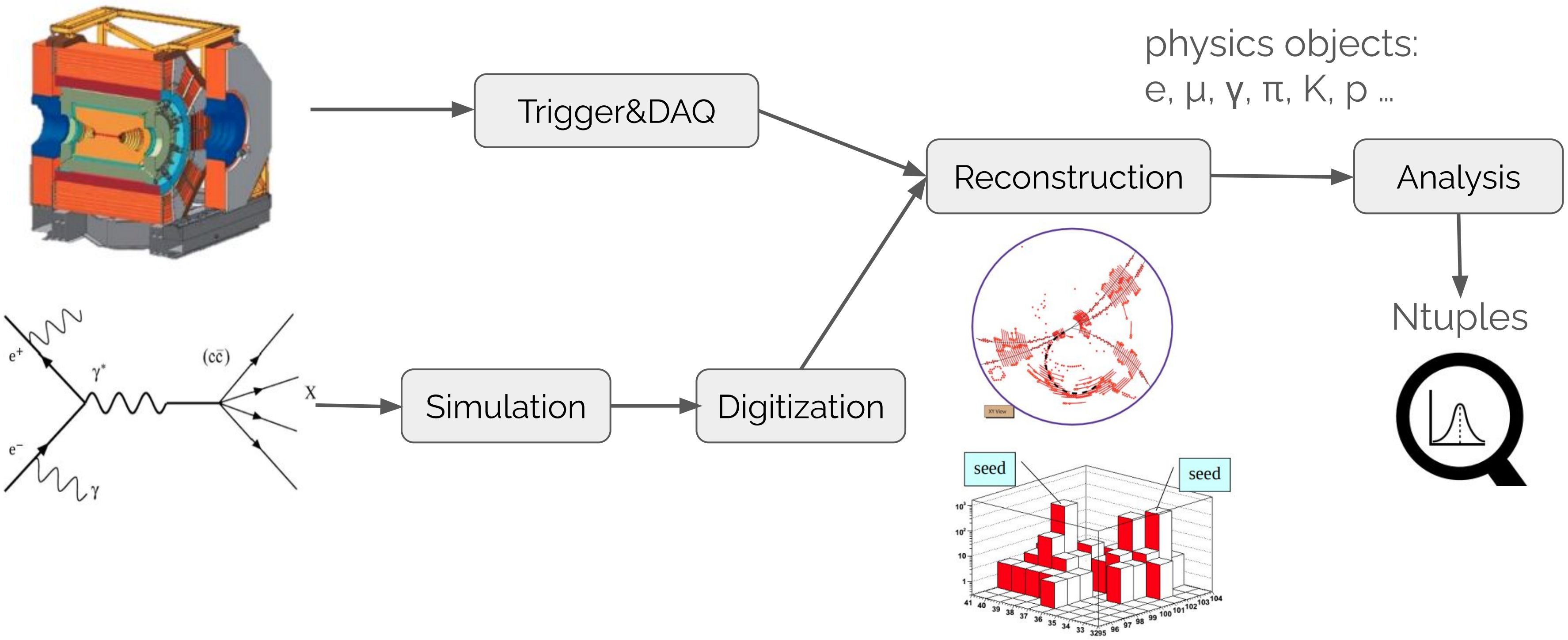
Status of STCF Offline Software

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

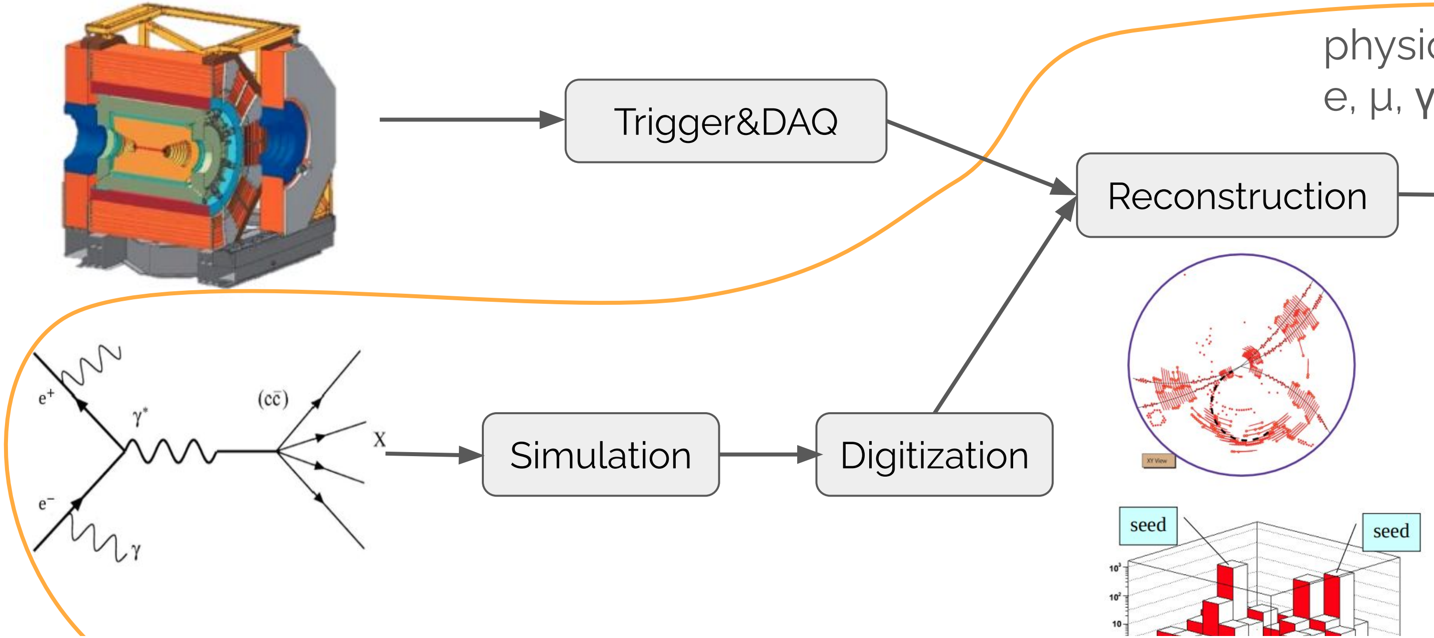
1. Shandong University
2. Zhengzhou University

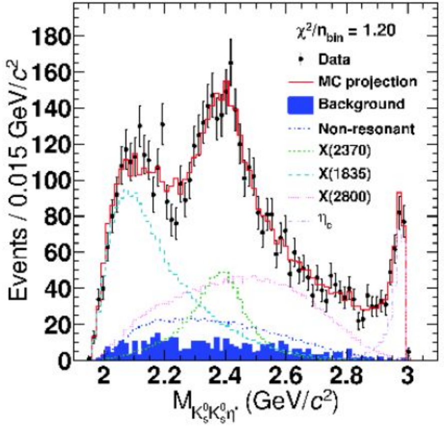
2024年超级陶粲装置研讨会, 2024年7月8日, 兰州

The crucial role of HEP offline software



HEP offline software is about physics discoveries



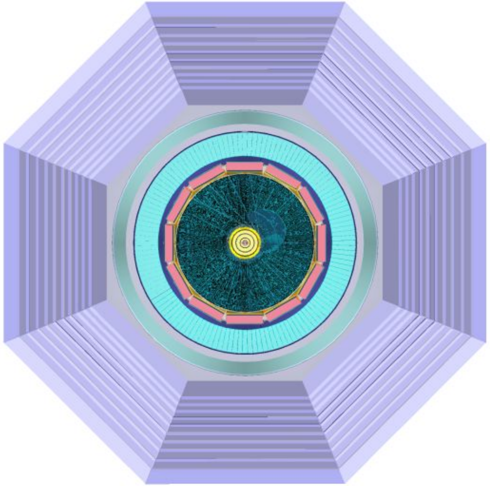


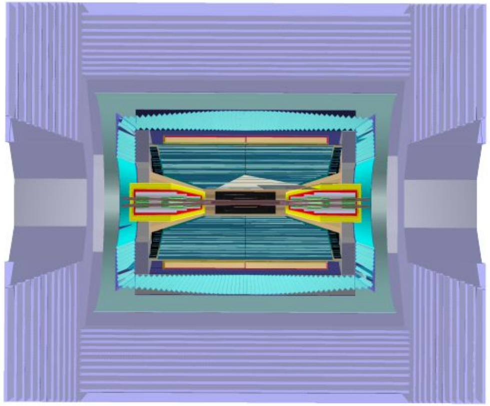
PhysRevLett.132.181901

The offline software is the "**converter**" from detector data to physics data to make physics discoveries (basically all physics analysis) possible!

HEP offline software is about detectors

- To **guide** the design of often very sophisticated detectors
- To **exploit** (i.e. not to spoil) the maximum performance of the detectors
- To **detect** possible defects, malfunction, aging … of the detectors





➔ See A. Guo's talk about EicC detector optimization
➔ See L. Wu' talk about BESIII detector alignment

Geometry of the STCF detector

Physics requirements of STCF

Requirements of offline software at STCF

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

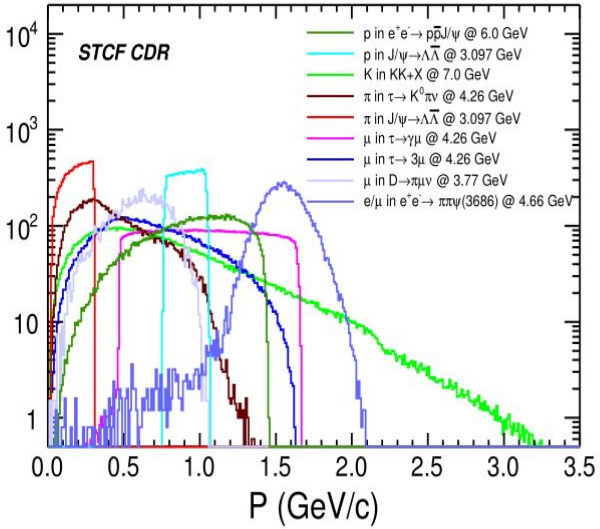
→ i.e. we need **reconstruct** the tracks and photon **with good efficiency and resolution**, and **identify** them **at high accuracy**, with **good speed**

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	$\sigma_T = \frac{300}{\sqrt{p^3(\text{GeV}^3)}}$ ps

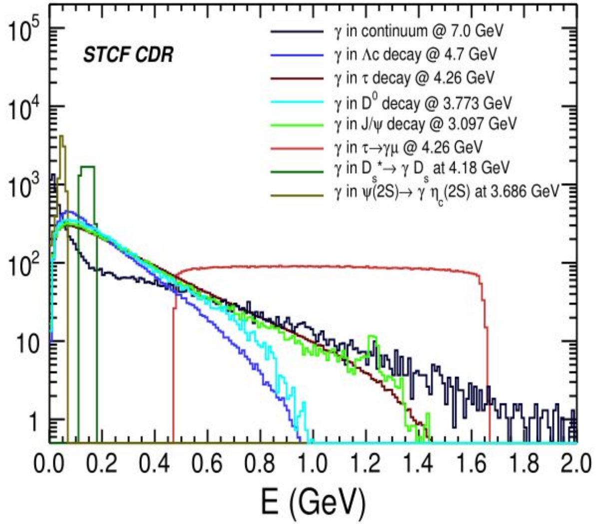
Particles at STCF

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

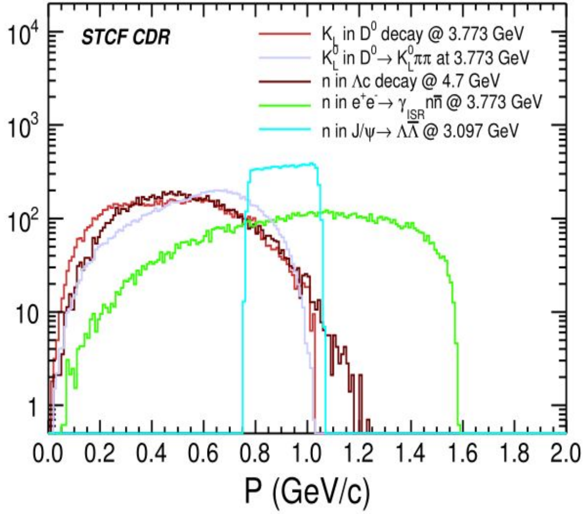
Charged particle momentum



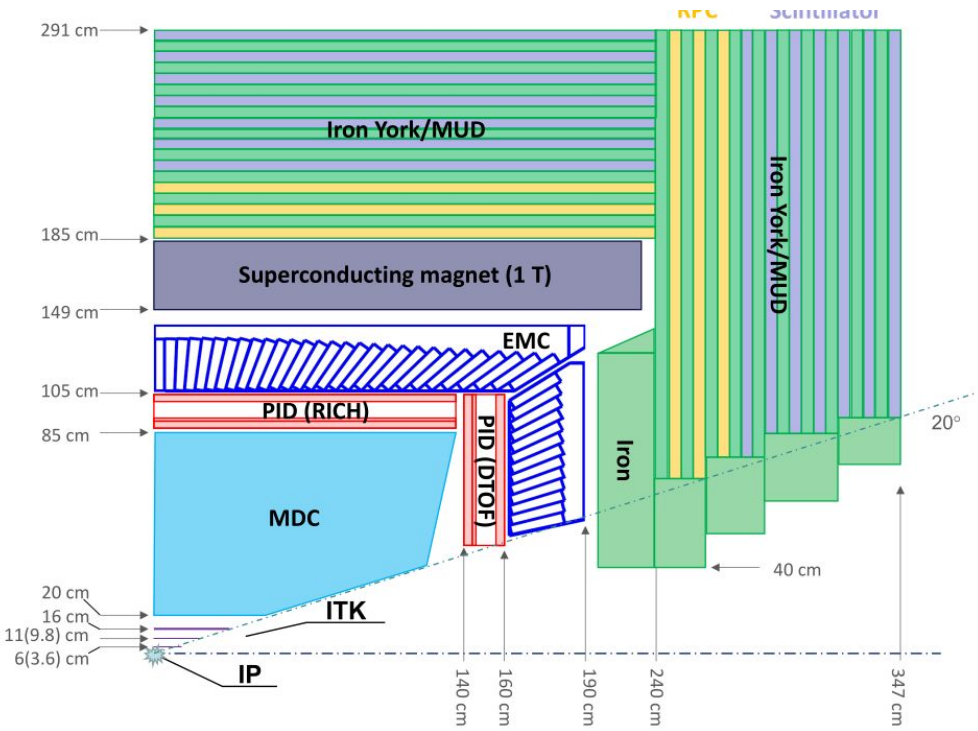
Photon energy



K_L, n momentum



The detector and performance requirements



ITK

- Material < 0.01 X0, σ_{xy} < 100 um

MDC

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

RICH & **DTOF**

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

EMC

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

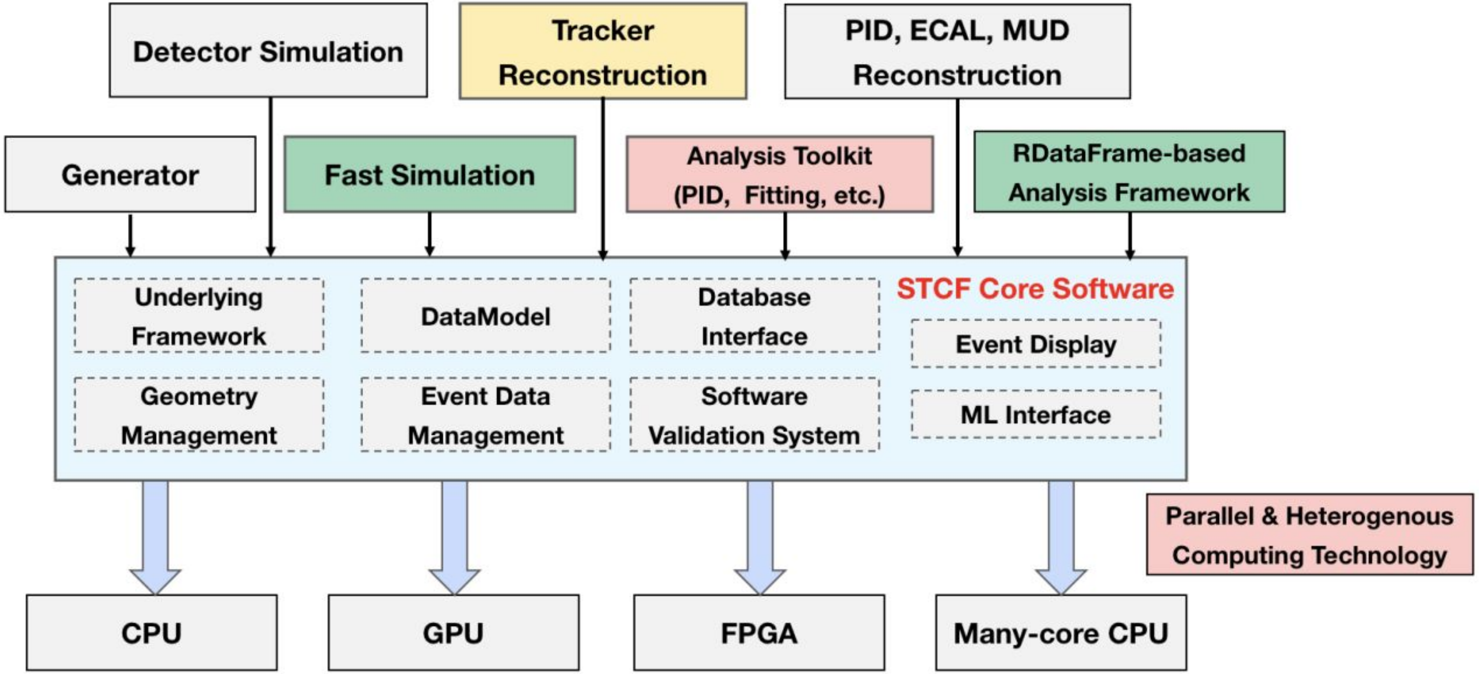
MUD

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

STCF offline software

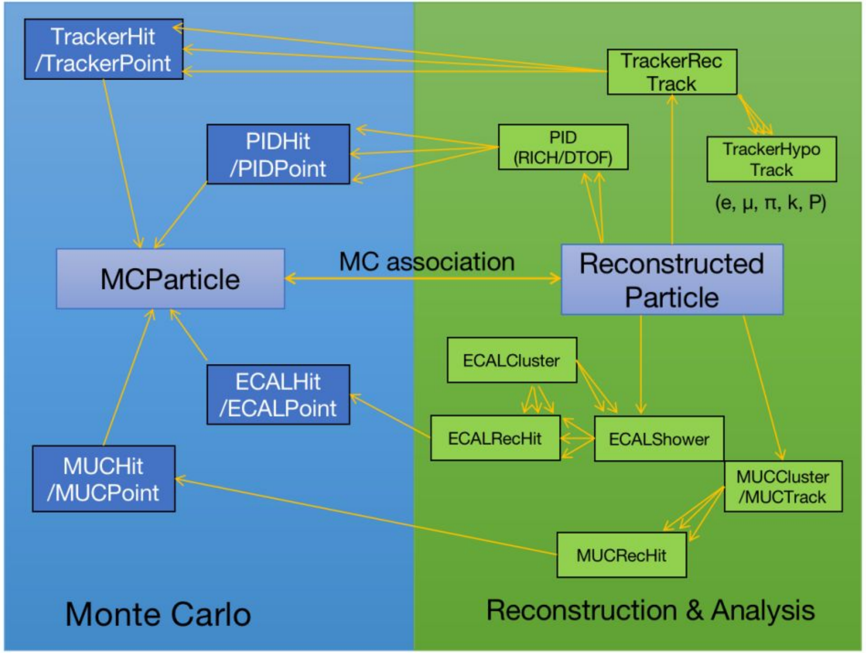
The Offline Software of Super Tau-Charm Facility (OSCAR)

- Provides common functionalities for **detector design**, **offline event simulation**, **reconstruction**, **calibration** and **physics analysis** at STCF



OSCAR Core Software features

- Underlying event loop control using **SNiPER** (adopted also by JUNO, LHAASO, nEXO, HERD)
- Event Data Model (EDM) based on **podio** (key4hep adopted by CEPC, ILC, FCC…)
- Detector description using **DD4hep**
- Supports **multithreading**, **Machine Learning** and **heterogeneous computing**
- Supports **event display**, database, tests…

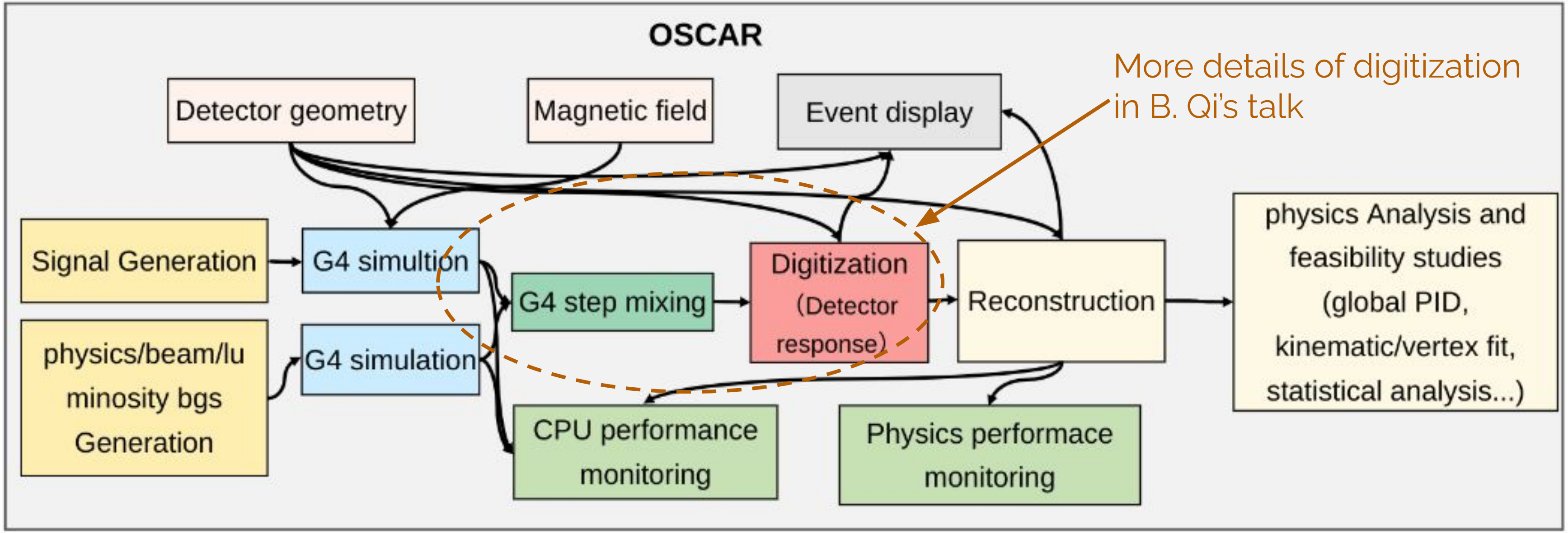


More details about core software in T. Li's talk and event display in Q. Zhang's talk

Event processing workflow with OSCAR

- A full chain of simulation + digitization + reconstruction + analysis has been established
- ML techniques exploited in simulation, reconstruction and analysis

My talk is non-exhaustive and focus more on performance

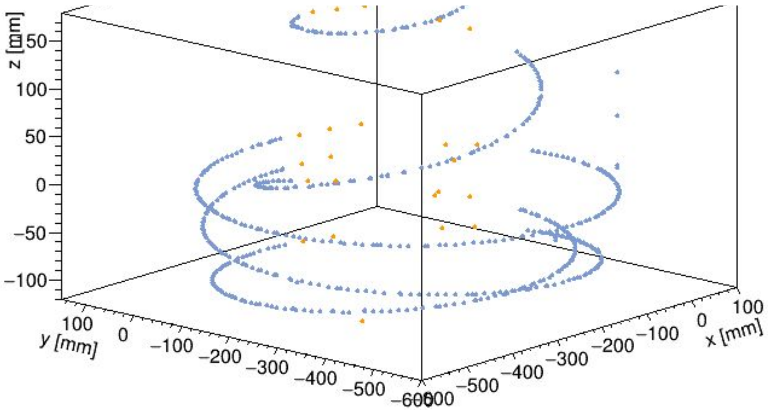


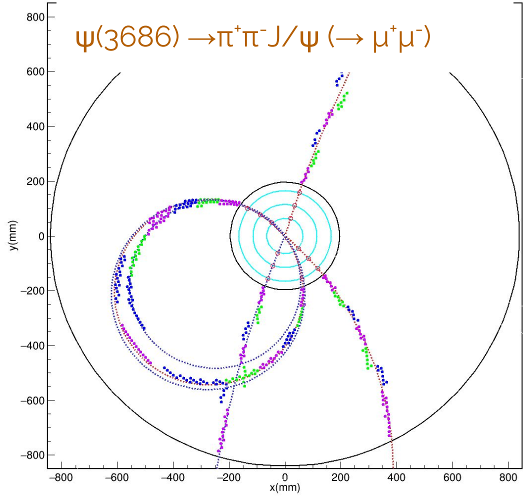
STCF tracking performance

STCF tracking challenges

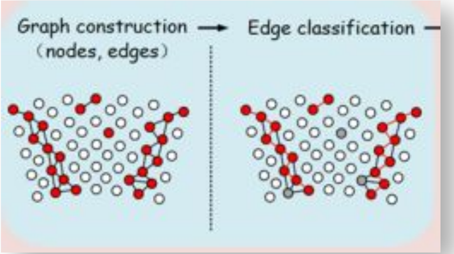
- Most physics processes have charged particles with p_T < 500 MeV/c
  - More material effects → worse resolution
  - Looping tracks with p_T < 130 MeV/c → fake/duplicate tracks
- High backgrounds and noise → worse efficiency and resolution
- Long-lived particles (non-trivial task in all HEP experiments!)

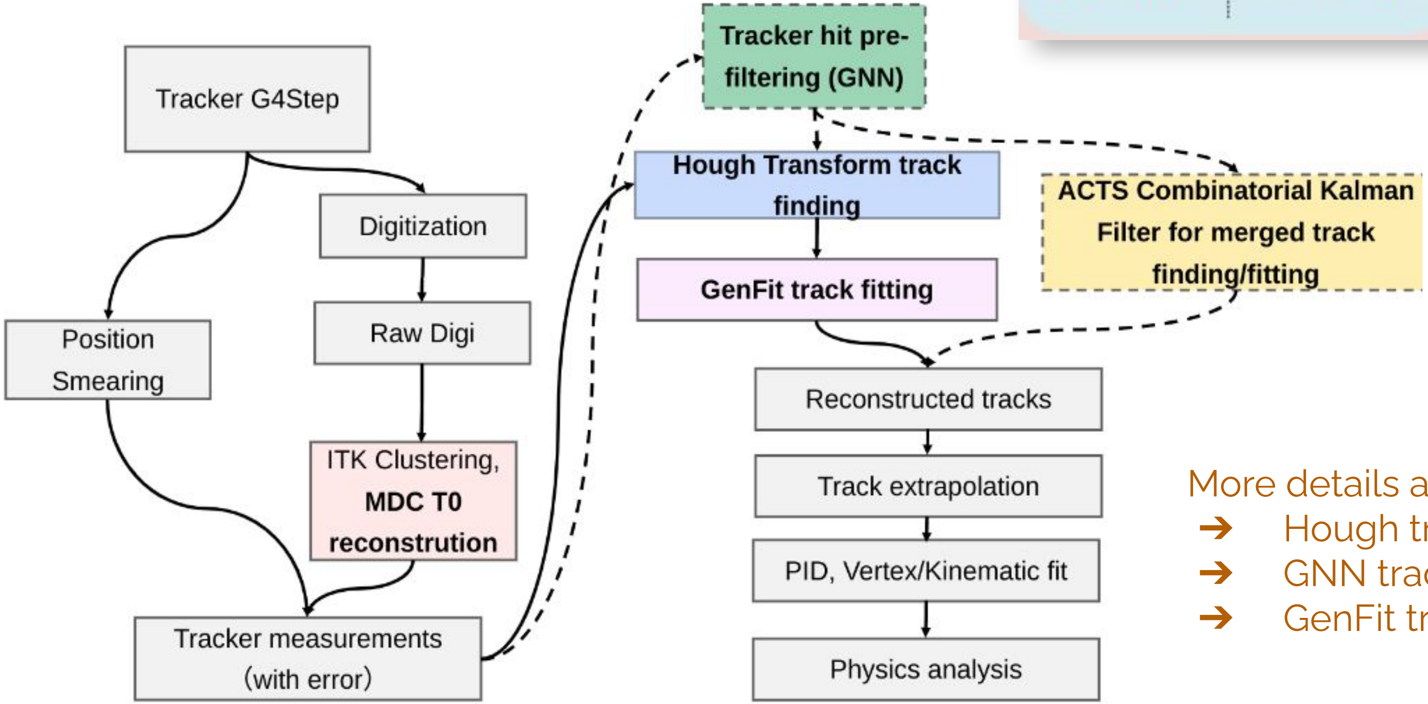
Simulated trajectory of single muon with p_T = 100 MeV, θ = 90°





STCF tracking landscape



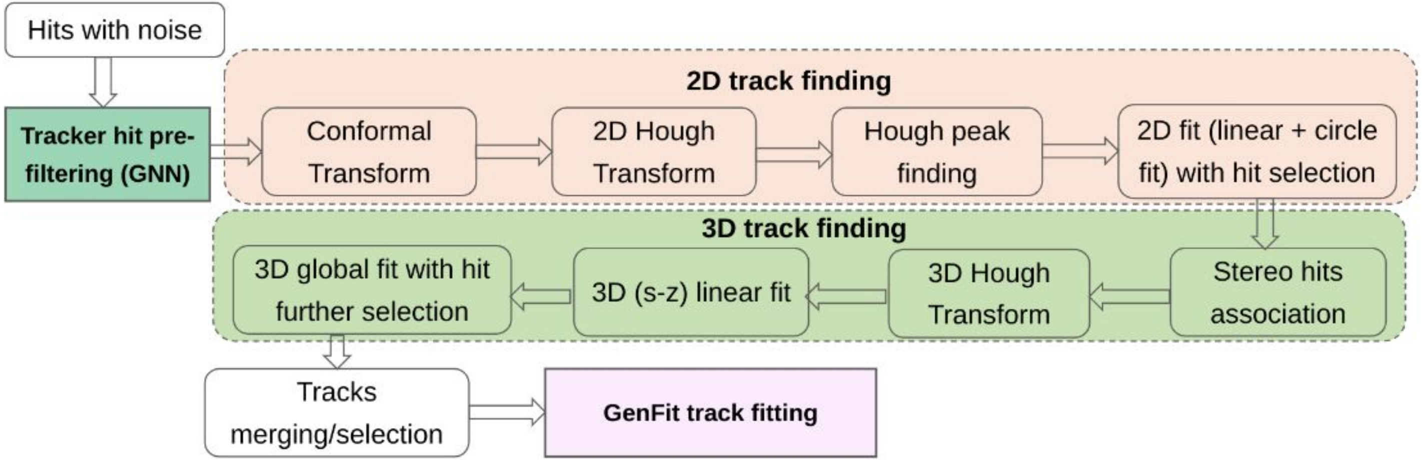


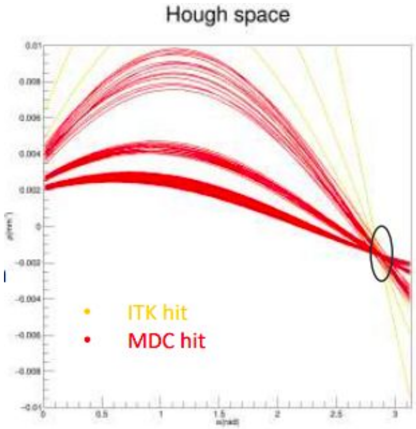
More details about:

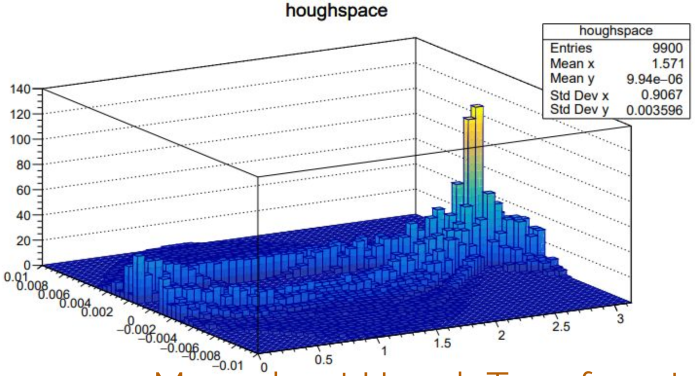
- ➔ Hough tracking: H. Zhou's talk
- ➔ GNN tracking: X. Jia's talk
- ➔ GenFit track fitting: Z. Lu' talk

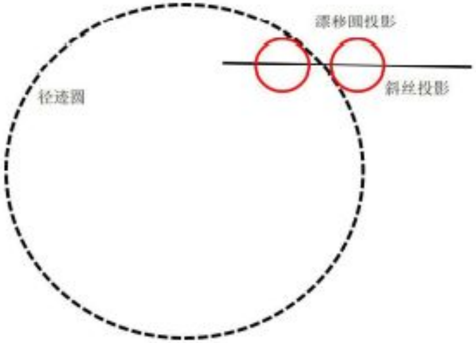
Dashed lines denote functionalities to be integrated into OSCAR

Track finding with Hough Transform



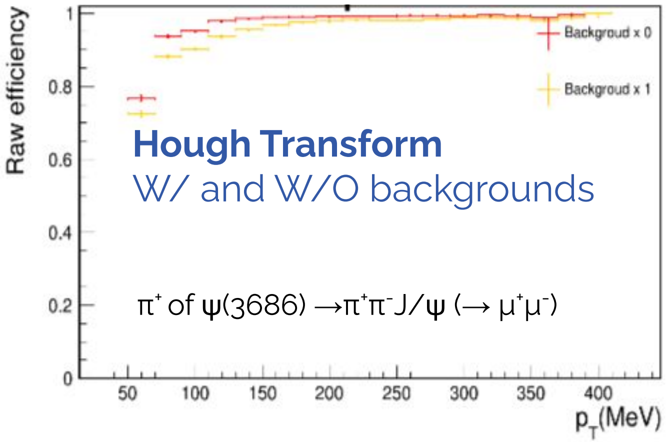


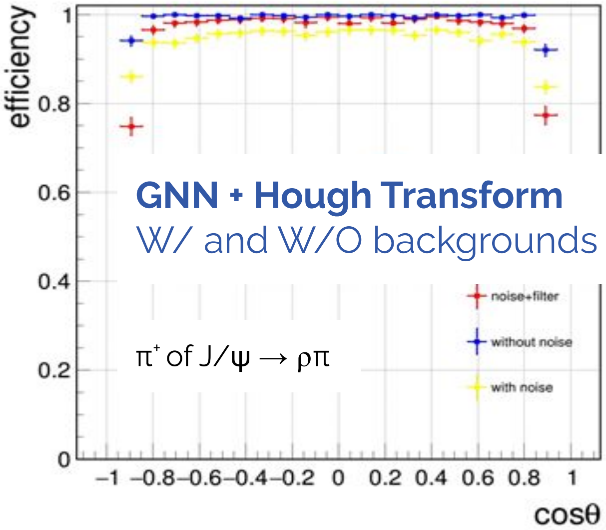


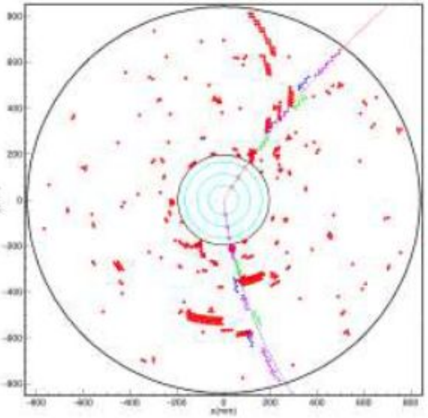


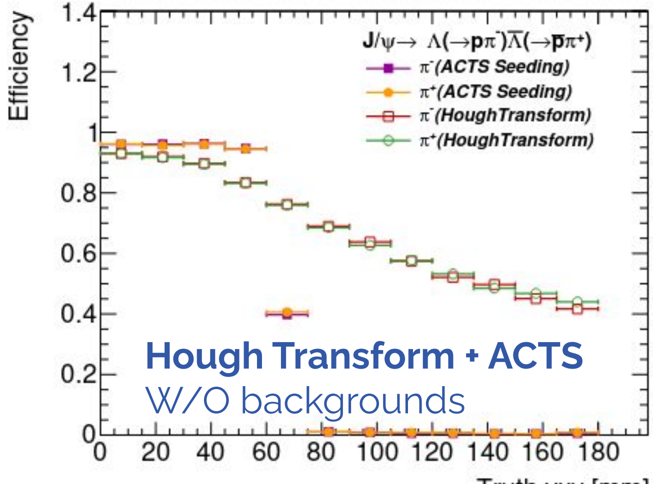
More about Hough Transform tracking in H. Zhou's talk

Track finding efficiency





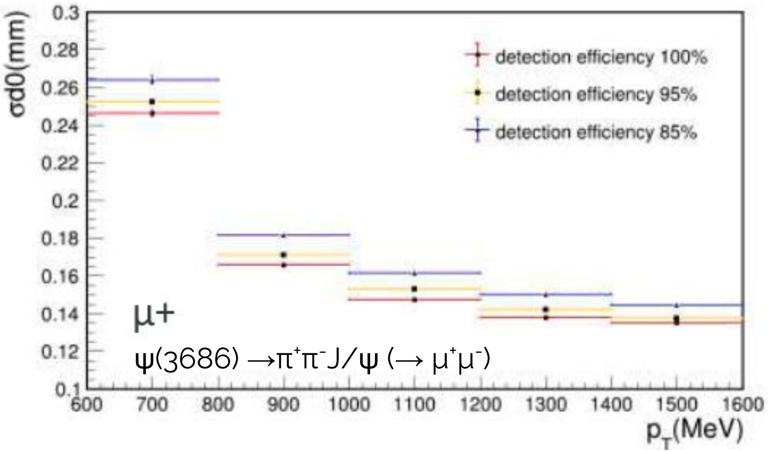


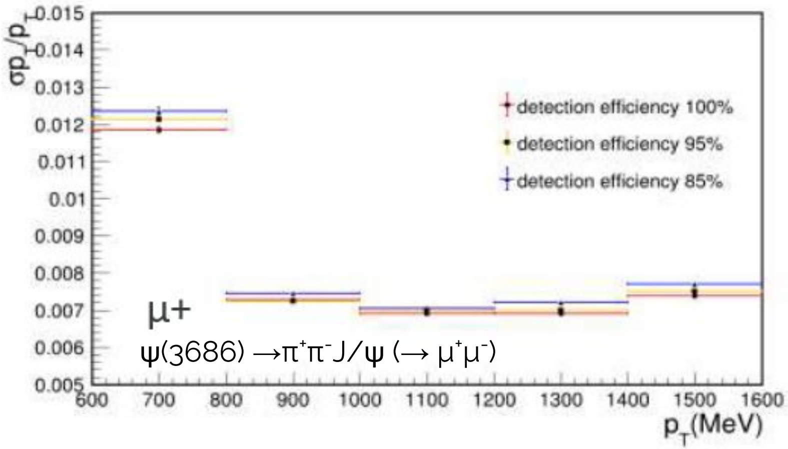


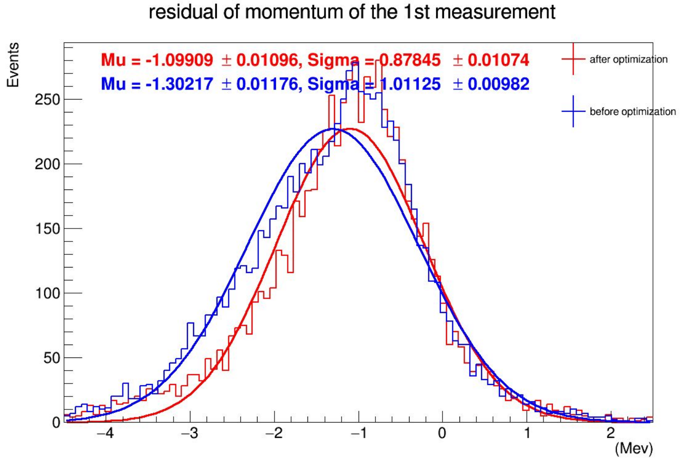
- Tracking efficiency is **above 95%** in central region for p_T > 100 MeV/c, even with backgrounds
- 99% noise hits can be removed by GNN (except first/secondary long tracks backgrounds)

Track parameters resolution

W/O backgrounds





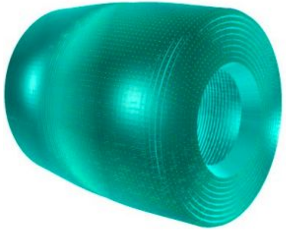


- Tracking resolution for low p_T looping tracks has been improved:
  - e.g. using only hits from first half loop, optimized GenFit workflow

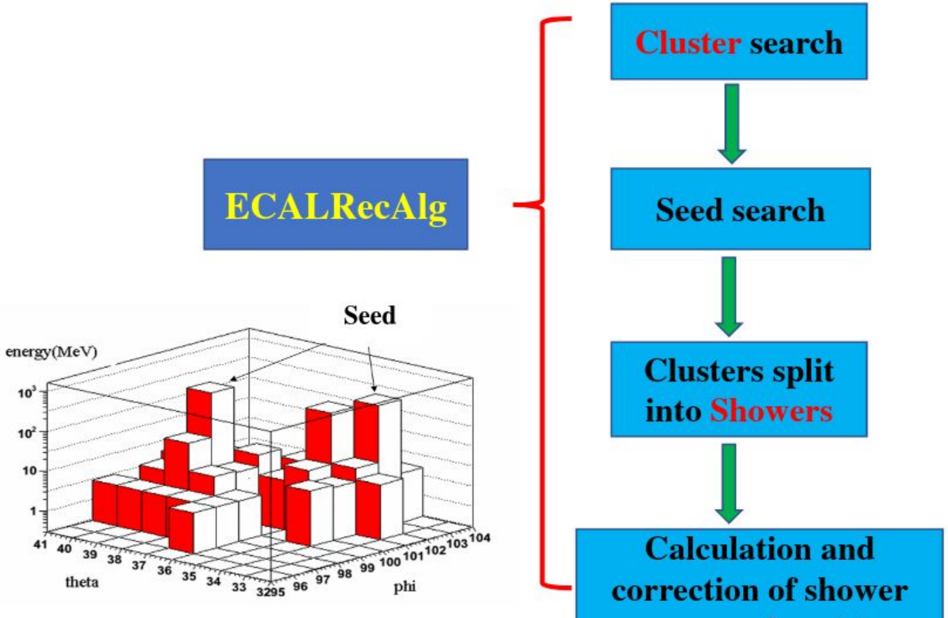
More about GenFit track fitting in Z. Lu's talk

STCF photon performance

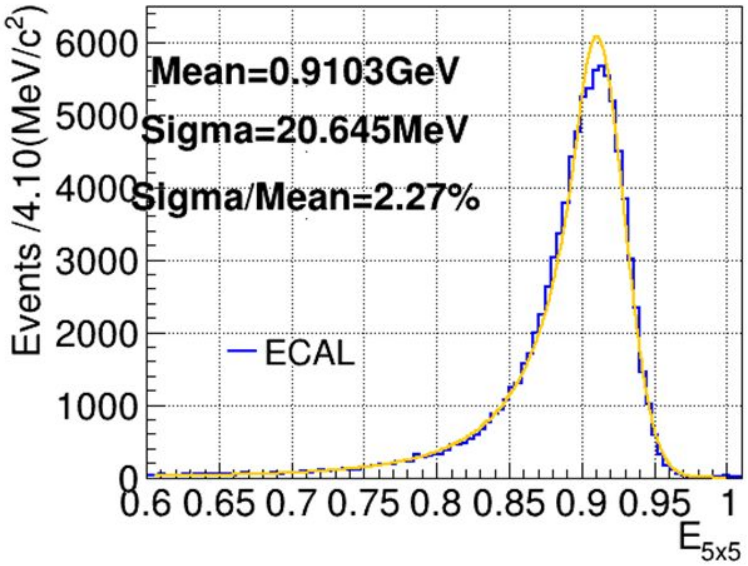
EMC reconstruction



More details in B. Wang's talk



π^0 Cluster (two photons)

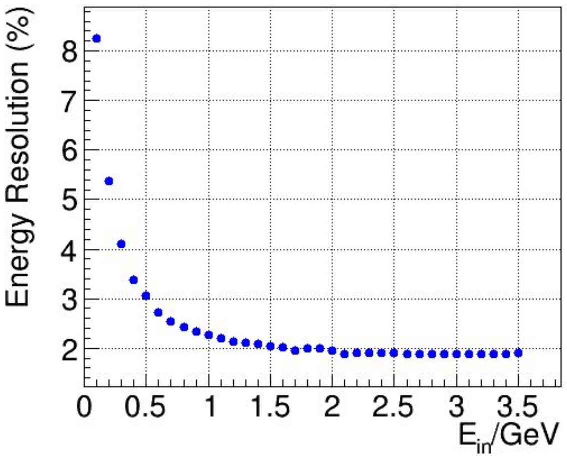


Energy distribution of 1 GeV γ (W/ backgrounds), fitted with crystal ball function

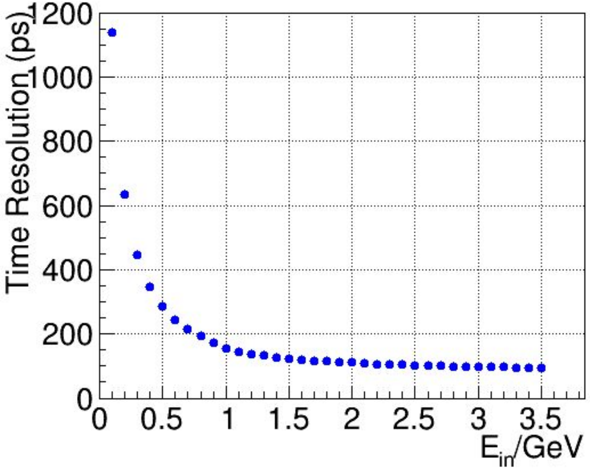
Photon performance

More details in B. Wang's talk

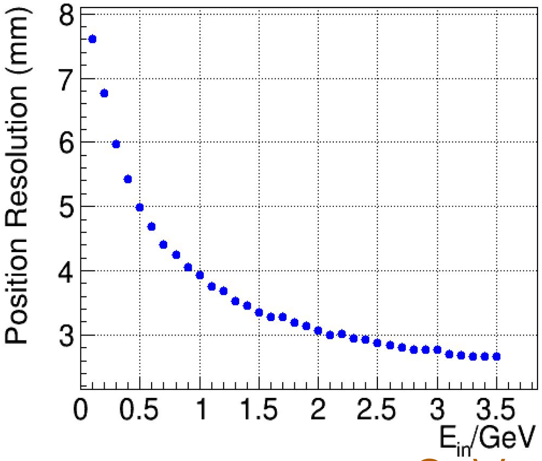
- The expected performance with backgrounds meet the physics requirements



σ_E = 2.27% @ 1 GeV



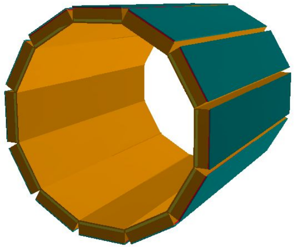
σ_t = 154 ps @ 1 GeV



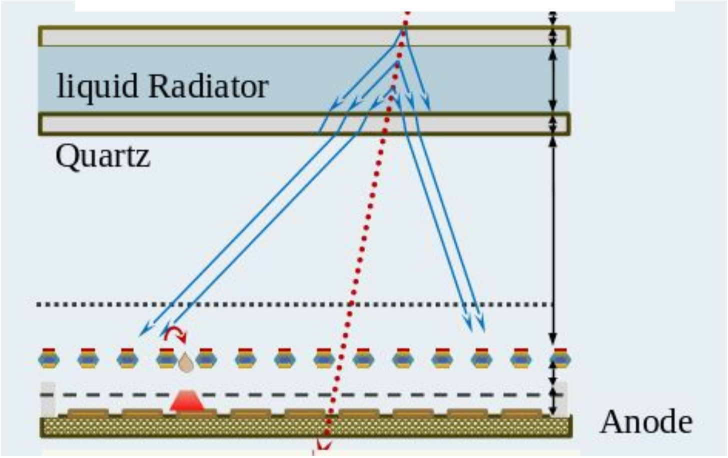
σ_{pos} = 4 mm @ 1 GeV

STCF PID performance

RICH reconstruction



More details in Q. Huang's talk

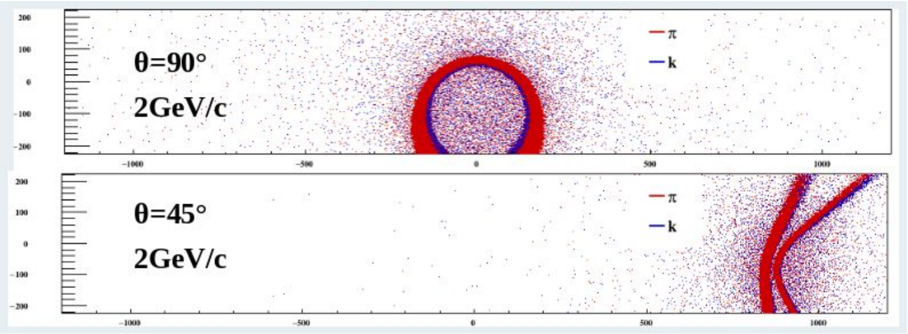


- Different distribution of number of cherenkov photons for different particles
- For each particle hypothesis h , log-likelihood defined as

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

simplified from Poisson

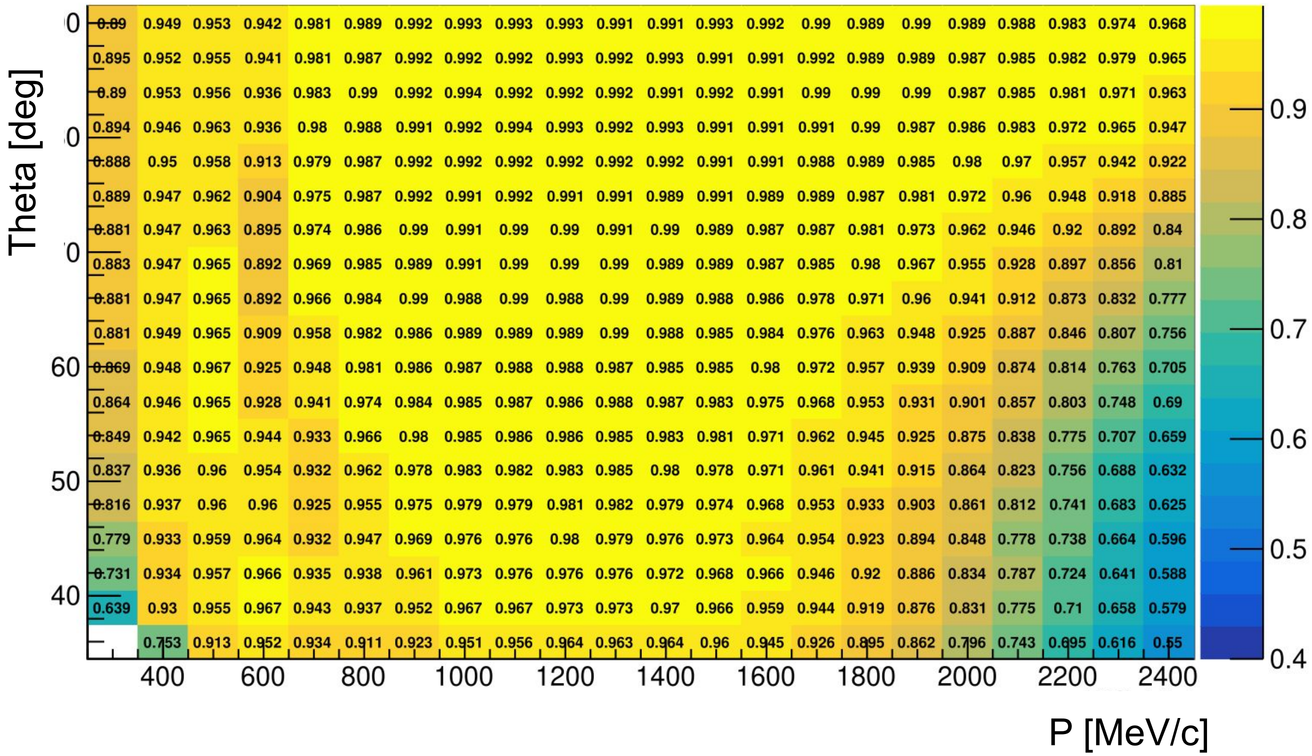
  - PDF_{Ckv} is calculated on-the-fly based on extrapolated track momentum/position
  - $DLL_{\pi K} = \ln \mathcal{L}_\pi - \ln \mathcal{L}_K$ provides particle ID



π/K PID efficiency with RICH

More details in Q. Huang's talk

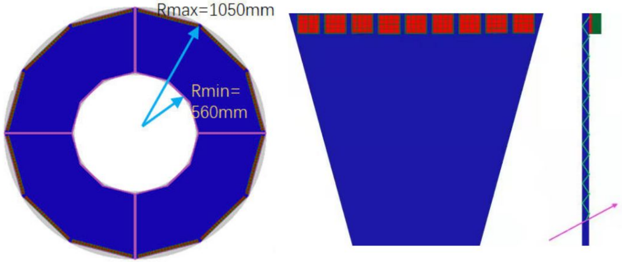
- 97% π/K PID efficiency for 0.7 GeV < p < 2 GeV with θ > 74°
- PID Efficiency for 2 GeV with smaller θ is less satisfactory



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

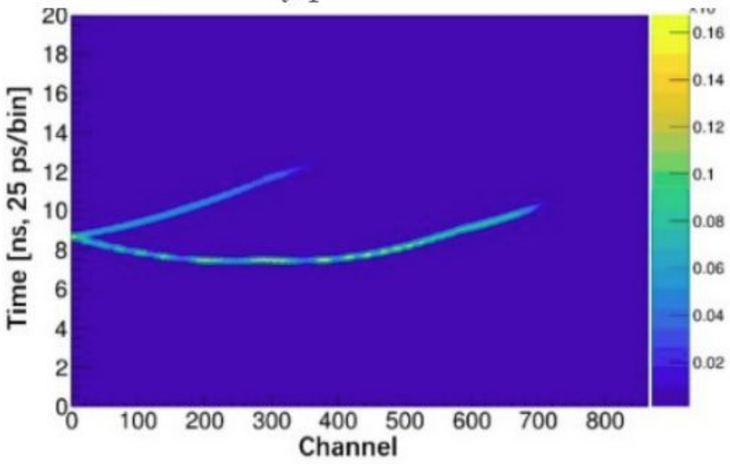
DTOF reconstruction

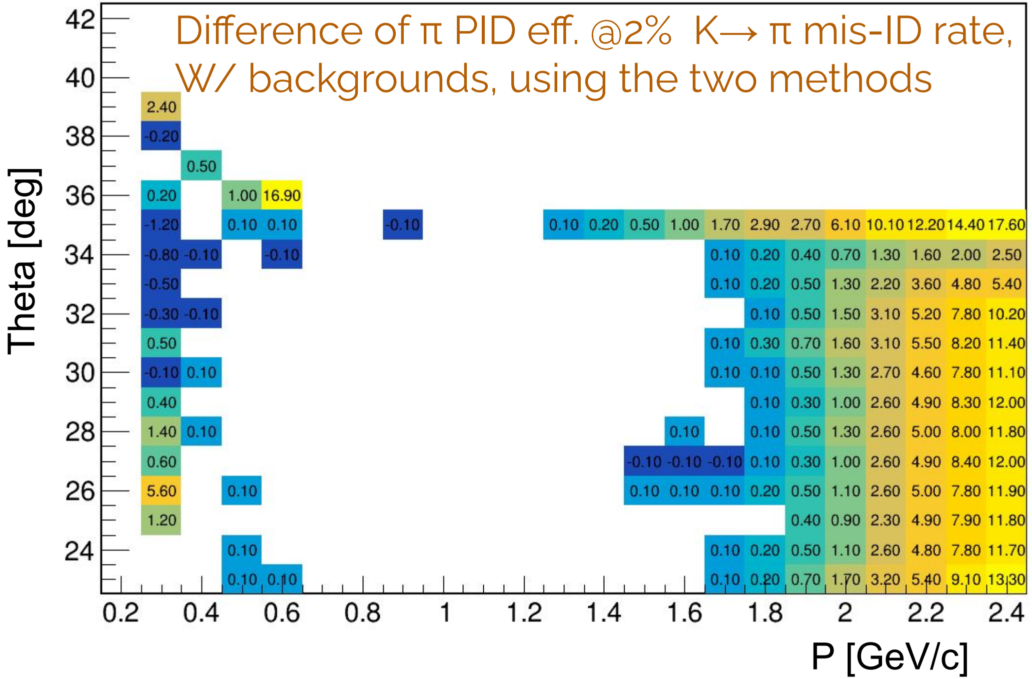
More details in Y. Feng's talk



- 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 :

$$L_h = \prod_{i=1}^{N_{p.e.}} \widetilde{N_h} S_h(ch_i, t_i) + B$$

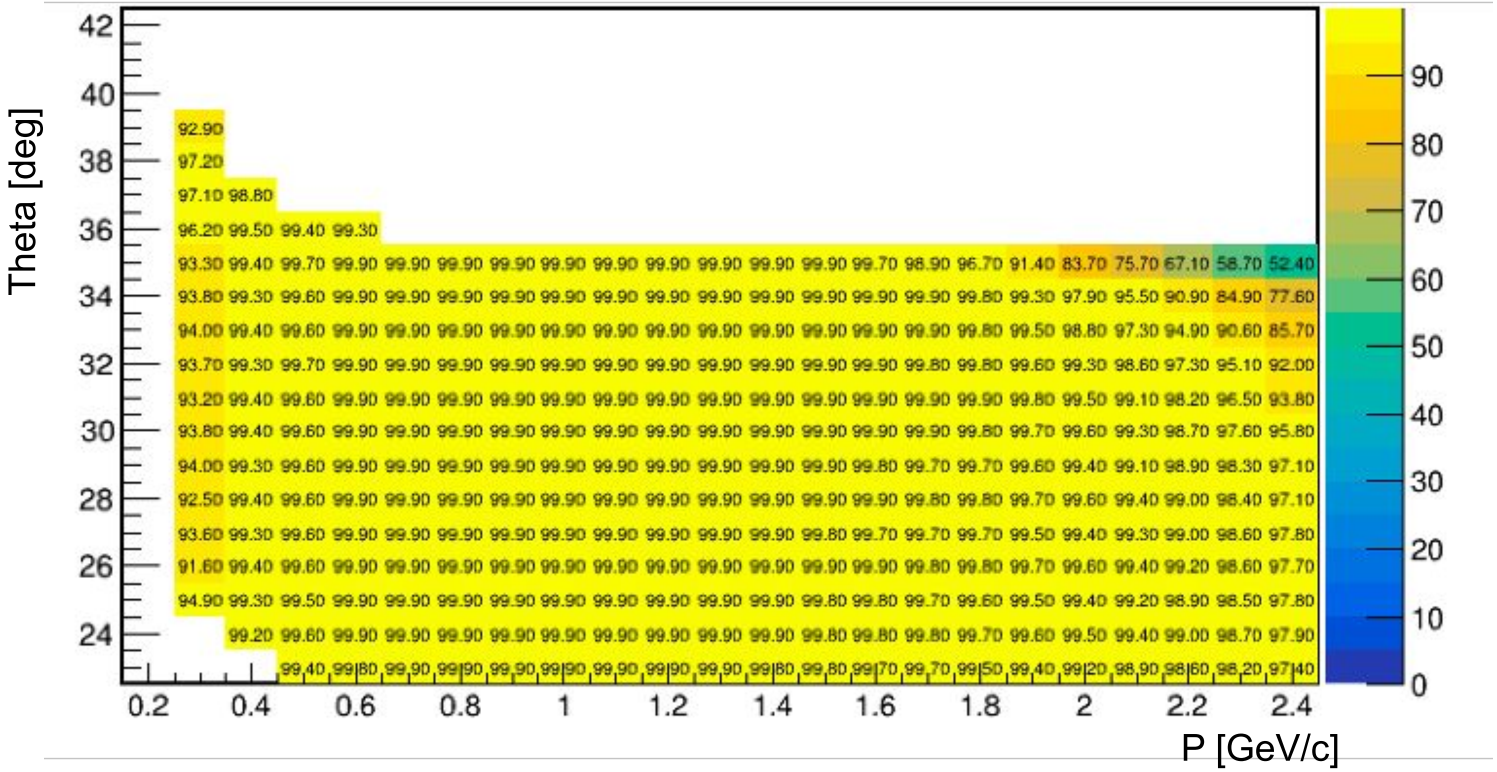




π/K PID efficiency with DTOF (Image method)

More details in Y. Feng's talk

- > 97% π/K PID efficiency for 0.35 GeV < p < 2 GeV with 24° < θ < 35°



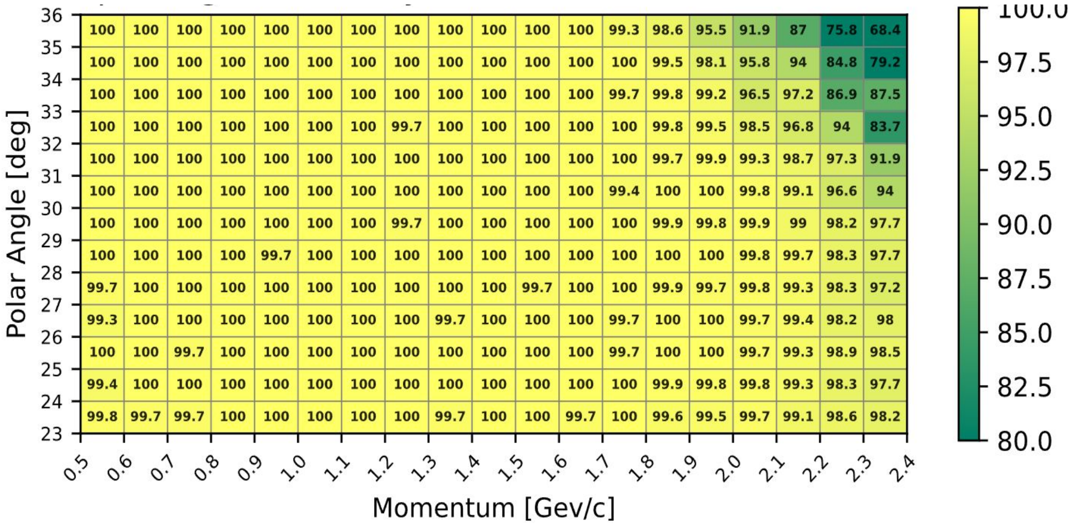
π PID eff. @2% K→ π mis-ID rate, W/ backgrounds

π/K PID efficiency with DTOF (CNN method)

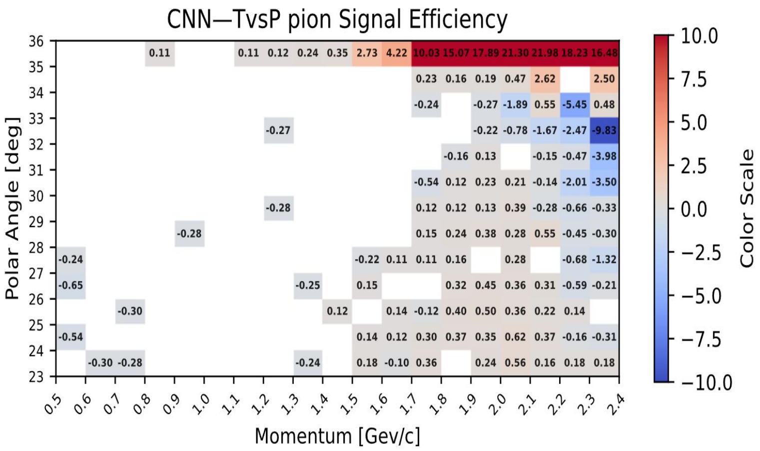
More details in Z. Yao's talk

- Using 2D map of TOF vs. channel as inputs
  - Momentum and position extrapolated to DTOF fully connected layer
- π/K PID efficiency ~99% @ p = 2.0 GeV/c (backgrounds not considered yet)

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



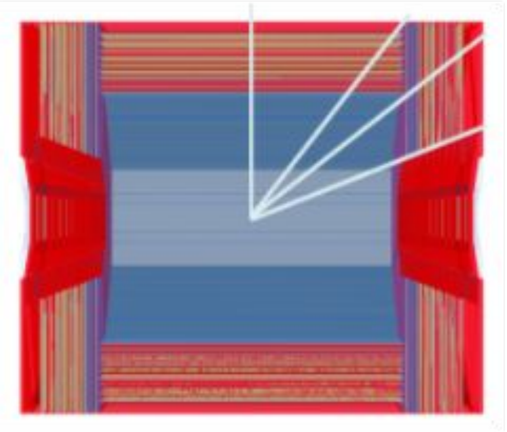
CNN might further improve performance at large theta and p



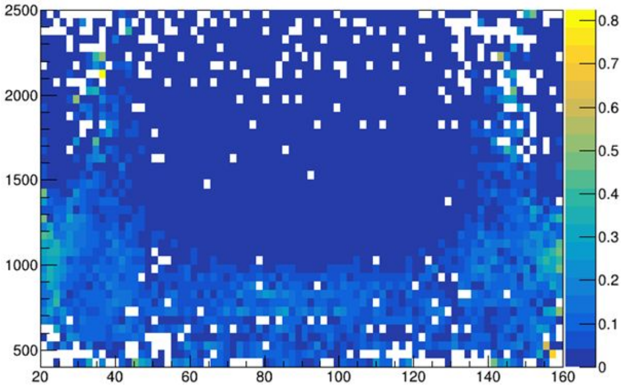
MUD reconstruction

More details in Y. Liu's talk

- 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



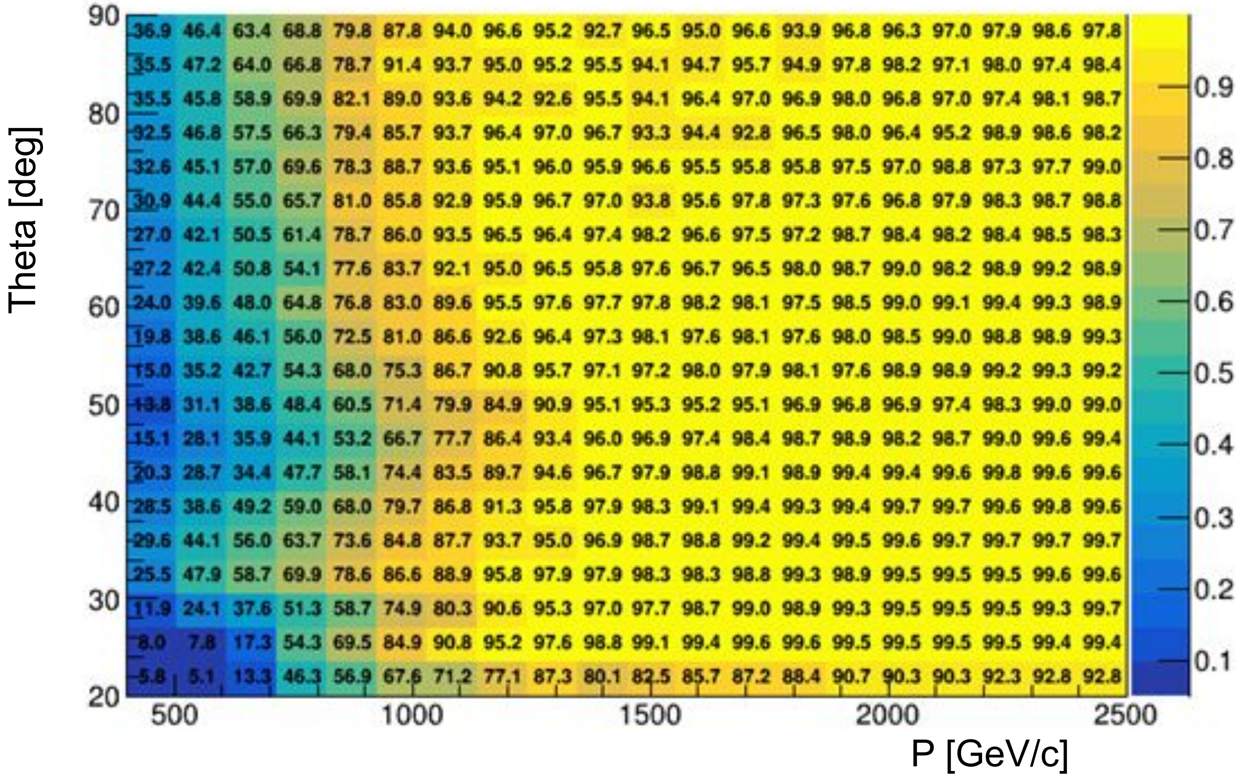
TrackInfo	ClusterInfo
MomentumMag	SeedTheta
SeedTheta	SeedPhi
SeedPhi	DeltaTheta
DeltaTheta	DeltaPhi
DeltaPhi	LargestDistance
LargestDistance	Velocity
Velocity	MaxHitLayer
MaxHitLayer	MaxHit
MaxHit	LastLayer
LastLayer	PSHitCenter
HitAverageDistance	LowHitCenter
HitEntries	HitEntries
HitInRPC	HitInRPC
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

μ/π PID efficiency with MUD

More details in Y. Liu's talk

- >95% μ PID efficiency for p > 1.2 GeV
- Needs optimization for lower momentum



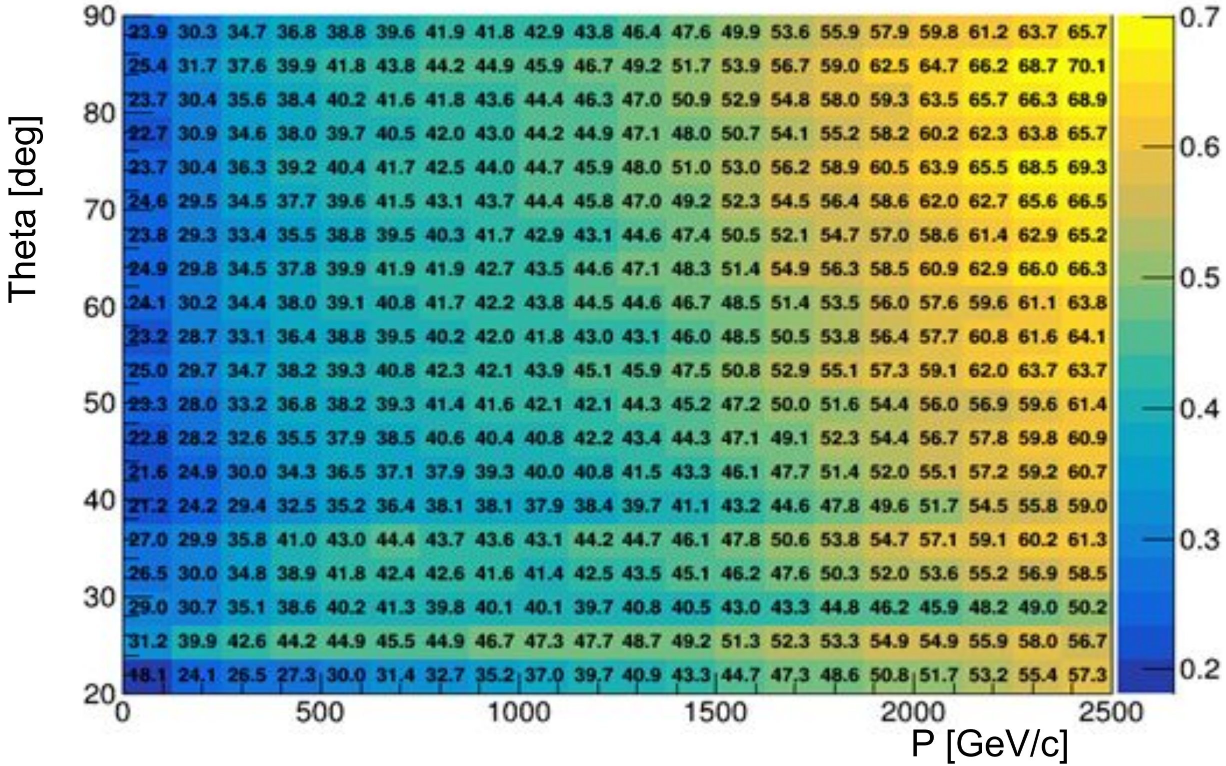
μ PID efficiency

@3.3% π→μ mis-ID rate, W backgrounds

(n, K_L)/γ PID efficiency with MUD

More details in Y. Liu's talk

- Still very preliminary. Below 70% for now. >95% is expected.



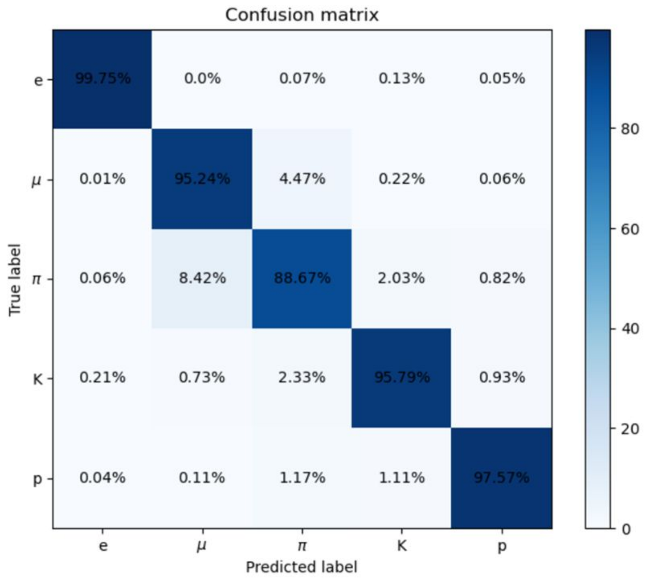
n/K_L PID efficiency

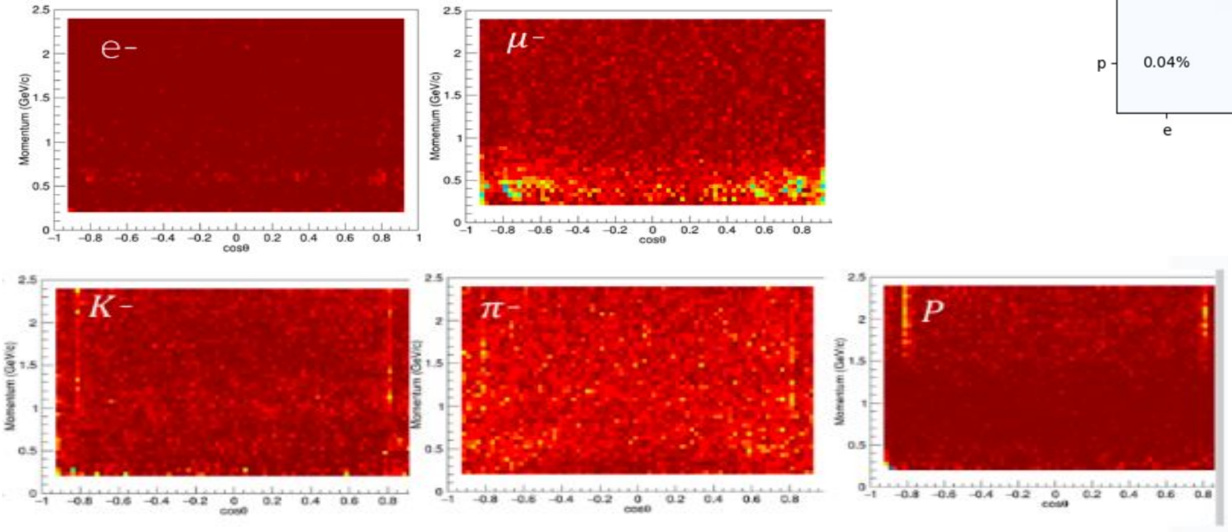
@3.3% γ→n/K_L mis-ID rate,
W/ backgrounds

Global PID using BDT

More details in Y. Zhai's talk

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



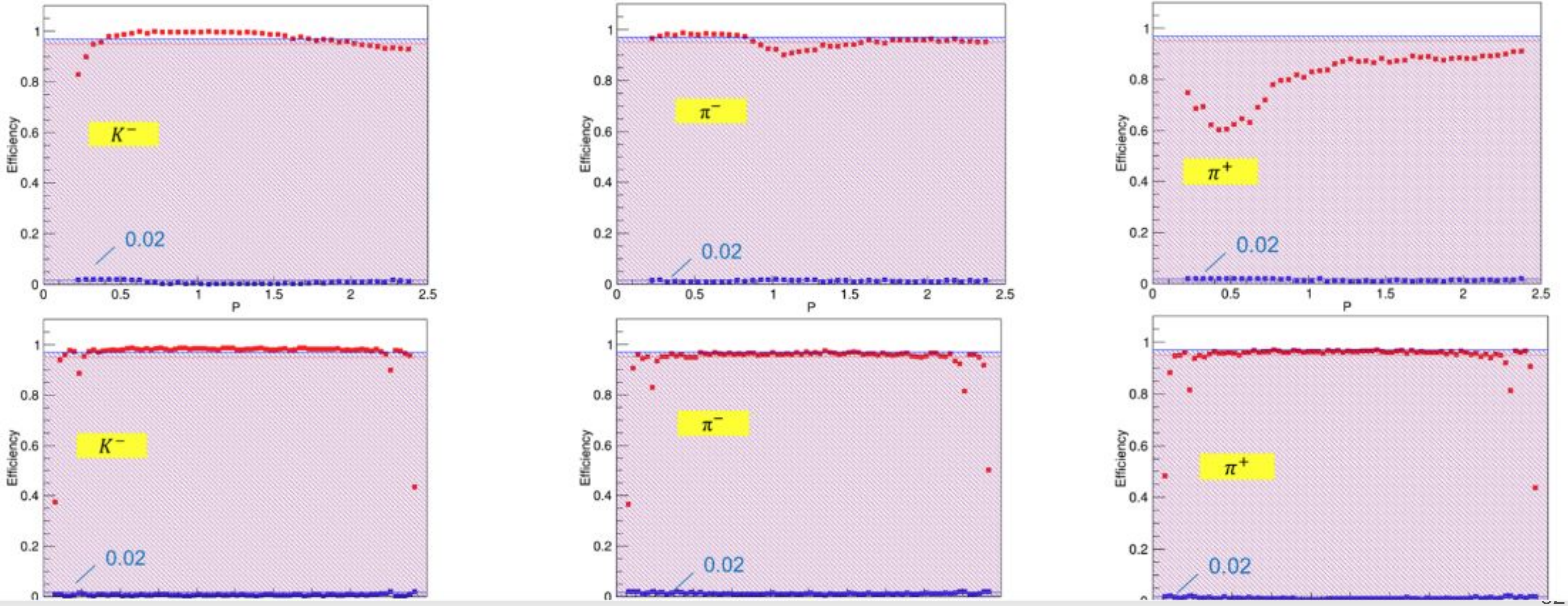


Backgrounds not considered yet

Global PID performance

More details in Y. 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%



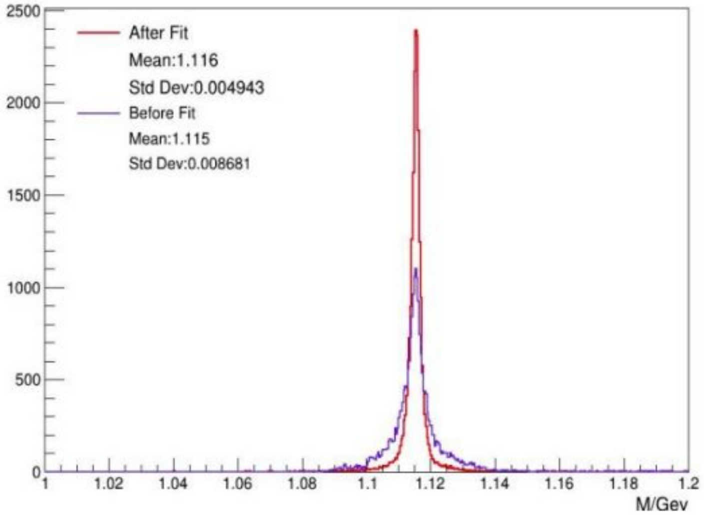
STCF Analysis tools performance

Vertex fit

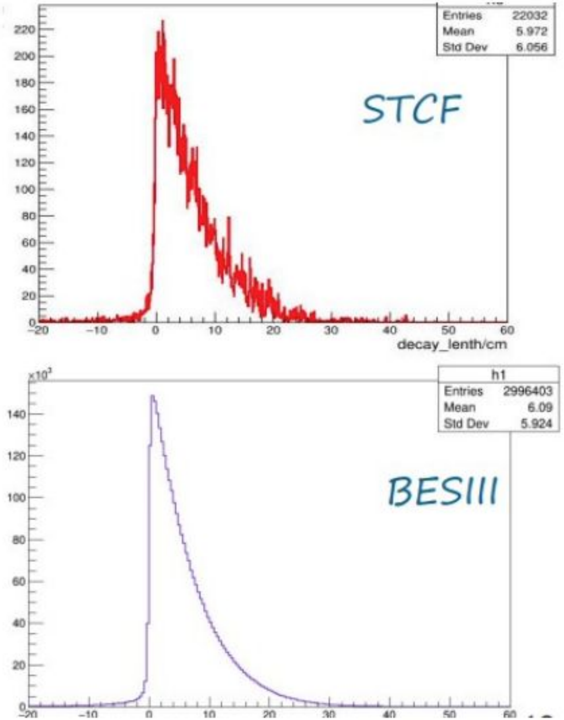
- Vertex fit transcribed from BESIII has been validated

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

Invariant mass of Λ



Decay length

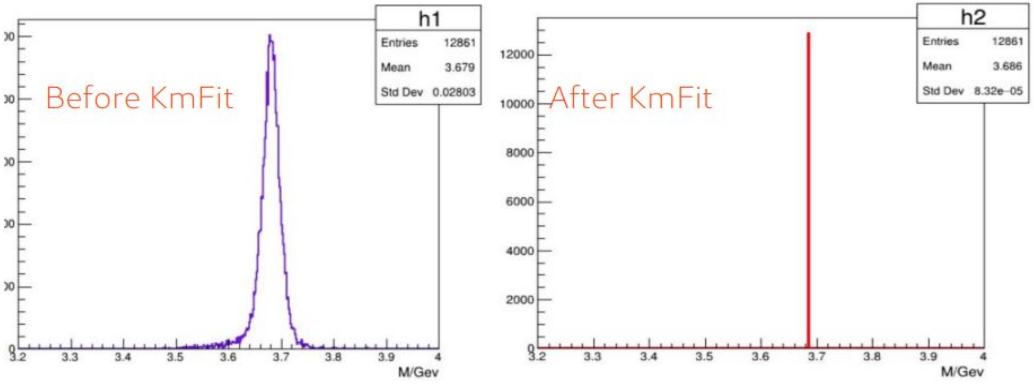


Kinematic fit

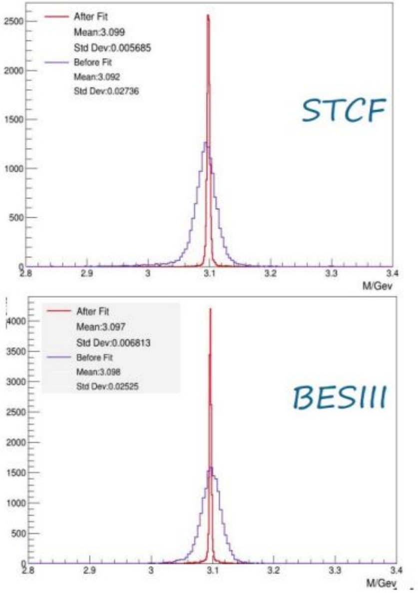
- Kinematic fit transcribed from BESIII has been validated
- GlobalVertexFit which combines vertex and kinematic fit has been transcribed from Belle-II recently. Validation is in progress.

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

Invariant mass of ψ(2S)



Invariant mass of J/ψ



Summary

Status of the full event processing chain

✅ Finalised   🟠 Working, under optimization   ❗ 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		✅	✅	🟠	🟠	1) Single particles ✅ 2) Physics Processes 🟠	1) Single particles 🟠 2) Physics Processes ❗	✅
MDC		✅	✅					
RICH		✅	✅	🟠	🟠			
DTOF		✅	✅	✅	✅			
ECAL		✅	✅	✅	✅			
MUD	Pre-geometry	✅	🟠	🟠	🟠			
	New geometry	✅	❗					

Summary

- Much progress has been made towards building **a full event processing chain** for STCF in the past year
  - Background simulation + mixing, digitization, reconstruction optimization…
- **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

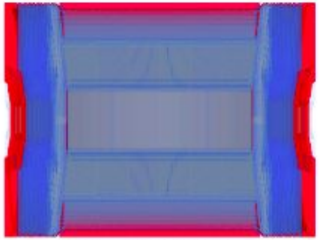
backup

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 : **$J/\psi \rightarrow \rho\pi$**

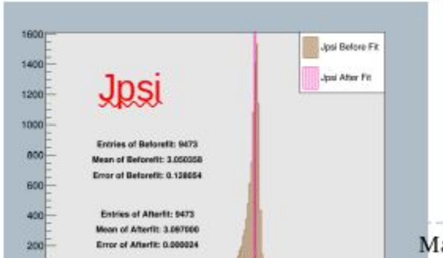
* 事例选择

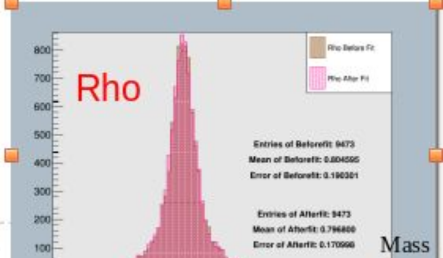
- 带电径迹的选择
  - $N_{\text{good charge track}} = 2$
  - Total charge=0
- 中性径迹的判断
  - 桶部光子 Eγ>25MeV (|cosθ|<0.8)
  - 端盖光子 Eγ>50MeV (0.86<|cosθ|<0.92)
  - $N_{\text{good neutral track}} > 2$
- Global PID
  - 判选条件
- 顶点拟合 & 运动学拟合
  - $\chi^2 > 200$

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