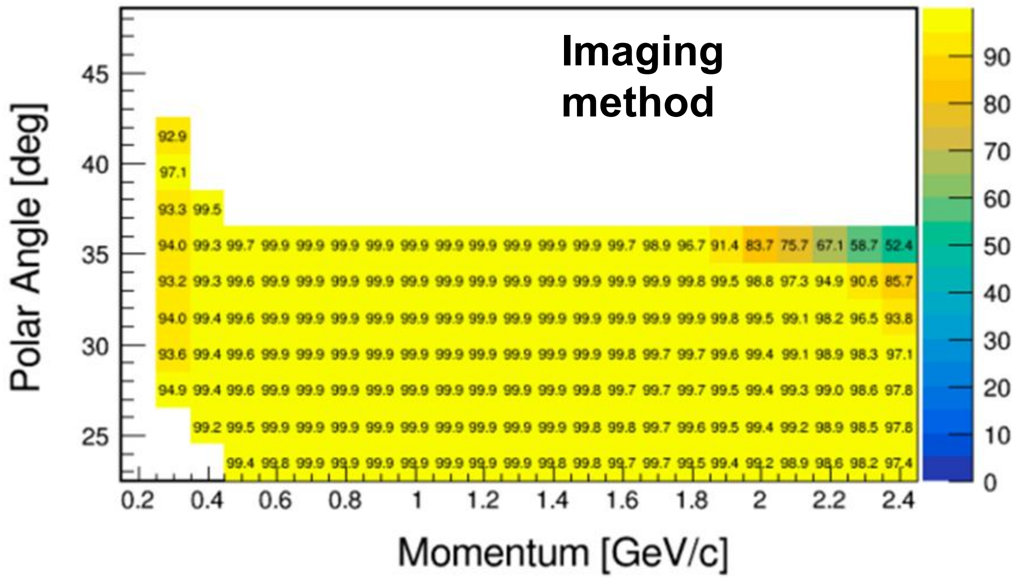


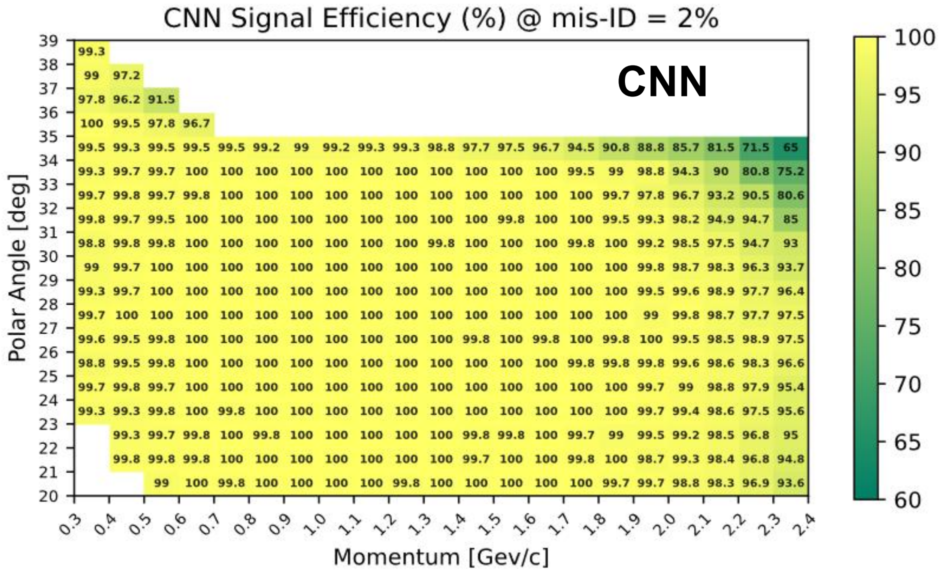
π/K PID performance (DTOF)

- Performance with optimized geometry is validated
- **CNN** is deployed to improve the performance in region with large theta and momentum



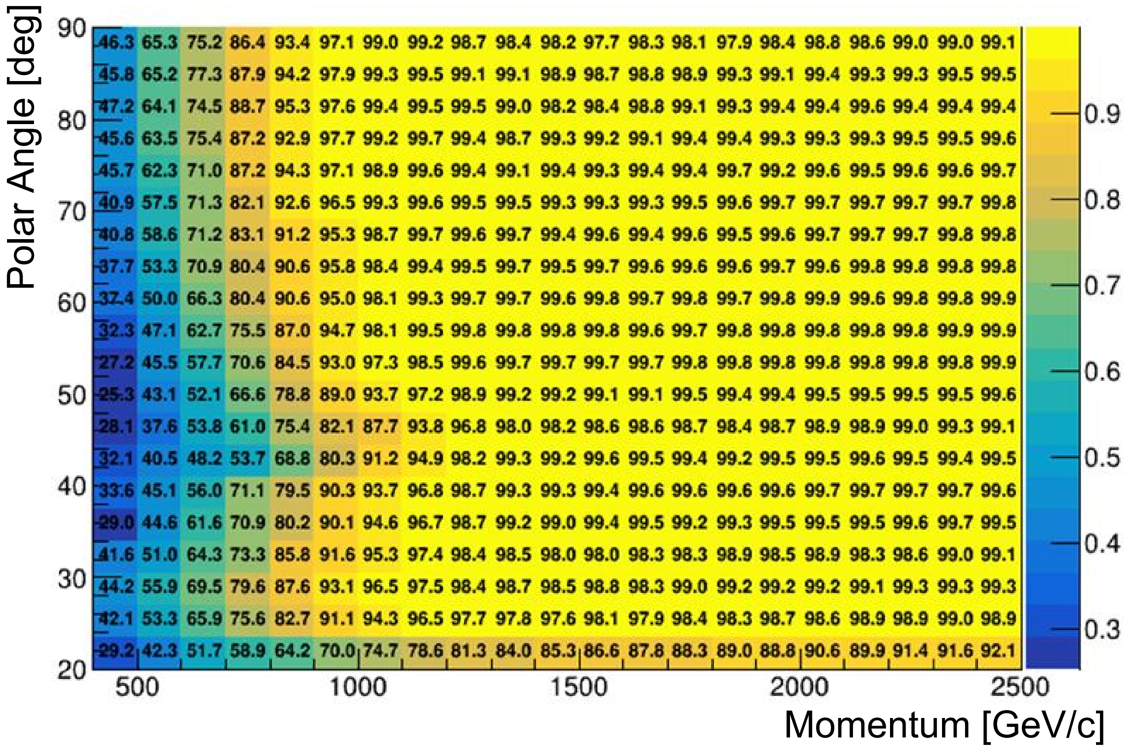
Old Geo



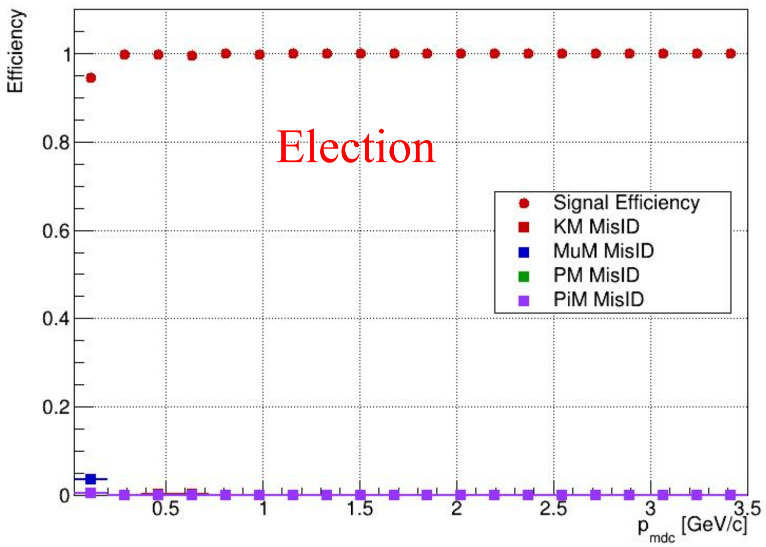


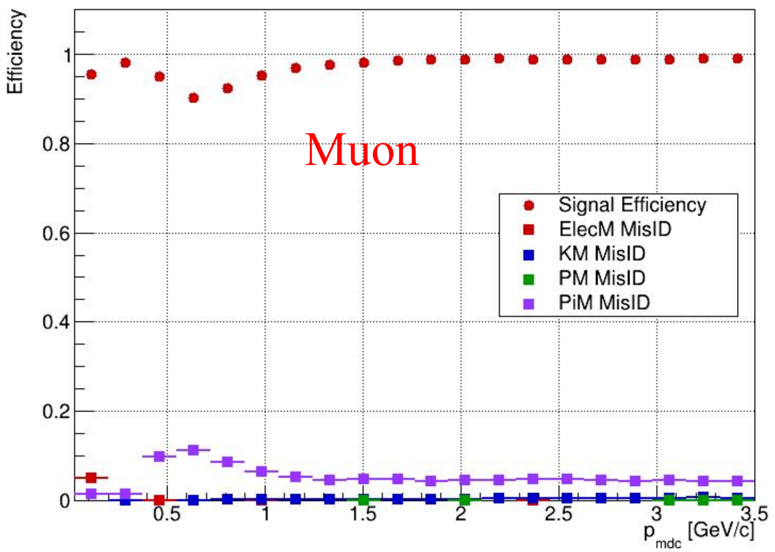
μ/π PID performance (MUD)

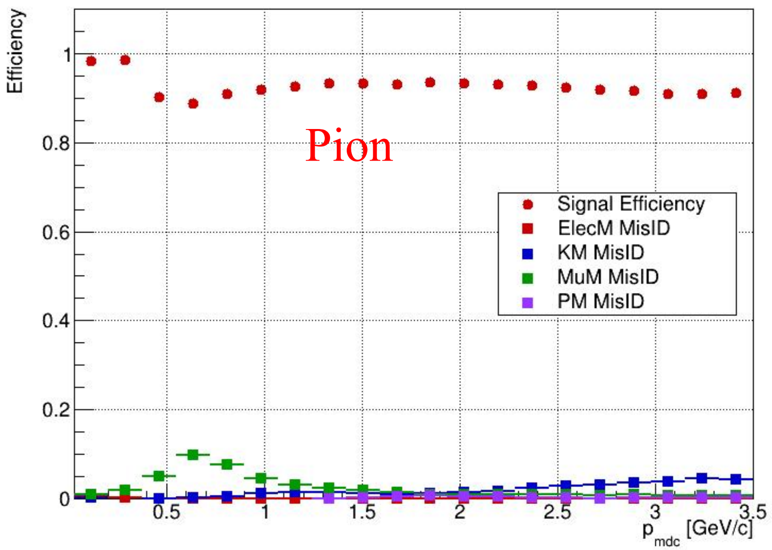
- Obtained μ/π identfication efficiency using **BDT o**n top of realistic simulation and digitization
  - μ efficiency is above **95%** @suppression = 30 with momentum above **1 GeV** in barrel region
  - Detector geometry is under optimization to improve the performance in barrel-endcap transition region
- Identfication between neutral hadron (n, K_L) and photon is being optimized

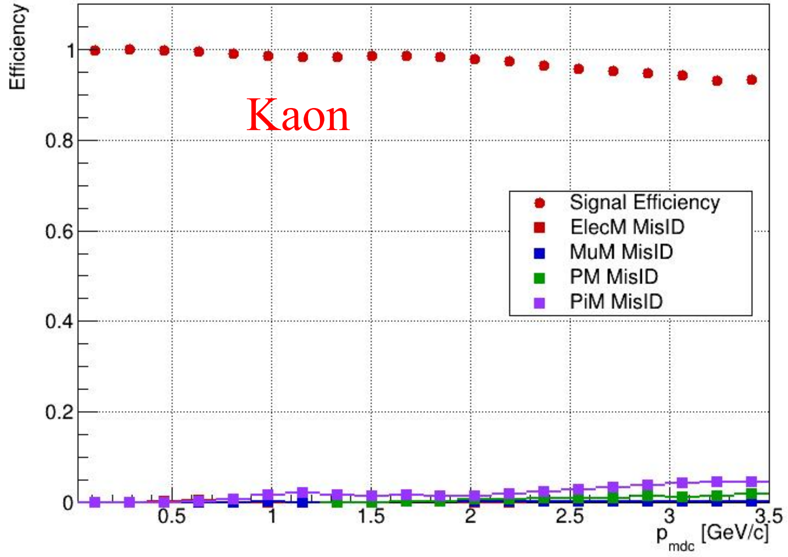


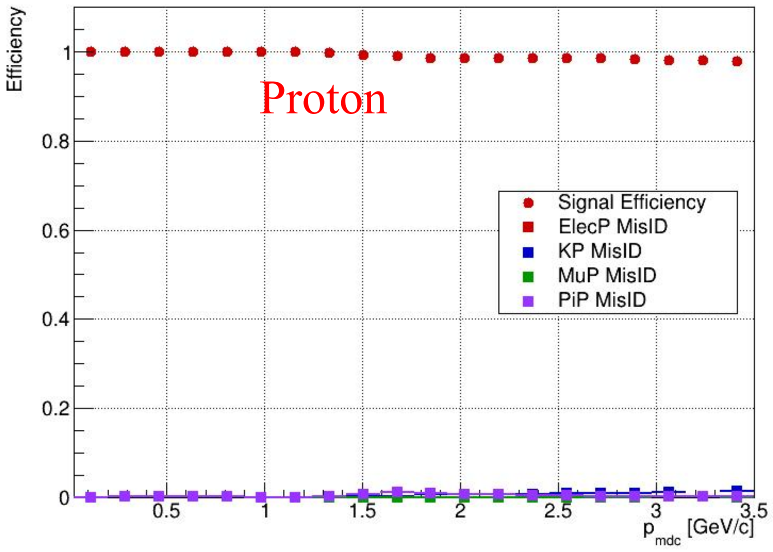
Machine Learning GlobalPID











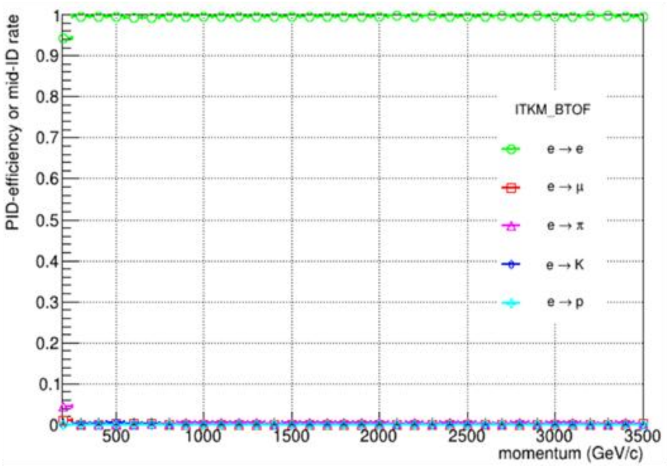
联合谱仪全部子系统，基于BDT训练机器学习模型

基本达到指标要求

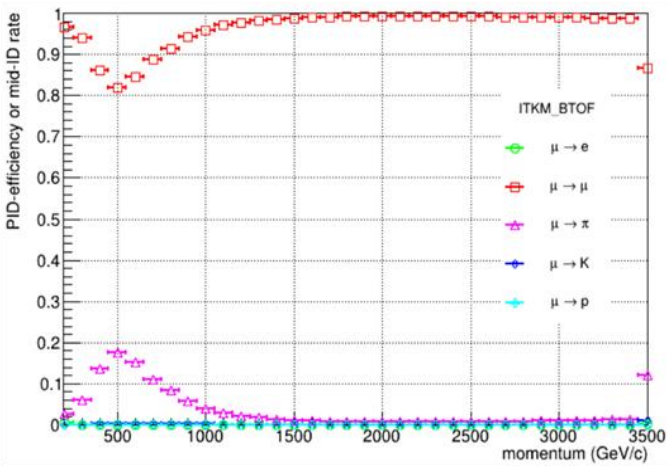
Weighted likelihood GlobalPID

- 联合所有探测器重建输出likelihood，实现**e/mu/pi/K/p五种带电粒子鉴别**
- 相较ML方法具备更好的**鲁棒性**
- 根据不同探测器对不同粒子的鉴别能力，分别**对不同likelihood设置不同权重**，实现性能提升

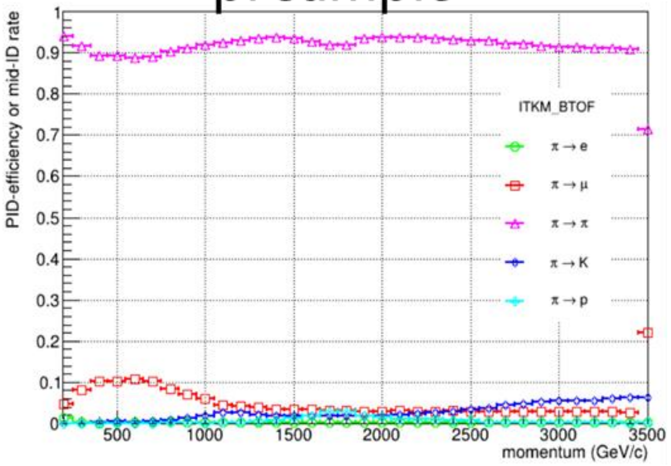
e sample



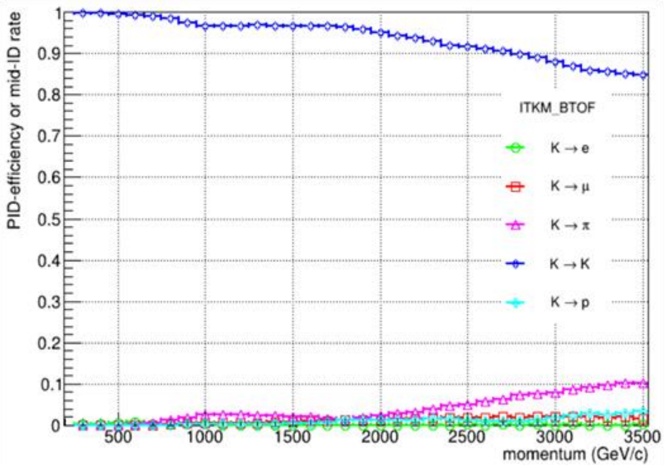
mu sample



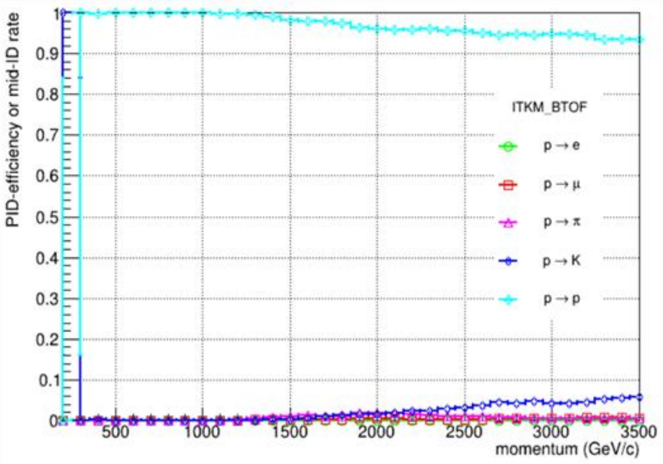
pi sample



K sample

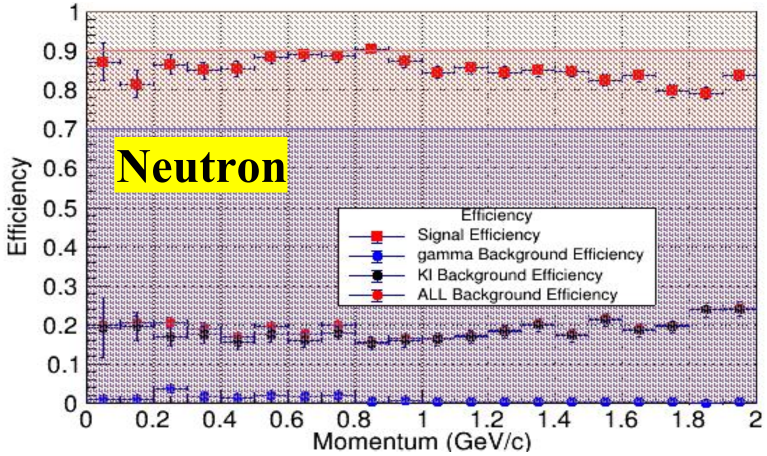


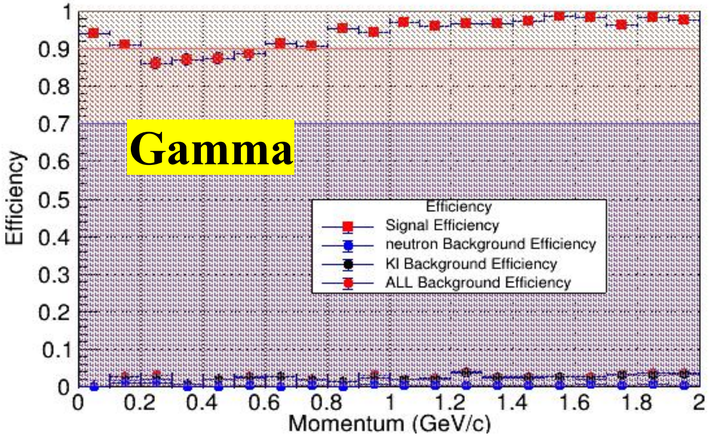
p sample

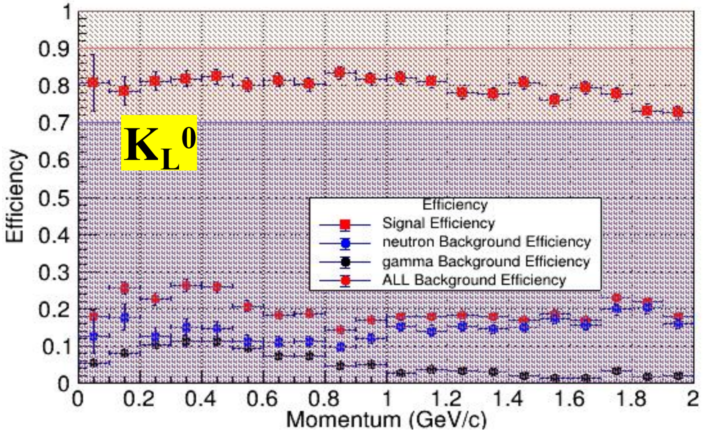


中性粒子鉴别

- 利用ECAL Shower形状，基于卷积神经网络初步开发γ/n/K_L鉴别算法

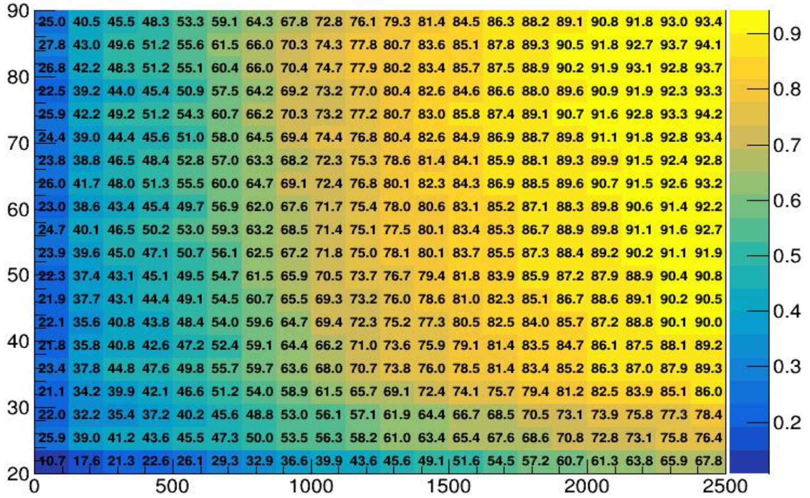






- 基于BDT开发MUD neutron/γ粒子鉴别算法

theta



neutron的鉴别效率 @ γ supp=30 momentum (MeV)

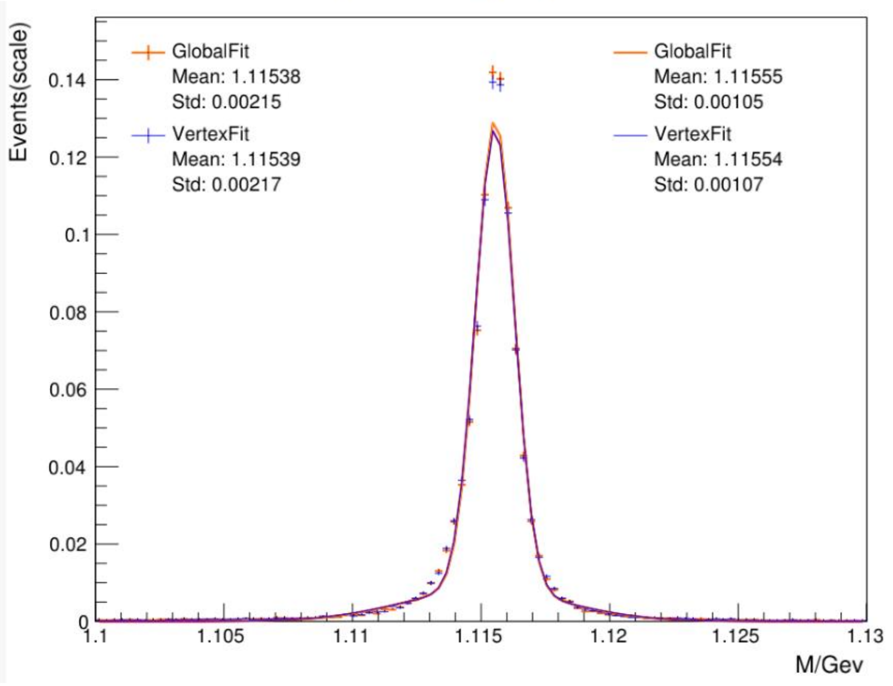
- 目前已经初步实现 γ /n/K_L鉴别
- 未来可联合ECAL和MUD增强中性粒子鉴别性能

分析工具

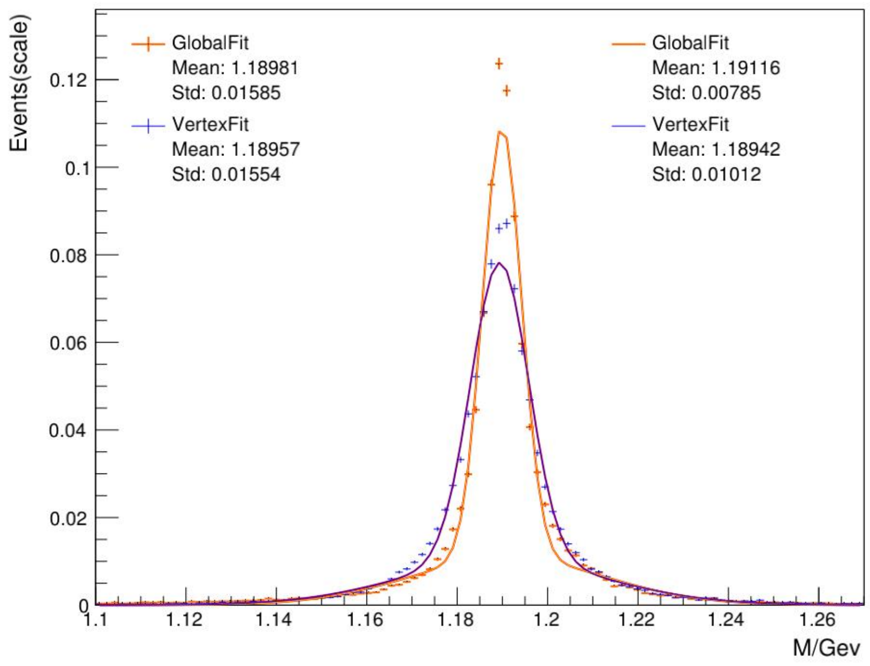
运动学和顶点拟合

- **VertexFix** package ported from BESIII, now stable
  - Further improvement being developed
- **GlobalFit** package for global vertex fitting algorithm ported from **Belle II**, showing slightly better performance

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



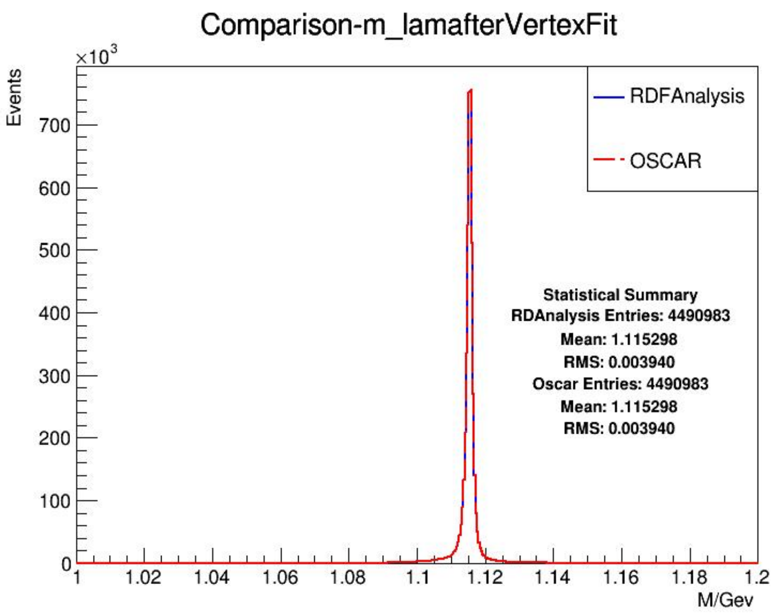
$J/\psi \to \Sigma^+(Pr)\bar{\Sigma}(\pi_0(rr)\bar{P})$

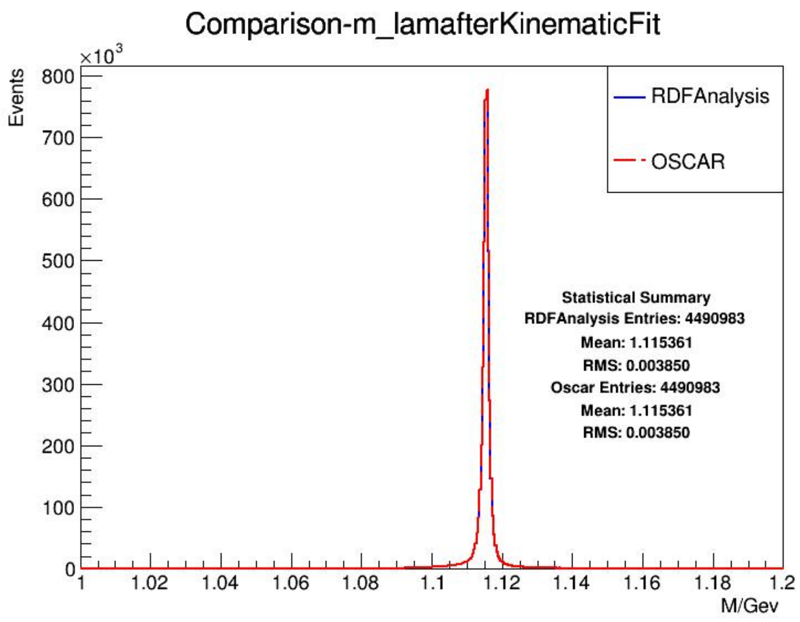


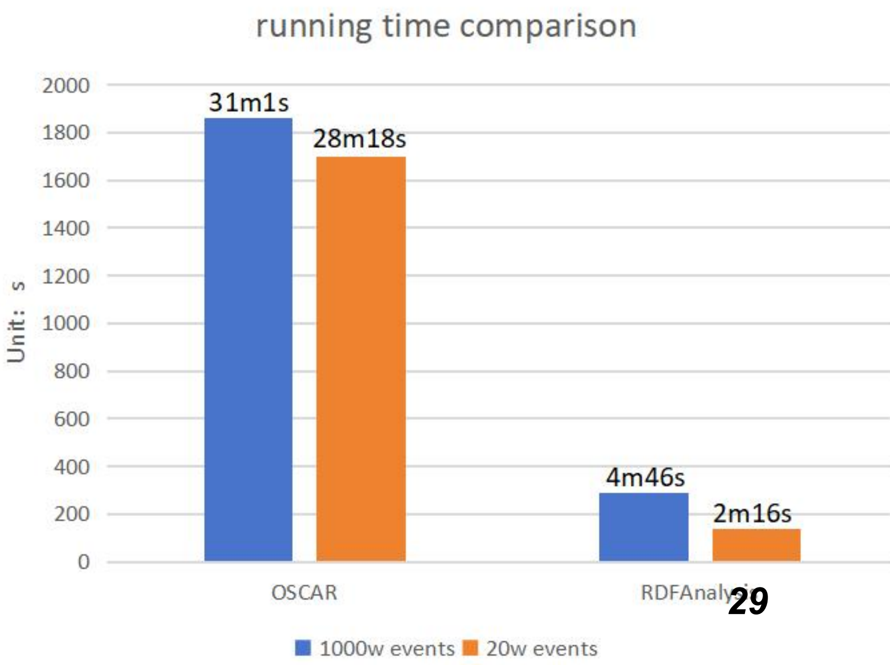
RDataFrame Analysis Framework

- The physical analysis results of process $J/\psi \to \Lambda\bar{\Lambda}$ using 10 million events are **consistent** with the traditional methods.
- Select the dataset and submit the job to test running time. The running speed of **16 parallel threads** is **14 times** that of the traditional method.
- Higher performance will be achieved with the **analysis EDM**

dataset	
total number of events	1×10^7
total size	169GB
number of events per file	2×10^5
size per file	3.38GB

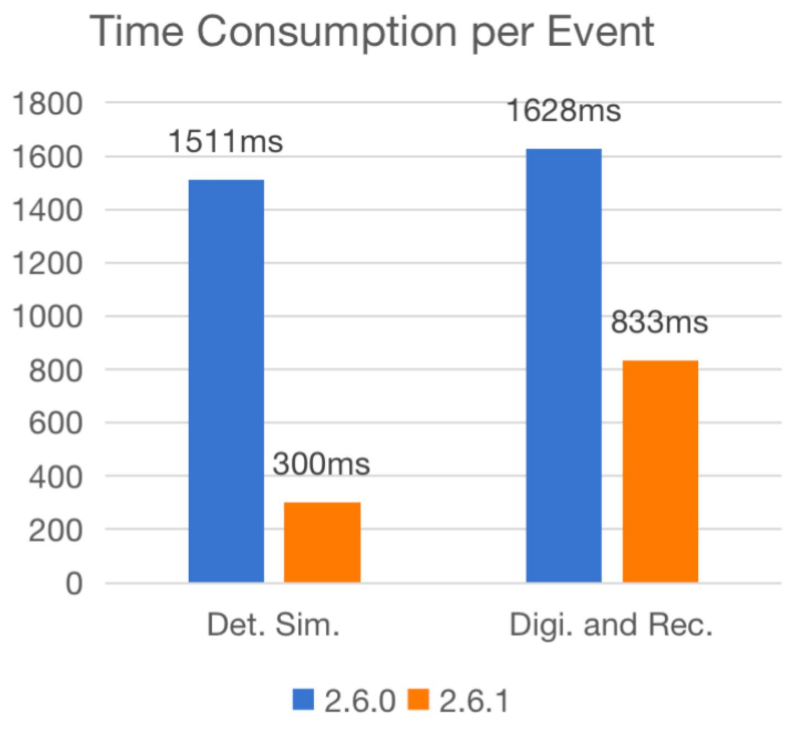


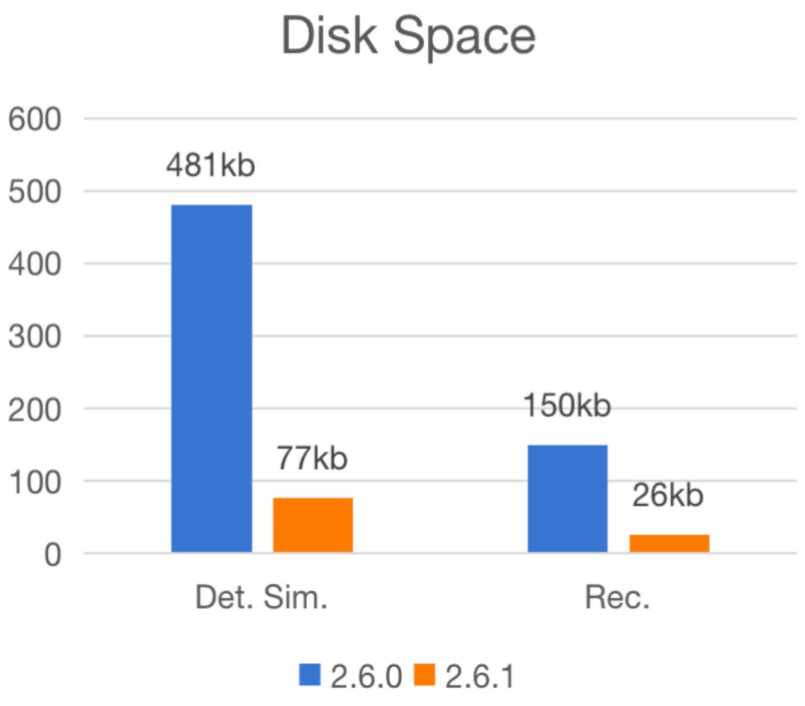


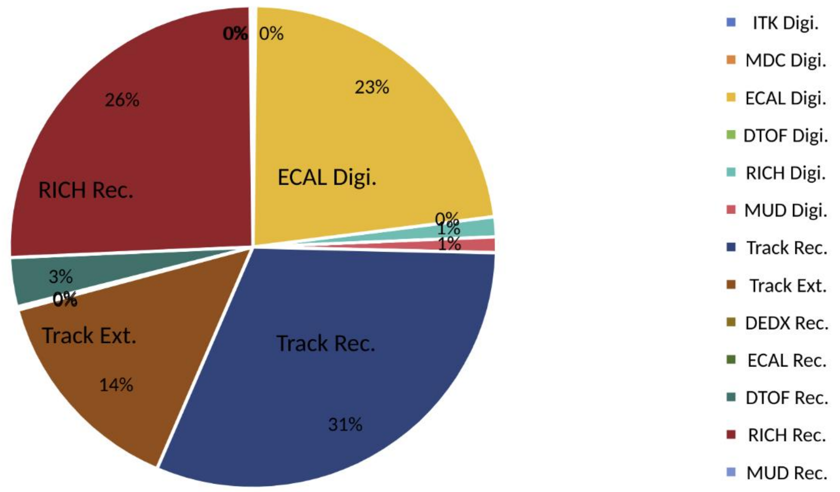


OSCAR性能

当前版本：2.6.2







e+e- -> pipijpsi -> pipimumu

探测器模拟：**速度 300ms/evt，空间 77kb/evt**

数字化和重建：**速度 833ms/evt，空间 26kb/evt**

Summary and Plan

- OSCAR (2.6.2) can stably support the full MC simulation chain (det. sim. -> digi. -> rec. -> analysis)
  - Performance has been improved a lot since 2.5.0
- Recent development plans of OSCAR
  - Support hardware optimization as first priority
  - Improve the fast simulation framework for quick iteration
  - Core software updates:
    - Further improvement of the EDM
    - Upgrade of the outdated external libraries & support for the el9 os
    - Upgrade of the automated validation system