STCF Offline Simulation and Reconstruction Software

Xingtao Huang (黄性涛)[1], Xiaocong Ai (艾小聪)[2]
on behalf of the STCF Software Group

1. Shandong University
2. Zhengzhou University

中国物理学会高能物理分会第十四届全国粒子物理会议，2024年8月15日，青岛

The Super Tau Charm Facility

From H.P. Peng's talk

Requirements of offline software at STCF

Higher event rate, background, CPU consumption at STCF than BESIII

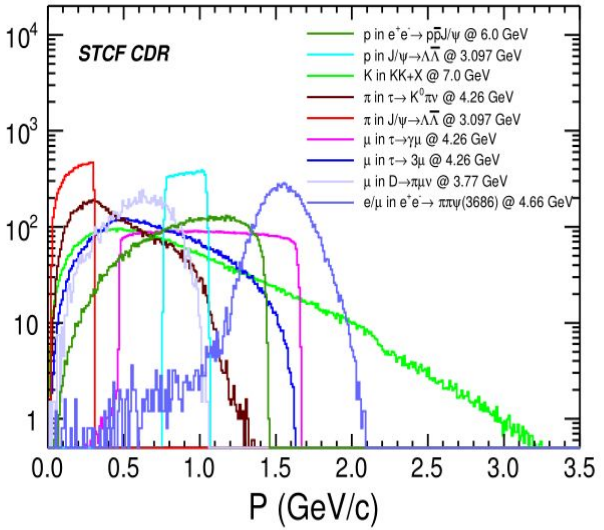
→ i.e. we need **reconstruct** the tracks and photon **with good efficiency and resolution**, and **identify** them **at high accuracy**, with **good speed**

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

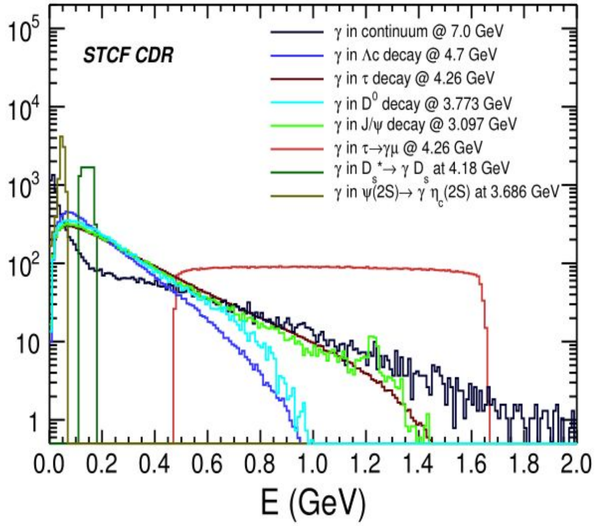
Particles at STCF

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

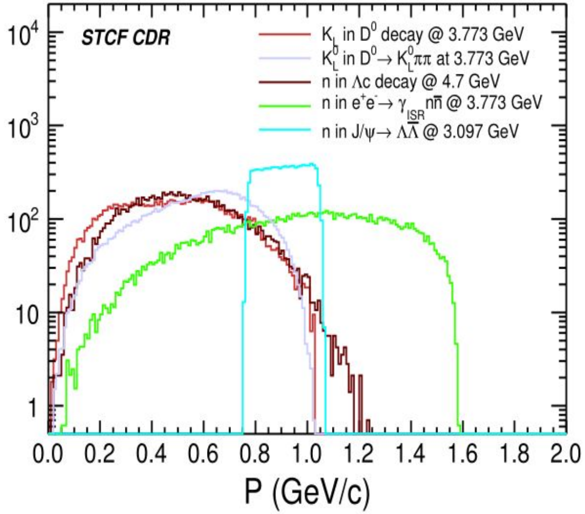
Charged particle momentum



Photon energy

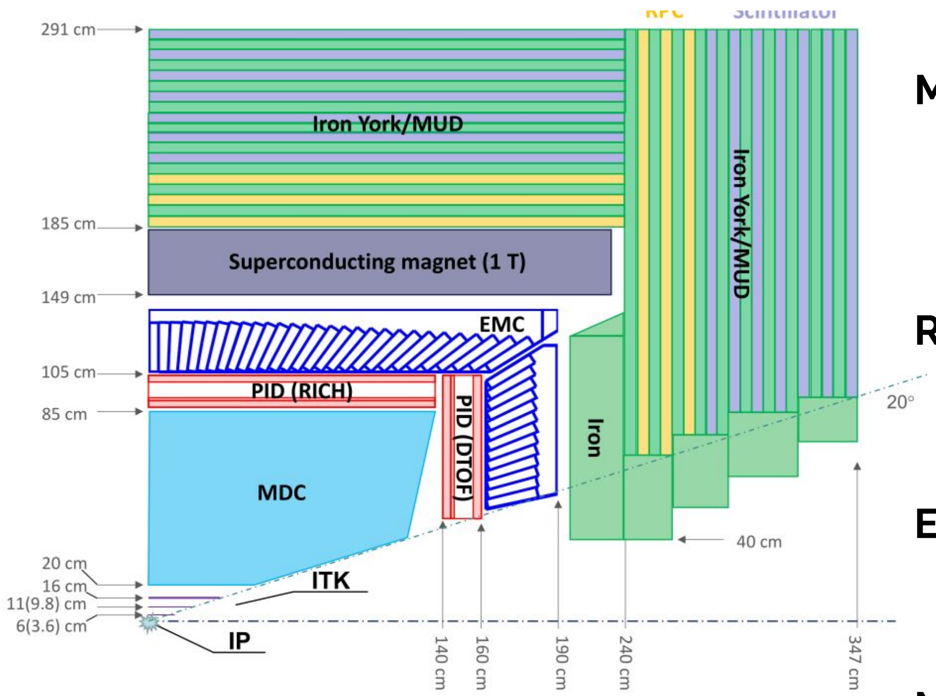


K_L, n momentum



From STCF CDR, Front. Phys. 19, 14701 (2024)

The detector and performance requirements



See an overview in J.B. Liu's talk

ITK (cylindrical MPGD/ CMOS MAPS)

- Material < 0.01 X0, σ_{xy} < 100 um

MDC (drift chamber) (more in M. Peng's talk)

- Material < 0.05 X0
- σ_{xy} < 130 um, σ_p/p < 0.5% at 1 GeV/c
- dE/dx resolution < 6%

RICH (CsI-MPGD) & **DTOF** (DIRC-like TOF)

- PID π/K PID efficiency > 97% up to 2 GeV/c @mis-ID rate 2% (more about RICH in Q. Liu's talk)

EMC (pure CsI + APD) (more in B. Wang's talk)

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

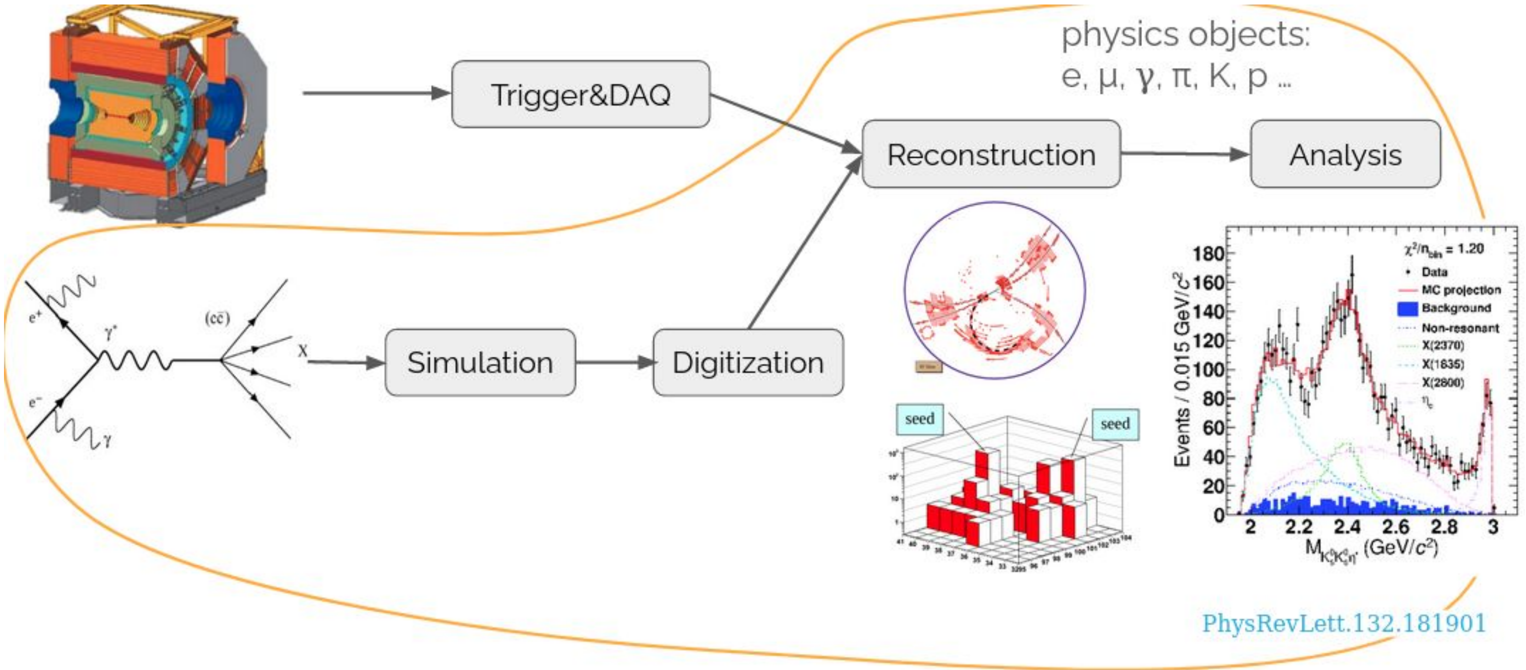
MUD (RPC + scintillator strips)

- μ PID efficiency > 95% with π→ μ mis-ID rate < 3.3% @ p = 1 GeV/c

STCF offline software OSCAR

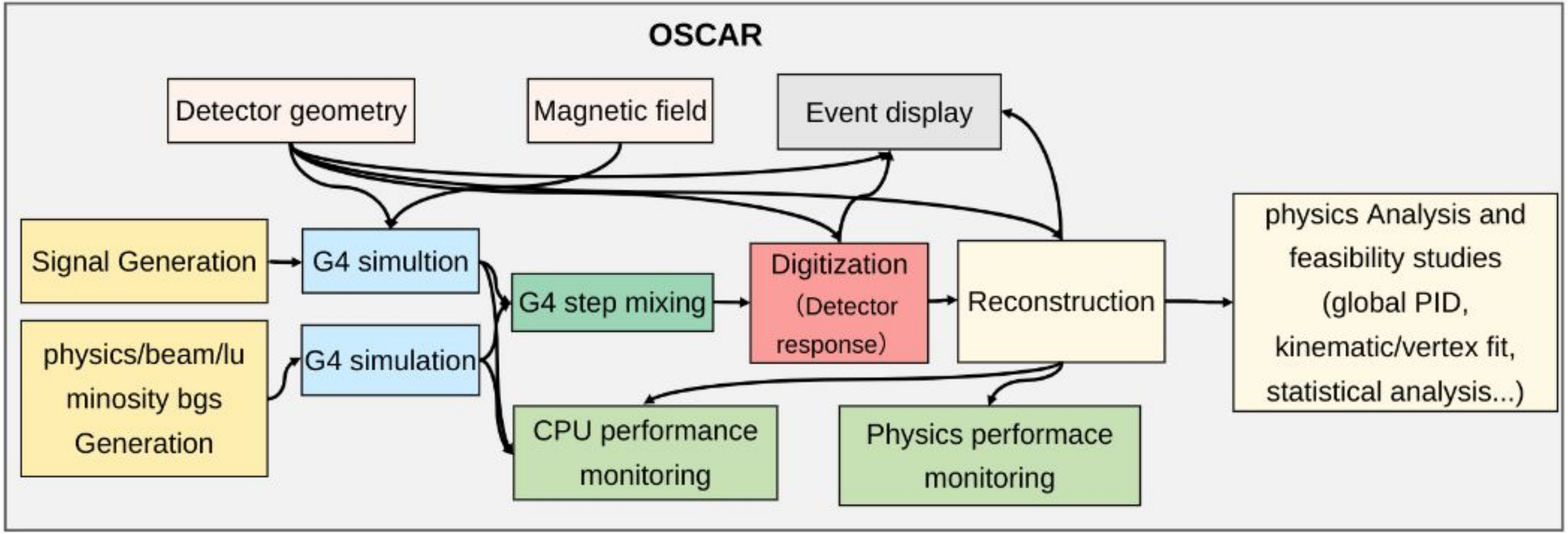
The crucial role of HEP offline software

- Offline software of experiment is the "**converter**" from detector data to physics data to make all physics analysis and possible!
- Plays a crucial role for detector **design**, **alignment** and defects/aging **detection**



The Offline Software of Super Tau-Charm Facility (OSCAR)

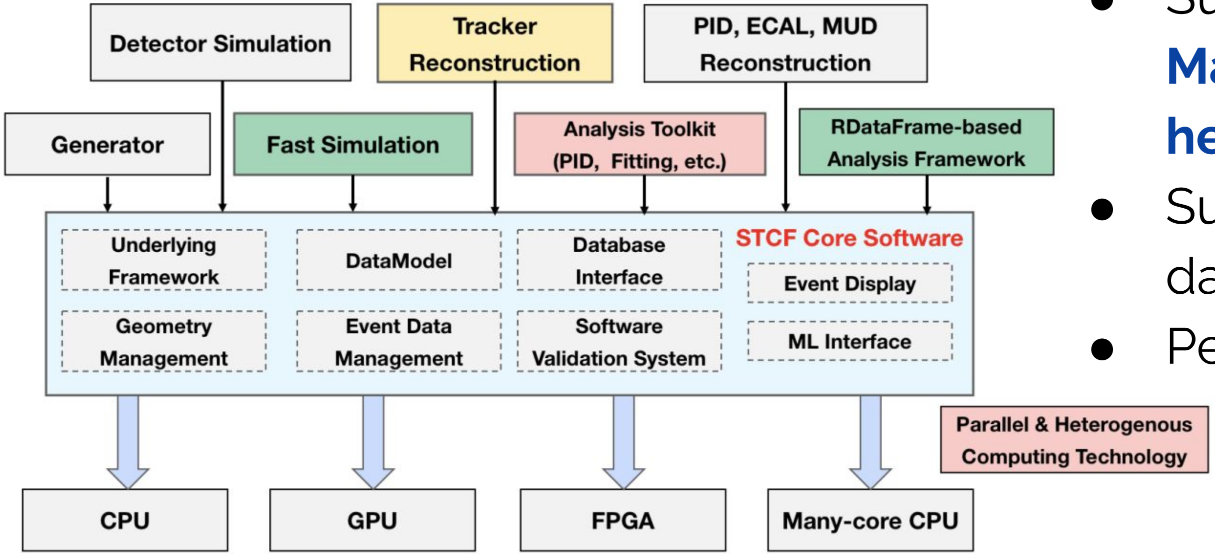
- OSCAR Provides common functionalities for **detector design**, **offline event simulation**, **reconstruction**, **calibration** and **physics analysis** at STCF
  - A full chain of simulation + digitization + reconstruction + analysis has been established, driving physics feasibility studies at STCF!
  - **Machine learning explored** in simulation (e.g. EMC), reconstruction (tracking, PID…) and analysis…



OSCAR Core Software features

- ➔ W.H. Huang et al 2023 JINST 18 P03004
- ➔ More in T. Li's slides

- Underlying event loop control using **SNiPER** (adopted also by JUNO, LHAASO, nEXO, HERD)
- Event Data Model (EDM) based on **podio** (key4hep adopted by CEPC, FCC…)
- Detector description using **DD4hep**



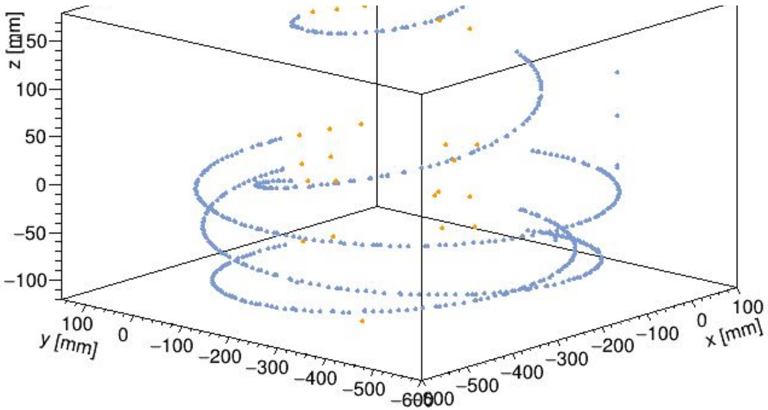
- Supports **multithreading**, **Machine Learning** and **heterogeneous computing**
- Supports **event display**, database, tests…
- Performance **monitoring**

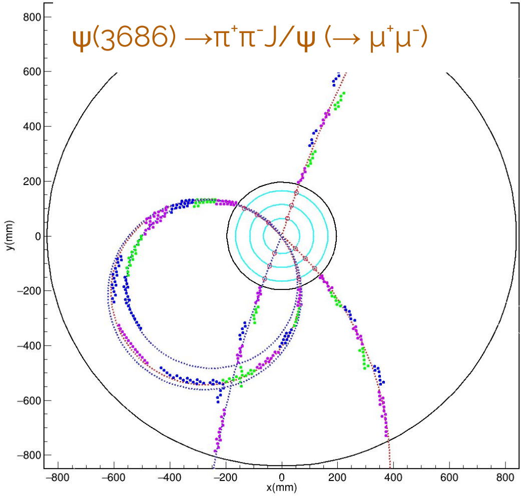
STCF tracking/γ performance

STCF tracking challenges

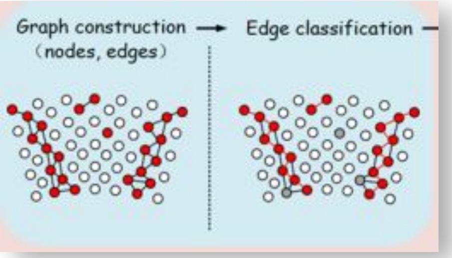
- Most physics processes have charged particles with p_T < 500 MeV/c
  - More material effects → worse resolution
  - Looping tracks with p_T < 130 MeV/c → fake/duplicate tracks
- High backgrounds and noise → worse efficiency and resolution
- Long-lived particles (non-trivial task in all HEP experiments!)

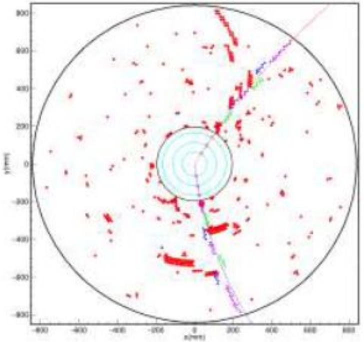
Simulated trajectory of single muon with p_T = 100 MeV, θ = 90˚

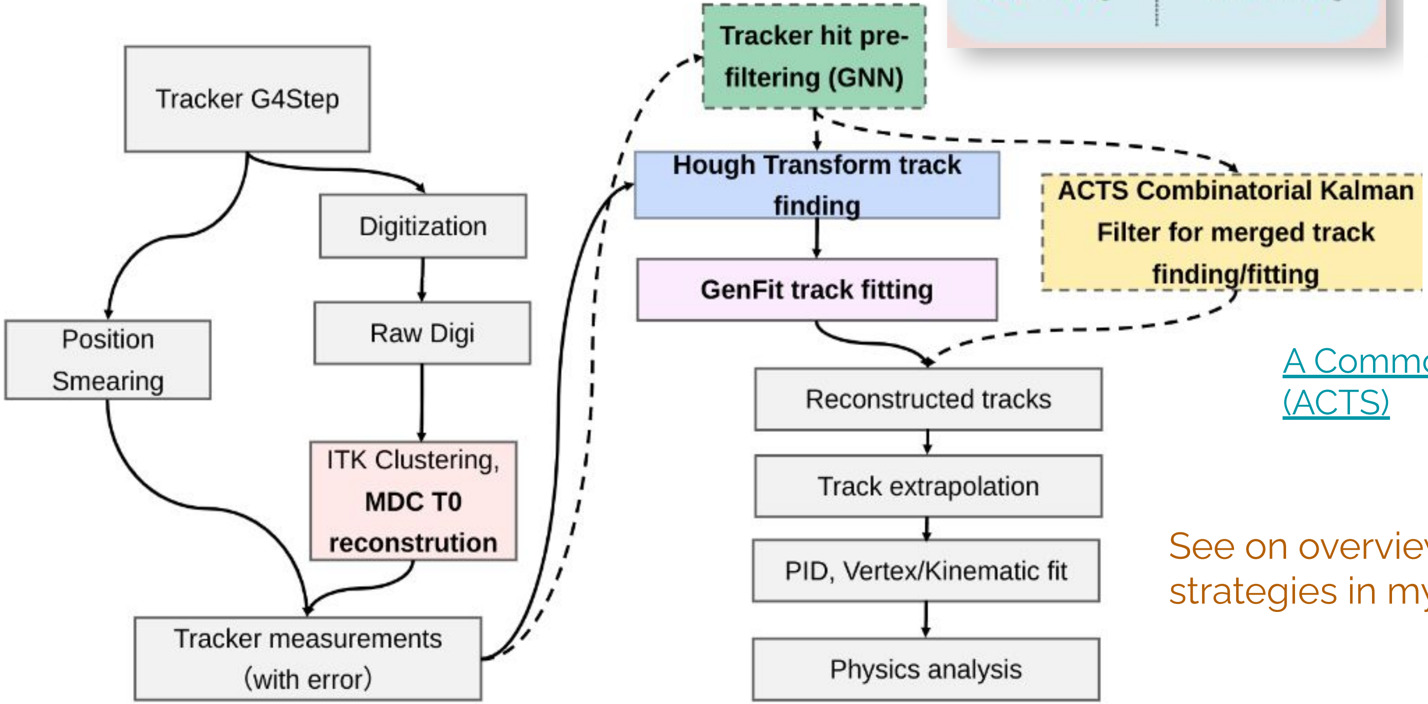




STCF tracking landscape







A Common Tracking Software (ACTS)

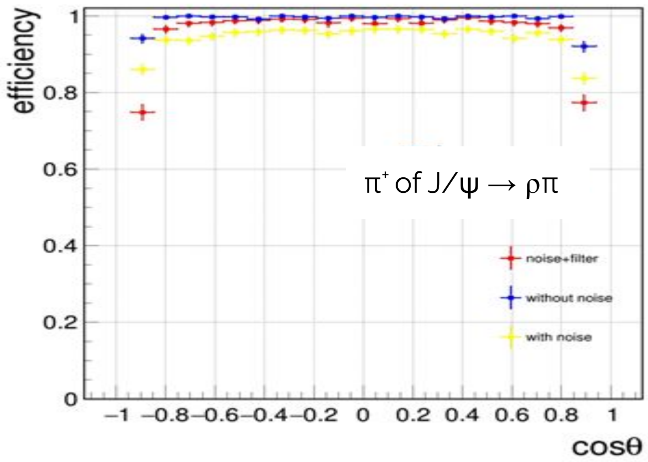
See on overview of HEP tracking strategies in my slides here

Dashed lines denote functionalities to be integrated into OSCAR

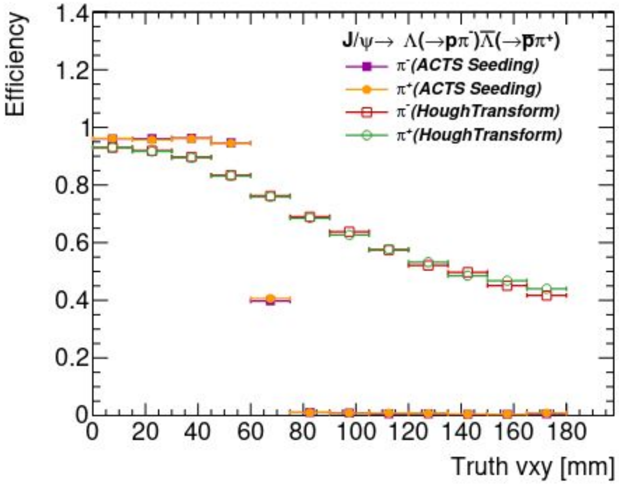
Tracking performance

More about STCF tracking in H. Zhou's slides

GNN + Hough Transform
W/ and W/O backgrounds



Hough Transform + ACTS
W/O backgrounds

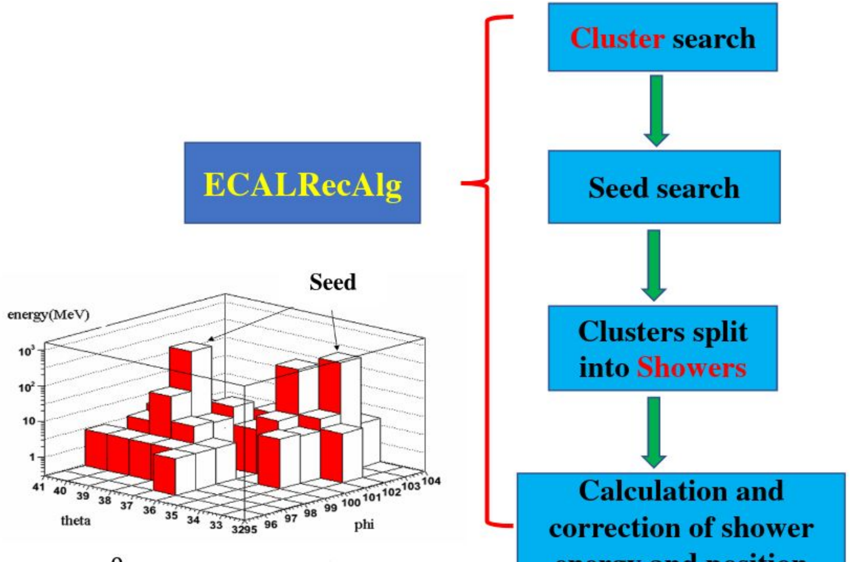


- Combination of global and local tracking strategies to achieve the maximum performance (Hough Transform, ACTS, GenFit, GNN)
  - Tracking efficiency is **above 95%** in central region for p_T > 100 MeV/c, even with backgrounds
  - 99% noise hits can be removed by GNN (more in X.Q. Jia's talk)

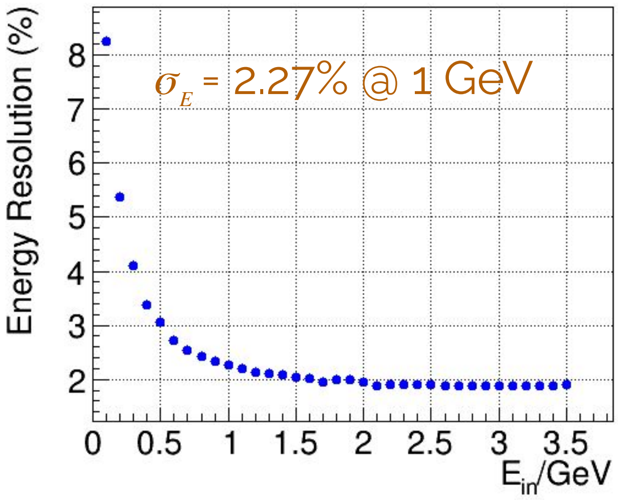
γ performance

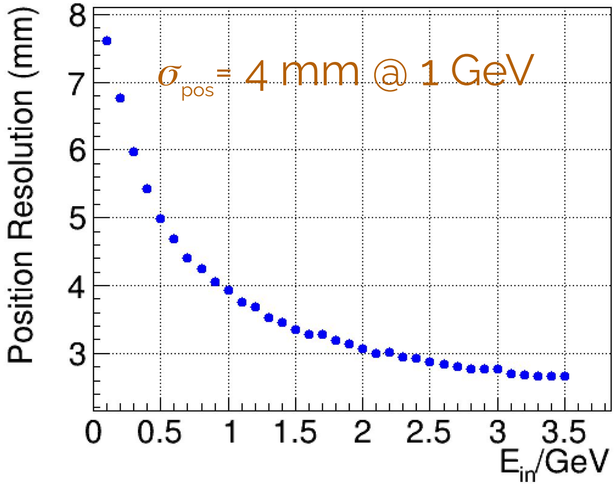
More about EMC in B. Wang's talk

- Performance with sophisticated simulation and reconstruction meets the CDR requirements



π^0 Cluster (two photons)

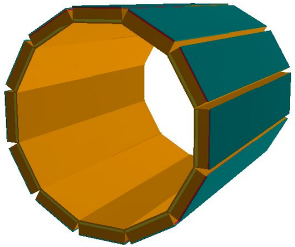




STCF PID performance

RICH reconstruction

More about RICH in [Q. Liu's talk]

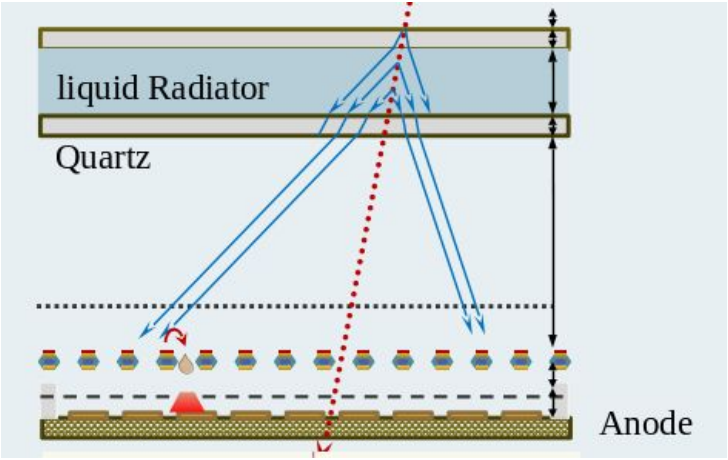


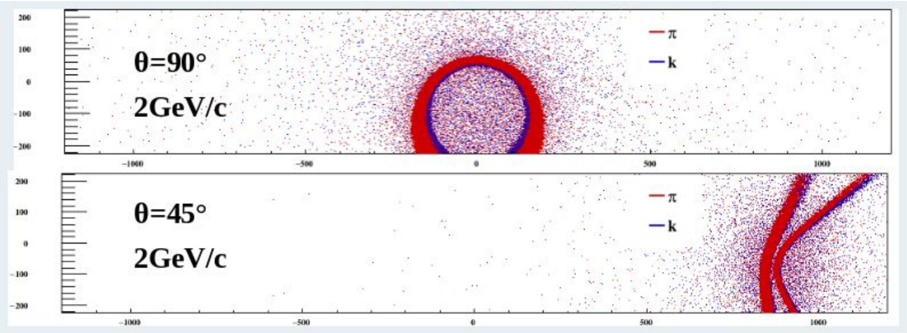
- Different distribution of number of cherenkov photons for different particles
- For each particle hypothesis h , log-likelihood defined as

$$\ln \mathcal{L}_h = \sum_{signal} \ln(PDF_{Ckv} + PDF_{Bkg})$$

simplified from Poisson

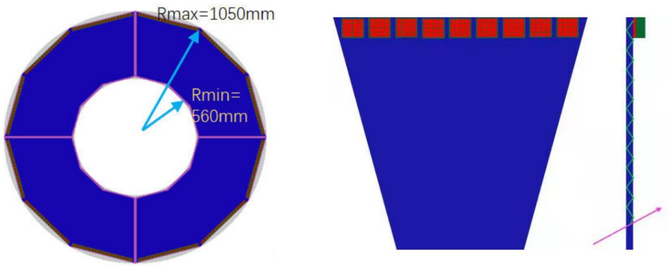
  - PDF_{Ckv} is calculated on-the-fly based on extrapolated track momentum/position
  - $DLL_{\pi K} = \ln \mathcal{L}_\pi - \ln \mathcal{L}_K$ provides particle ID



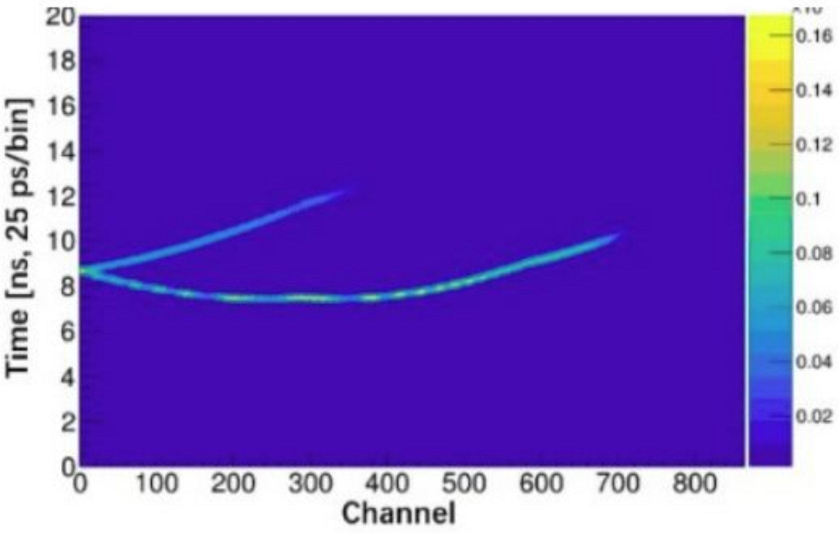


DTOF reconstruction

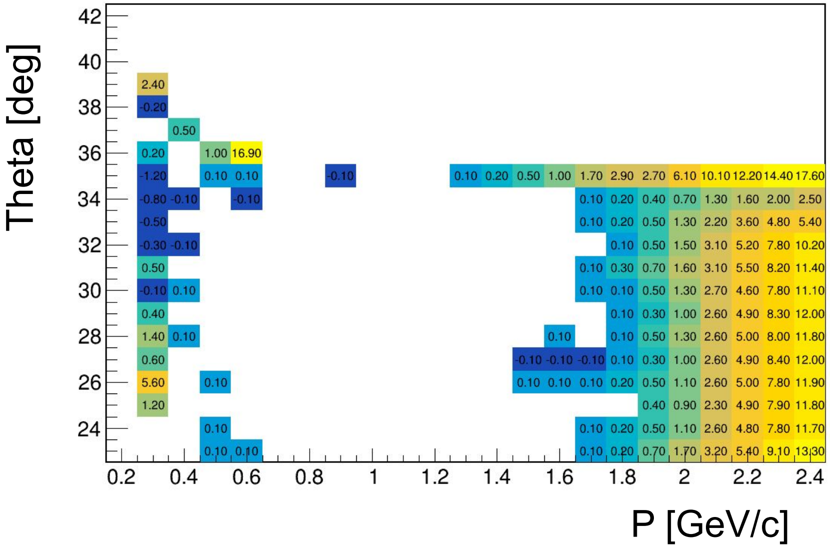
- Timing method (TOF): $$L_h = \prod_{i=1}^{N_{p.e.}} N_h S_h(TOF_i) + B$$
- Image method (Time vs Position) has better performance : $$L_h = \prod_{i=1}^{N_{p.e.}} \widetilde{N_h} S_h(ch_i, t_i) + B$$



More details in Y.T. Feng's slides



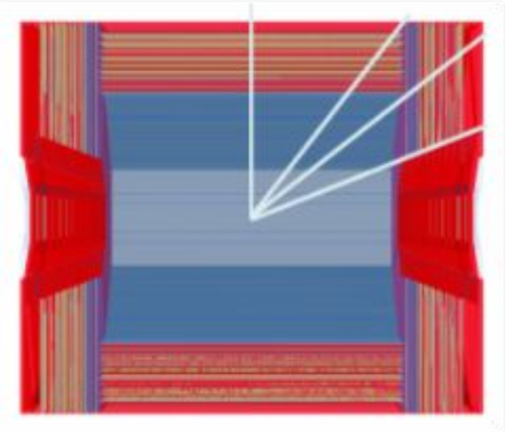
Difference of π PID eff. @2% K→ π mis-ID rate, W/ backgrounds, using the two methods



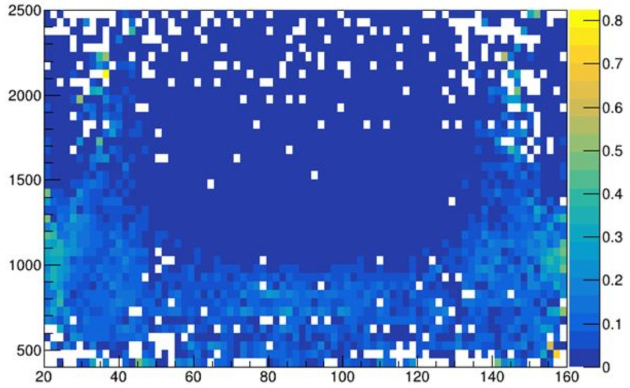
MUD reconstruction

More details in Y.L. Liu's slides

- Associate MUD hits/clusters to :
  - charged tracks based on extrapolated track position & momentum
  - n/K_L based on their ECAL showers
- Fine-binned BDT training using tracking + EMC + MUD reco features



PID eff. binned - PID eff. unbinned



TrackInfo	ClusterInfo
MomentumMag	SeedTheta
SeedTheta	SeedPhi
SeedPhi	DeltaTheta
DeltaTheta	DeltaPhi
DeltaPhi	LargestDistance
LargestDistance	Velocity
Velocity	MaxHitLayer
MaxHitLayer	MaxHit
MaxHit	LastLayer
LastLayer	PSHitCenter
HitAverageDistance	LowHitCenter
HitEntries	HitEntries
HitInRPC	HitInRPC
HitInPS	HitInPS
TrackType	PSHitNorm1
TrackQuality	PSHitNorm2
EcalEnergy	TrackType
EcalSeedEnergy	EcalEnergy
ESeed/E3x3	EcalSeedEnergy
ESeed/E5x5	ESeed/E3x3
E3x3/E5x5	ESeed/E5x5
EcalDev	E3x3/E5x5
	EcalDev

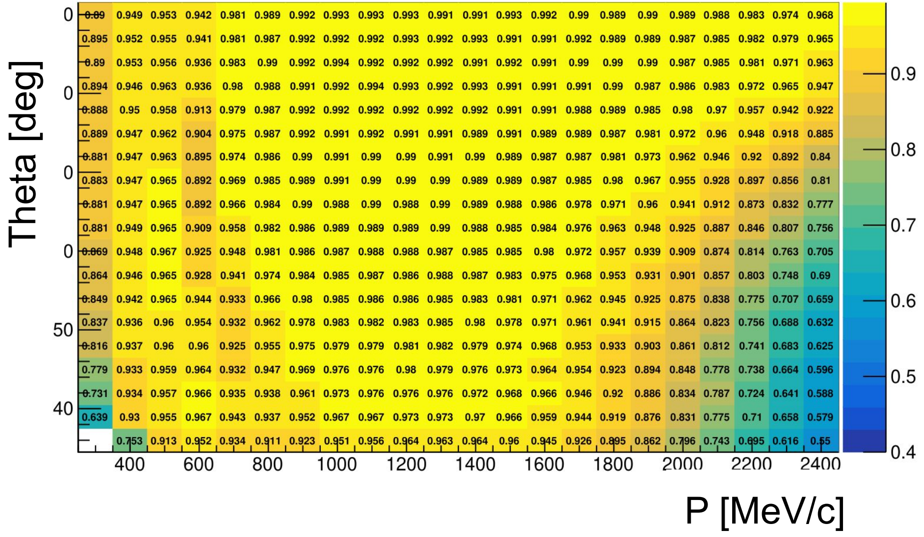
BDT features

π/K PID efficiency with traditional likelihood method

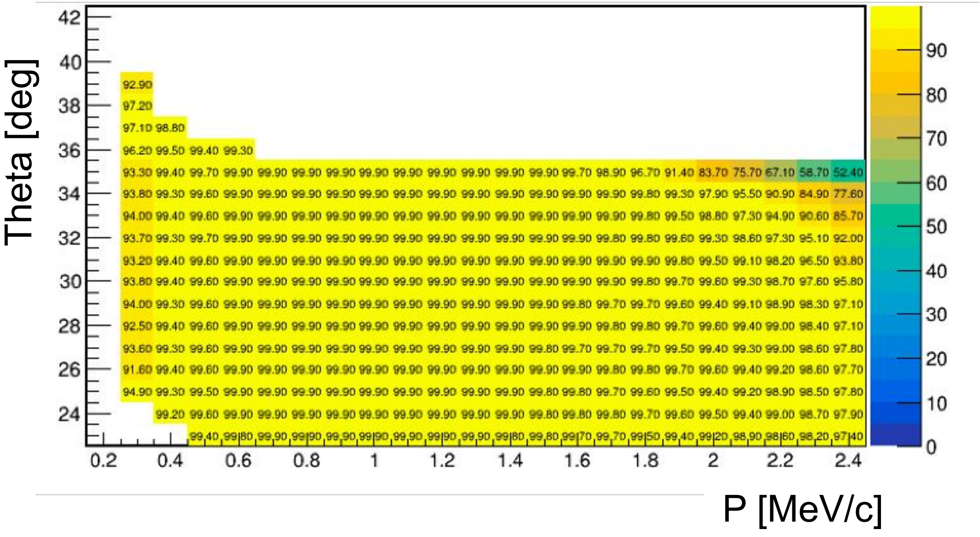
- 97% π/K PID efficiency for 0.7 GeV < p < 2 GeV with θ > 74˚
  - Room for improvement for 2 GeV with smaller θ
- > 97% π/K PID efficiency for 0.35 GeV < p < 2 GeV with 24˚ < θ < 35˚

π PID eff. @2% K→ π mis-ID rate, W/ backgrounds

From RICH



From DTOF (image method)

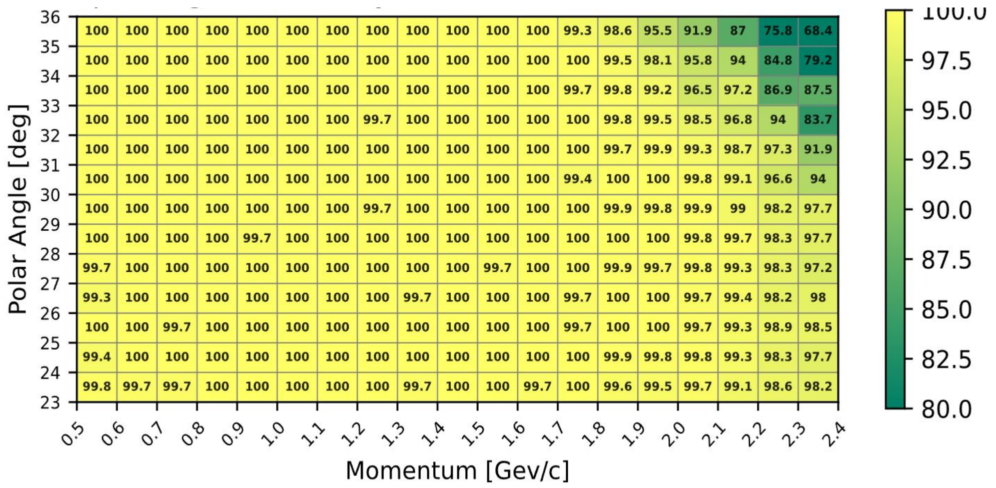


π/K PID efficiency with CNN

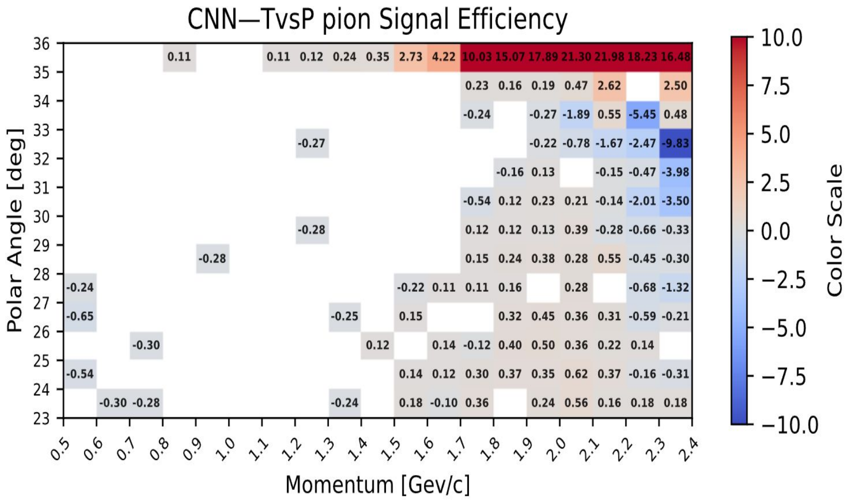
More details in Z. P. Yao's talk

- Using 2D map of TOF vs. channel of DTOF as inputs
  - Momentum and position extrapolated to DTOF fully connected layer
- CNN can further improve performance in certain regions

π PID efficiency @2% K→ π mis-ID rate, W/O backgrounds



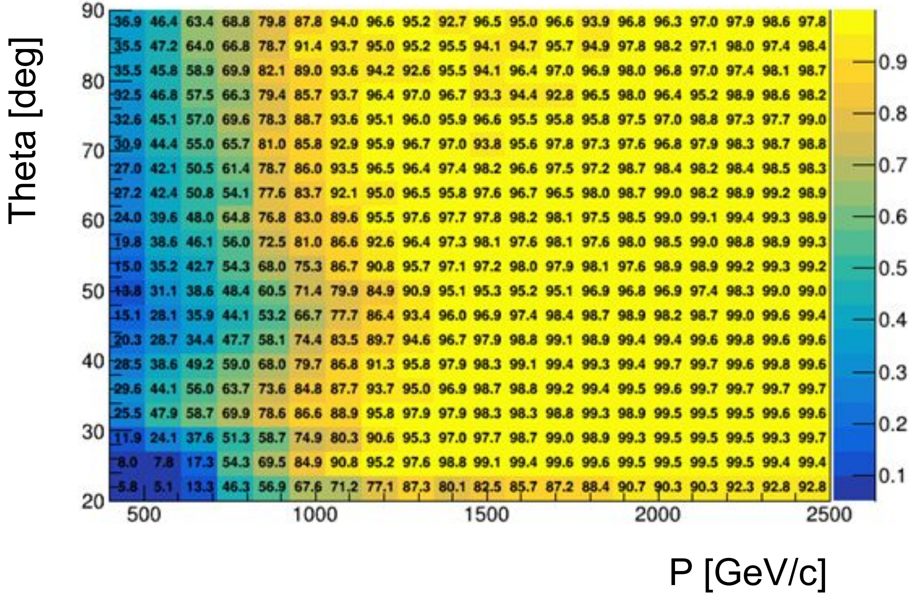
CNN has better performance at large theta and p



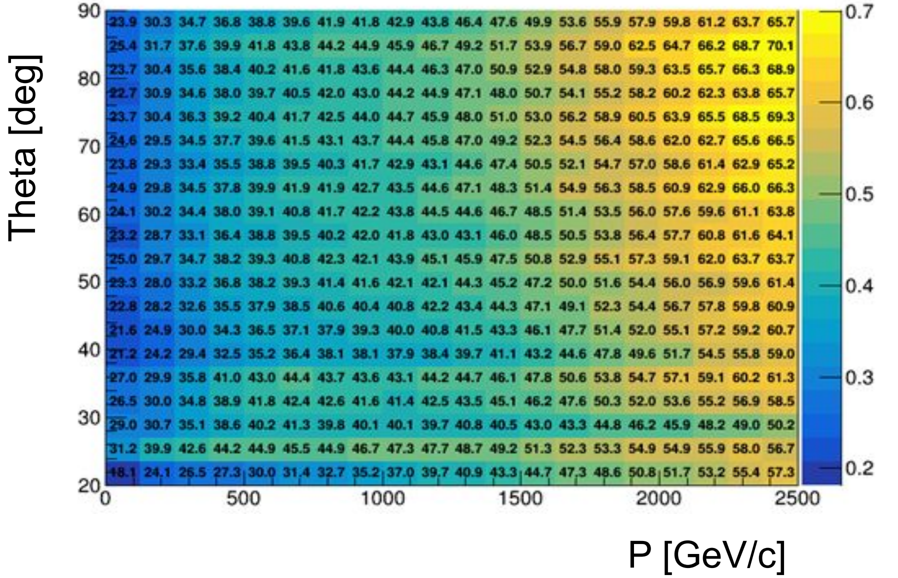
μ/π and (n, K_L)/γ PID efficiency with MUD

- >95% μ PID efficiency for p > 1.2 GeV (room for improvement for lower p)
- n/K_L PID performance is preliminary. Below 70% for now (>95% is expected)

μ PID efficiency @3.3% π→μ mis-ID rate, W backgrounds



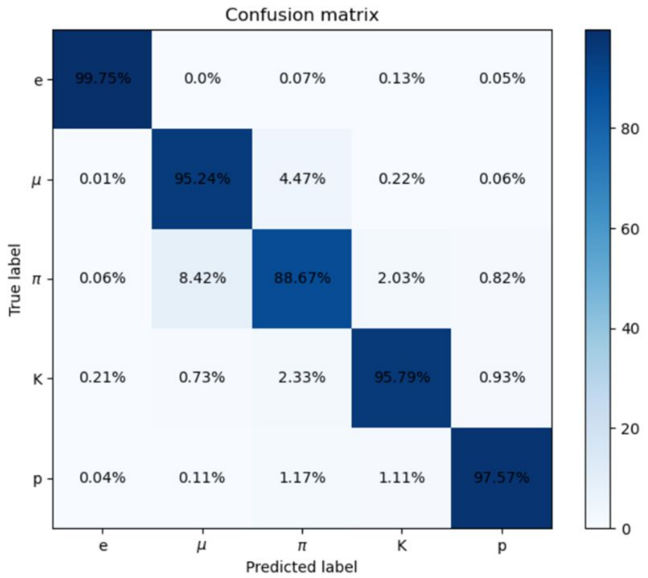
n/K_L PID efficiency @3.3% γ→n/K_L mis-ID rate, W/ backgrounds

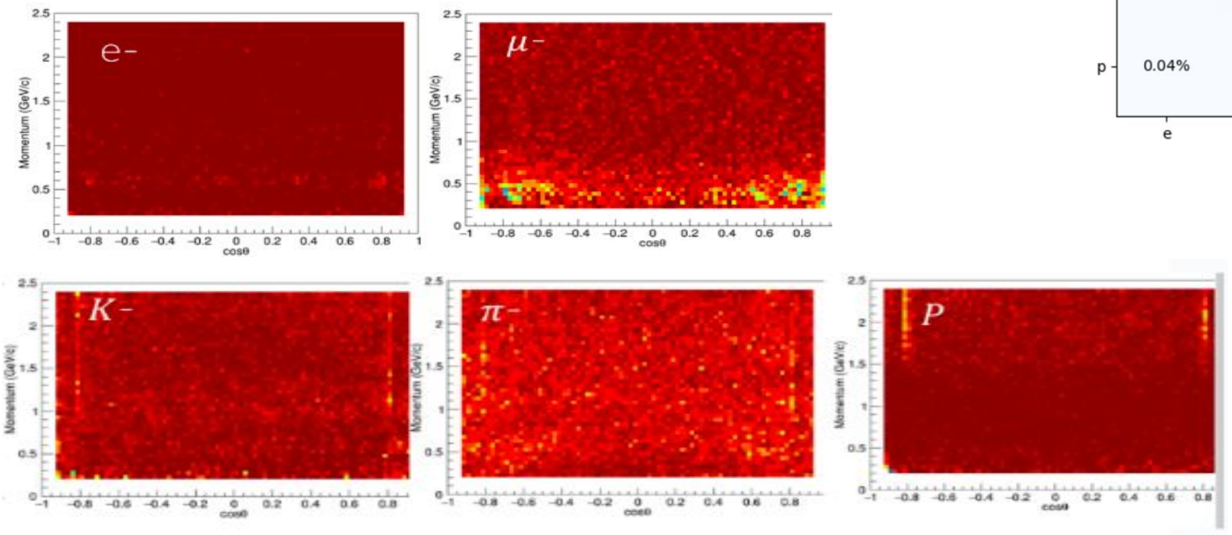


Global PID using BDT

More details in Y. C. Zhai's talk

- Using **45 features from Tracker/dEdx/RICH/DTOF/EMC/MUD**
- PID efficiency above 95% for other particles except π



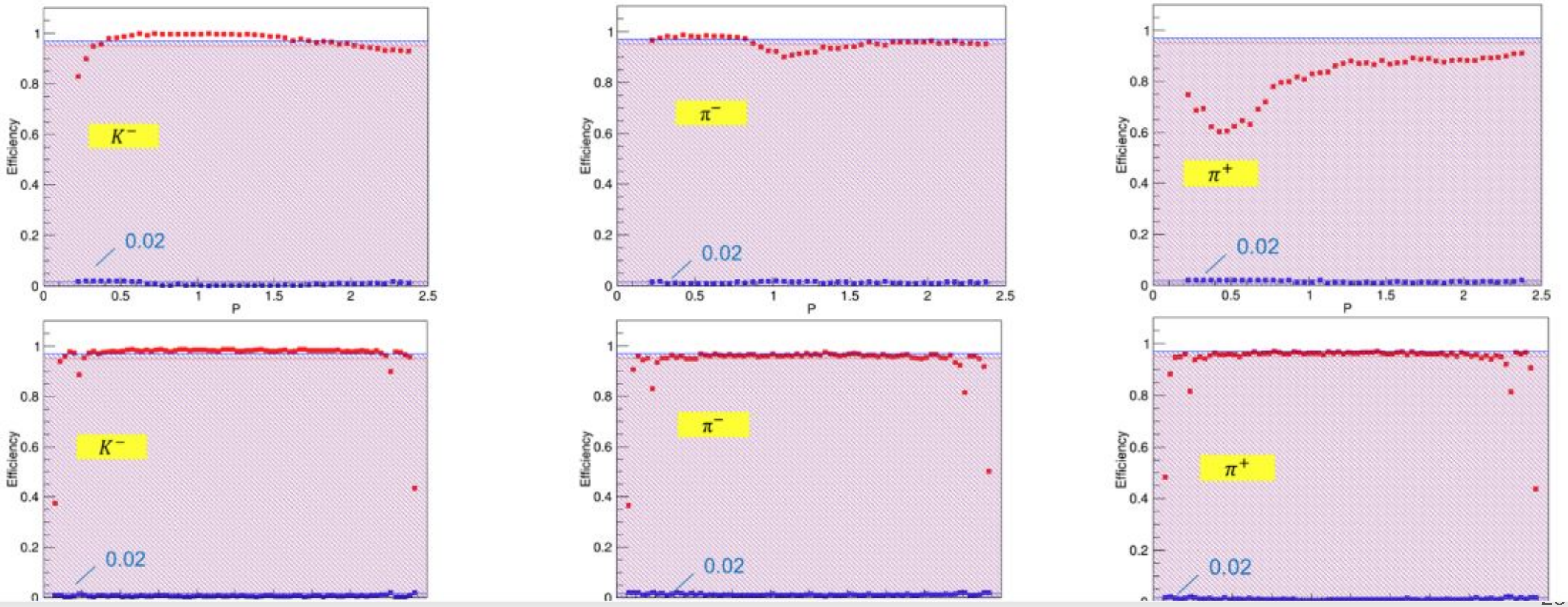


Backgrounds not considered yet

Global PID performance

More details in [Y. C. Zhai's talk]

- K/p (误鉴别<2%):
  - P < 0.4 Gev/c : PID 效率 > 80%
  - P 直到 2Gev/c : PID 效率 > 95%
- K/π (误鉴别<2%):
  - P < 0.8 Gev/c : PID 效率 ~ 97%
  - 动量大于 1.5Gev/c : PID 效率 ~ 95%
- μ/π (误鉴别<2%):
  - PID 效率 : > 60%



Summary

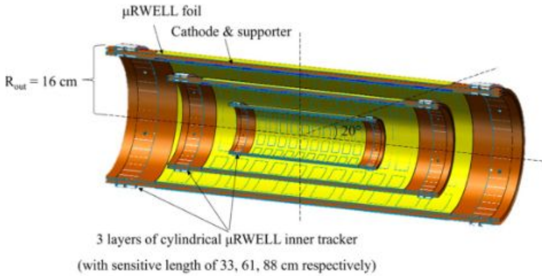
Summary

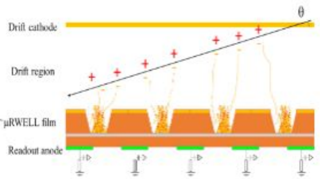
- The STCF offline software allows for a **full event processing chain** supporting STCF detector design, optimization and physics simulation
- **Good tracking, photon and PID performance already achieved** based on both **traditional** and **innovative ML** techniques
  - Still room for improvement in certain phase space region
- STCF **physics simulation studies in a realistic scenario** has started
- Currently, in full swing for both **physics and CPU optimization** to facilitate further **detector optimization**
  - Combination of different techniques/algorithms is the key for improvement

backup

From J.B. Liu's slides

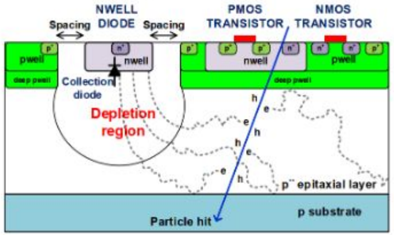
Gaseous option: MPGD



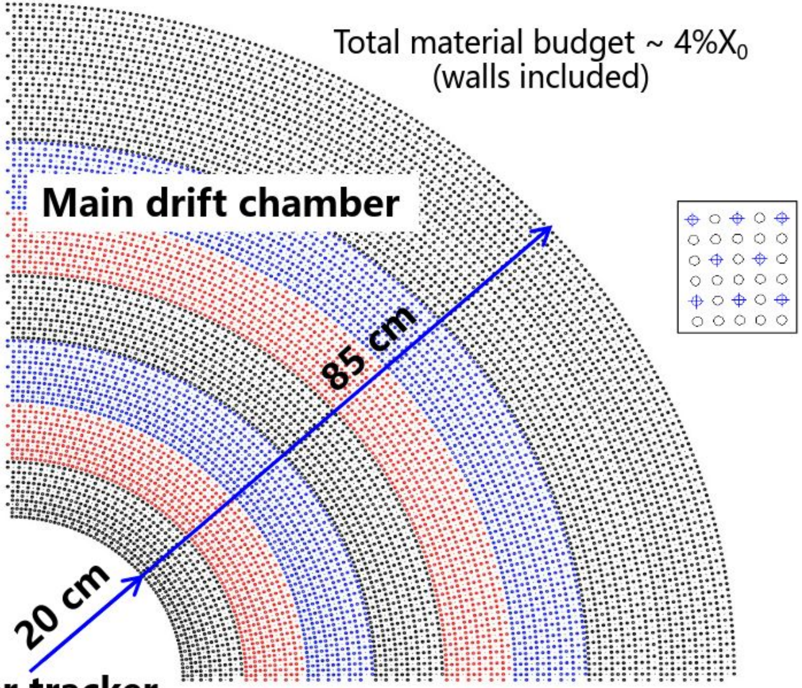


Material budget ~ 0.3%X_0/layer

Silicon option: CMOS MAPS



单片有源像素探测器



Inner tracker

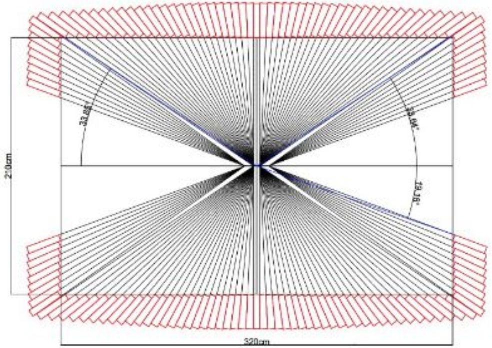
Inner-outer separate designs to accommodate different levels of radiation background

From J.B. Liu's slides

- **Barrel PID: A RICH detector using MPGD (THGEM with CsI + MM) for photon detection**

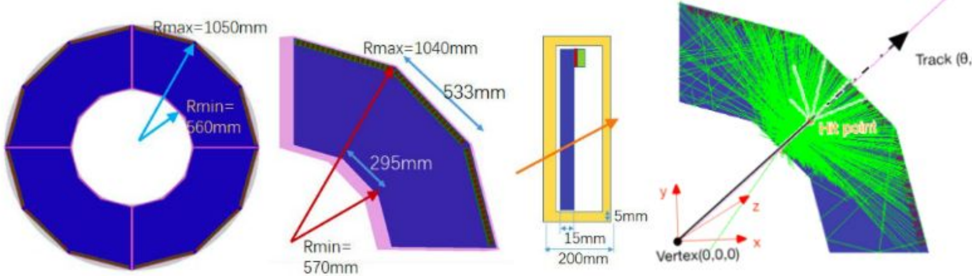
Material budget < $0.3X_0$

- **EMC: A pure-CsI crystal calorimeter**

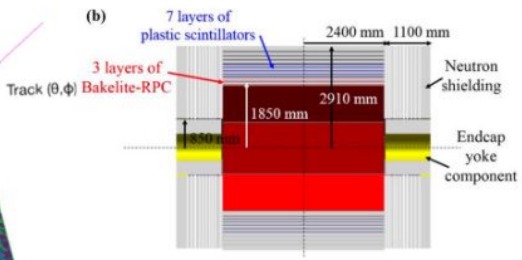


  - Crystal size
    28cm (15X0)
    5×5cm2
  - ~ 8670 crystals
  - 4 large area APDs (1×1cm2) to enhance light yield

- **Endcap PID: A DIRC-like high-resolution TOF detector (DTOF)**



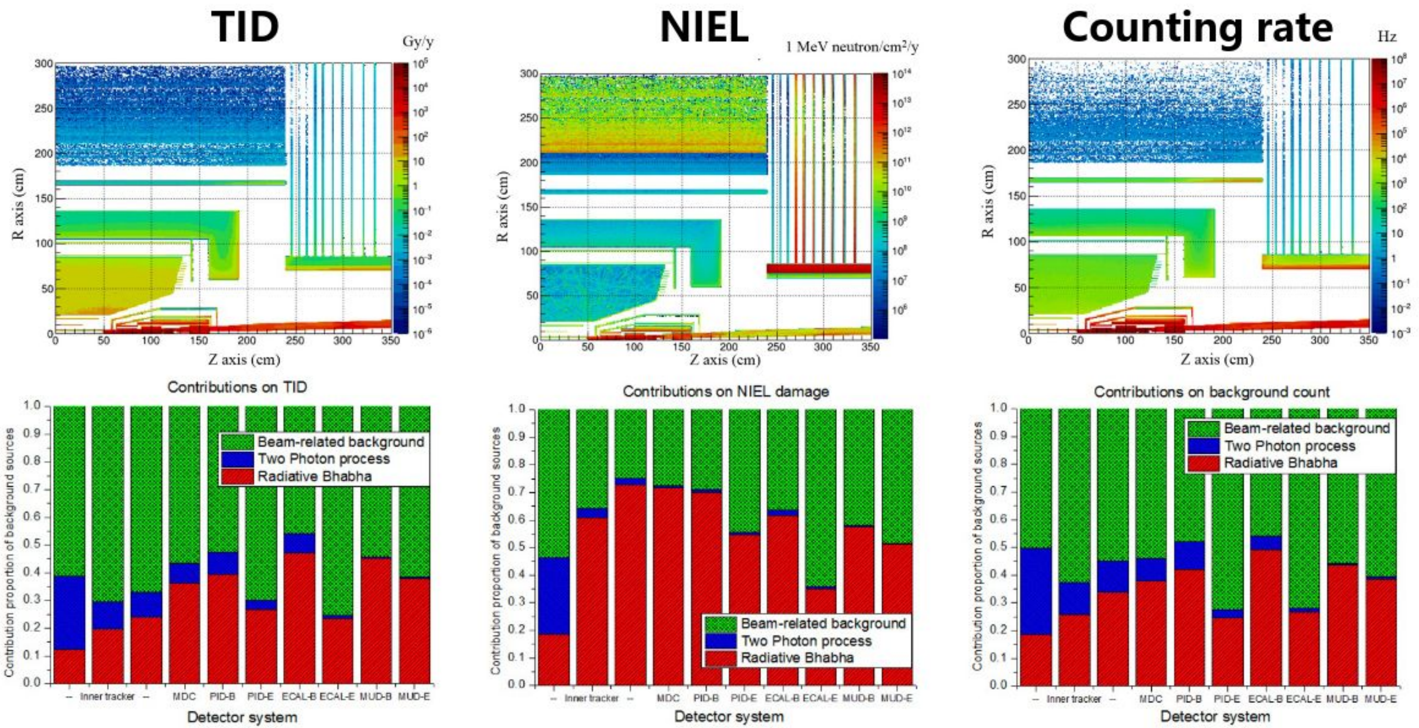
- **MUD: A RPC-scintillator hybrid detector**



Parameter	Baseline design
R_{in} [cm]	185
R_{out} [cm]	291
R_e [cm]	85
L_{Barrel} [cm]	480
T_{Endcap} [cm]	107
Segmentation in ϕ	8
Number of detector layers	10
Iron yoke thickness [cm]	4/4/4.5/4.5/6/6/6/8/8 cm
(λ=16.77 cm)	Total: 51 cm, 3.04λ
Solid angle	79.2%×4π in barrel 14.8%×4π in endcap 94%×4π in total
Total area [m^2]	Barrel ~717 Endcap ~520 Total ~1237

Beam-induced backgrounds

[Fig. from J.B. Liu's slides](#)



Inner most detector layer: ~3.5 kGy/y, ~2×10^{11} 1MeV n-eq/cm^2/y, ~1 MHz/cm^2

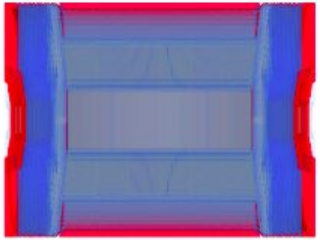
The major challenge is to maintain or even enhance the state of the art performance of τ-**c** detectors in much harsher experimental conditions.

MUD digitization

✧MUD：

✧旧版几何：模拟+数字化框架已经搭建，数字化参数需要更新

✧新版几何：模拟算法已经完成，数字化算法开发中



MUD新版几何

* MC 样本

- Oscar 版本 : 2.5.0
- GeV
- Exclusive MC : $J/\psi \to \rho\pi$

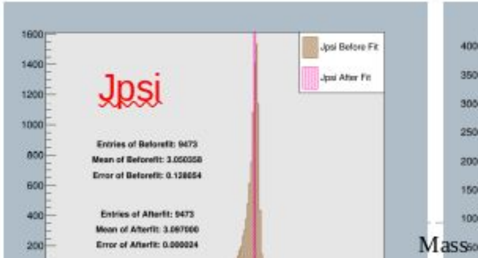
* 事例选择

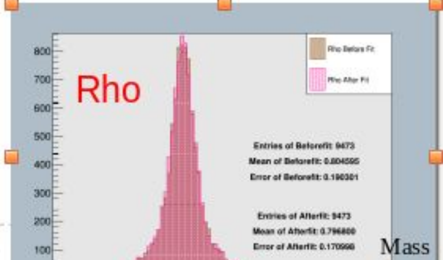
- 带电径迹的选择
  - $N_{\text{good charge track}} = 2$
  - Total charge=0
- 中性径迹的判断
  - 桶部光子 Eγ>25MeV (|cosθ|<0.8)
  - 端盖光子 Eγ>50MeV (0.86<|cosθ|<0.92)
  - $N_{\text{good neutral track}} > 2$
- Global PID
  - 判选条件
- 顶点拟合 & 运动学拟合
  - $\chi^2 > 200$

GlobalPID 软件包可以提供良好的鉴别能力

* 效率检查（初步结果）

选择条件	事例数目	效率
事例数目	24000	
带电径迹	14217	59.24%
中性径迹	14214	99.98%
带正电的径迹鉴别为 pion	12477	88.70%
带负电的径迹鉴别为 pion	12411	88.10%
两条径迹都鉴别为 pion	10869	78.02%
顶点拟合 & 运动学拟合	9473	66.65%





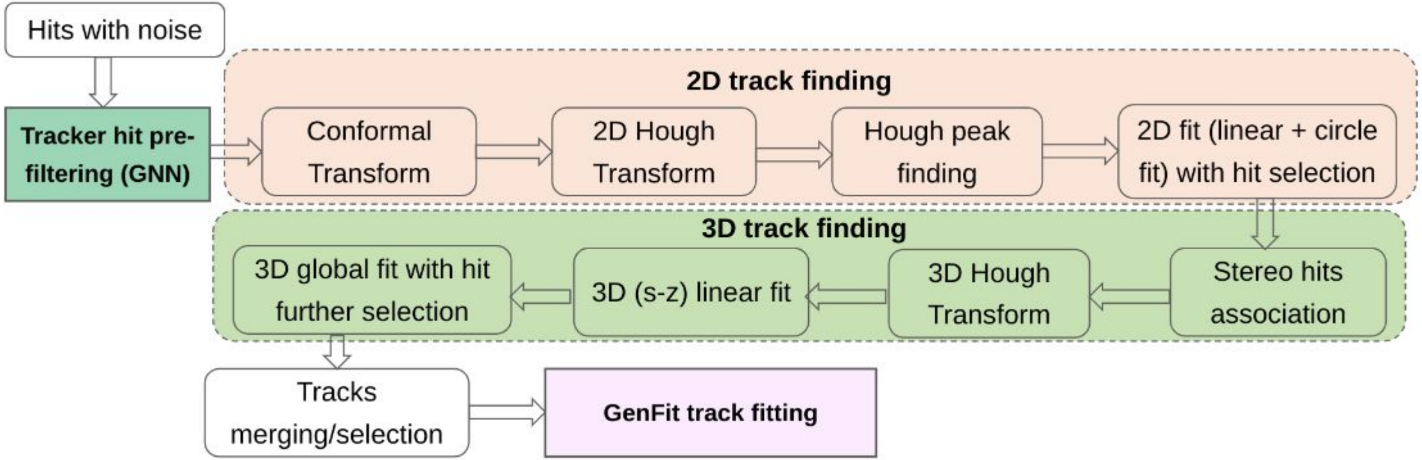
子探测器	每事例耗时（ms）	占比
ITK	67	16.6%
MDC	24	6.0%
RICH	46	11.4%
DTOF	130	32.3%
ECAL	55	13.6%
MUD	136	33.7%
ALL	403	100%

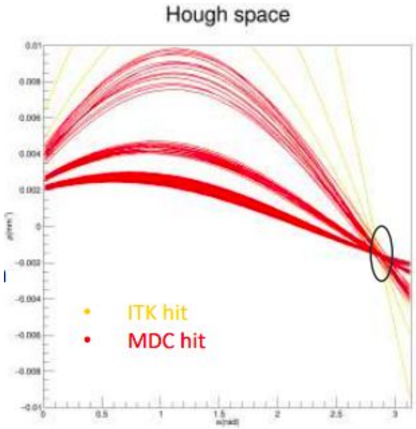
- 探测器模拟，相互作用模拟部分耗时
- 物理过程： e+e- --> j/psi $\mu^+\mu^-$ --> $\pi^+\pi\mu^+\mu^-$
- 单线程
- 关闭光吸收电离
- 模拟总耗时： ~1s/event

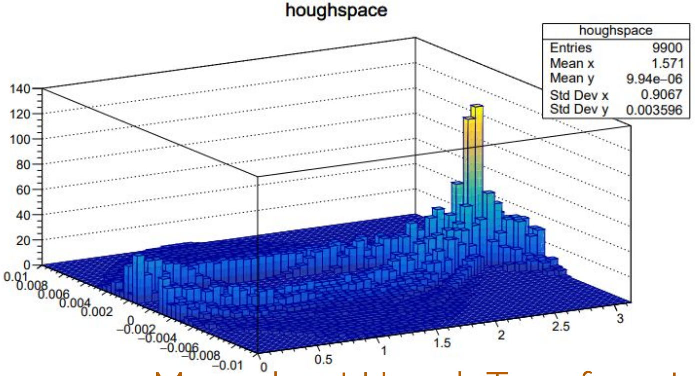
算法	每事例耗时 (ms)	标准差 (ms)	占比
ITK 数字化（简化）	0.48	0.95	0.00%
MDC 数字化（简化）	1.05	1.35	0.01%
ECAL 数字化	186.57	100.18	1.92%
RICH 数字化（简化）	3.28	0.55	0.03%
DTOF 数字化（简化）	0.51	3.87	0.01%
MUD 数字化（简化）	2.28	0.58	0.02%
ITK Cluster 重建	0.04	0.04	0.00%
MDC Hit 重建	0.44	0.13	0.00%
Hough Finder	404.52	306.1	4.16%
径迹拟合（to debug）	6657.52	14872.14	68.60%
dE/dx 重建	2.05	0.51	0.02%
径迹外推	206.98	38.61	2.13%
ECAL 重建	0.32	0.89	0.00%
RICH 重建 （to optimize）	2187.43	761.16	22.54%
DTOF 重建	38.41	62.97	0.40%
MUD 重建	2.80	14.90	0.03%
事例组装器	0.33	0.29	0.00%
事例取样器	0.64	2.30	0.01%
其它	9.51	/	0.10%
总和	9704.71	14937.79	100%

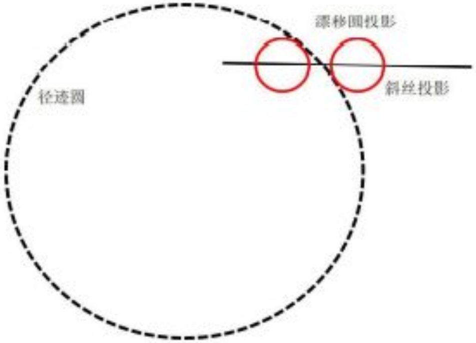
- 数字化 + 重建耗时
- 物理过程： e+e- --> j/psi $\mu^+\mu^-$ --> $\pi^+\pi^-\mu^+\mu^-$
- 总耗时： 9.7s/event

Track finding with Hough Transform









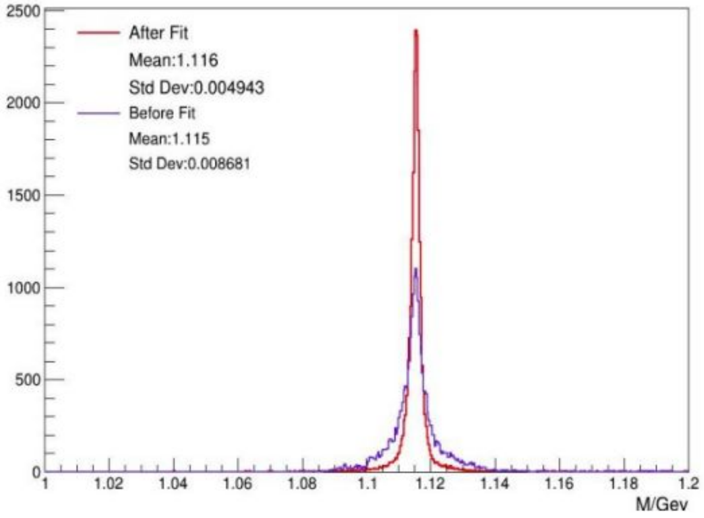
More about Hough Transform tracking in H. Zhou's talk

Vertex fit

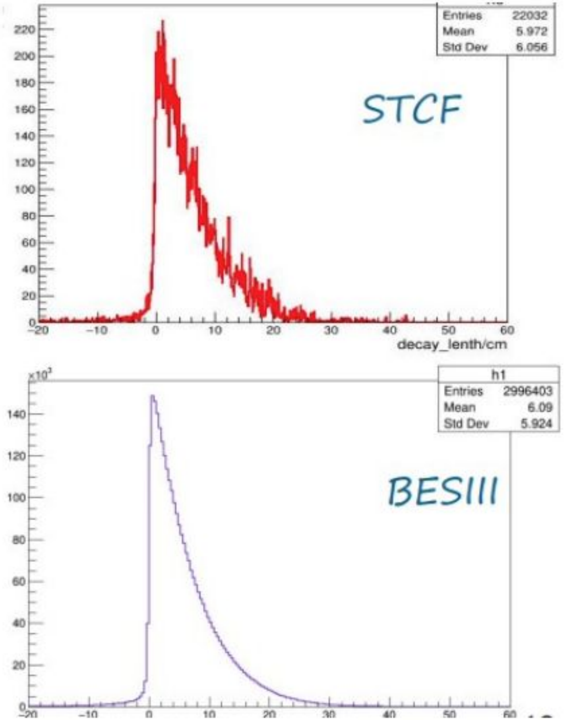
- Vertex fit transcribed from BESIII has been validated

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

Invariant mass of Λ



Decay length

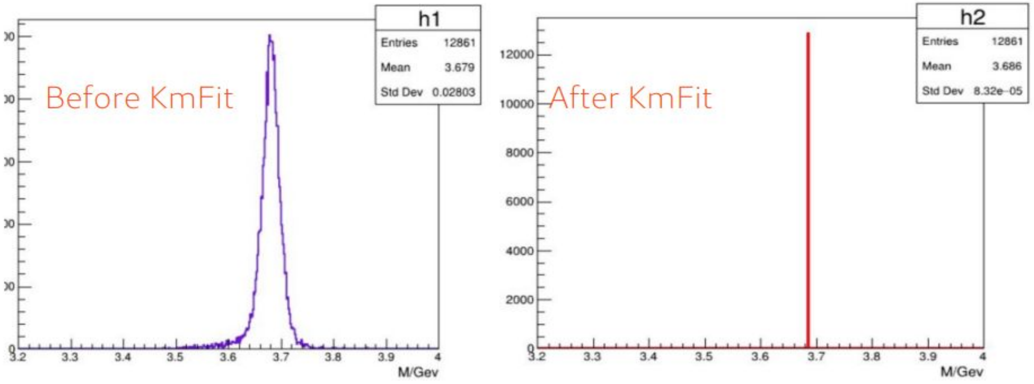


Kinematic fit

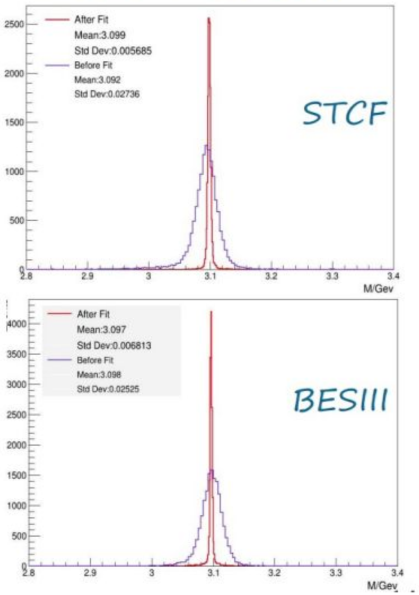
- Kinematic fit transcribed from BESIII has been validated
- GlobalVertexFit which combines vertex and kinematic fit has been transcribed from Belle-II recently. Validation is in progress.

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

Invariant mass of ψ(2S)



Invariant mass of J/ψ



Status of the full event processing chain

[Finalised] Finalised   [Working] Working, under optimization   [Developing] Developing or not started

		Simulation	Digitization	Reconstruction: On top of Digitization	Reconstruction: On top of Bkgs + Digitization	Analysis (W/O bkgs): Global PID Charged	Analysis (W/O bkgs): Global PID Neutral	Analysis (W/O bkgs): Vertex/Kinematic Fit
ITK		[Finalised]	[Finalised]	[Working]	[Working]	1) Single particles [Finalised]; 2) Physics Processes [Working]	1) Single particles [Working]; 2) Physics Processes [Developing]	[Finalised]
MDC		[Finalised]	[Finalised]	[Working]	[Working]			
RICH		[Finalised]	[Finalised]	[Working]	[Working]			
DTOF		[Finalised]	[Finalised]	[Finalised]	[Finalised]			
ECAL		[Finalised]	[Finalised]	[Finalised]	[Finalised]			
MUD	Pre-geometry	[Finalised]	[Working]	[Working]	[Working]			
	New geometry	[Finalised]	[Developing]	[Working]	[Working]			