中国科学技术大学 University of Science and Technology of China
中国科学院大学 University of Chinese Academy of Sciences
湖南科技大学 Hunan University of Science and Technology

2025年超级陶粲装置研讨会

湖南科技大学，湖南 · 湘潭，2025年7月2日-6日

STCF global particle identification based on combined likelihood

齐斌斌
中国科学技术大学

On behalf of the STCF PID group
2025.07.04

STCF
超级陶粲装置
Super Tau-Charm Facility

摘要

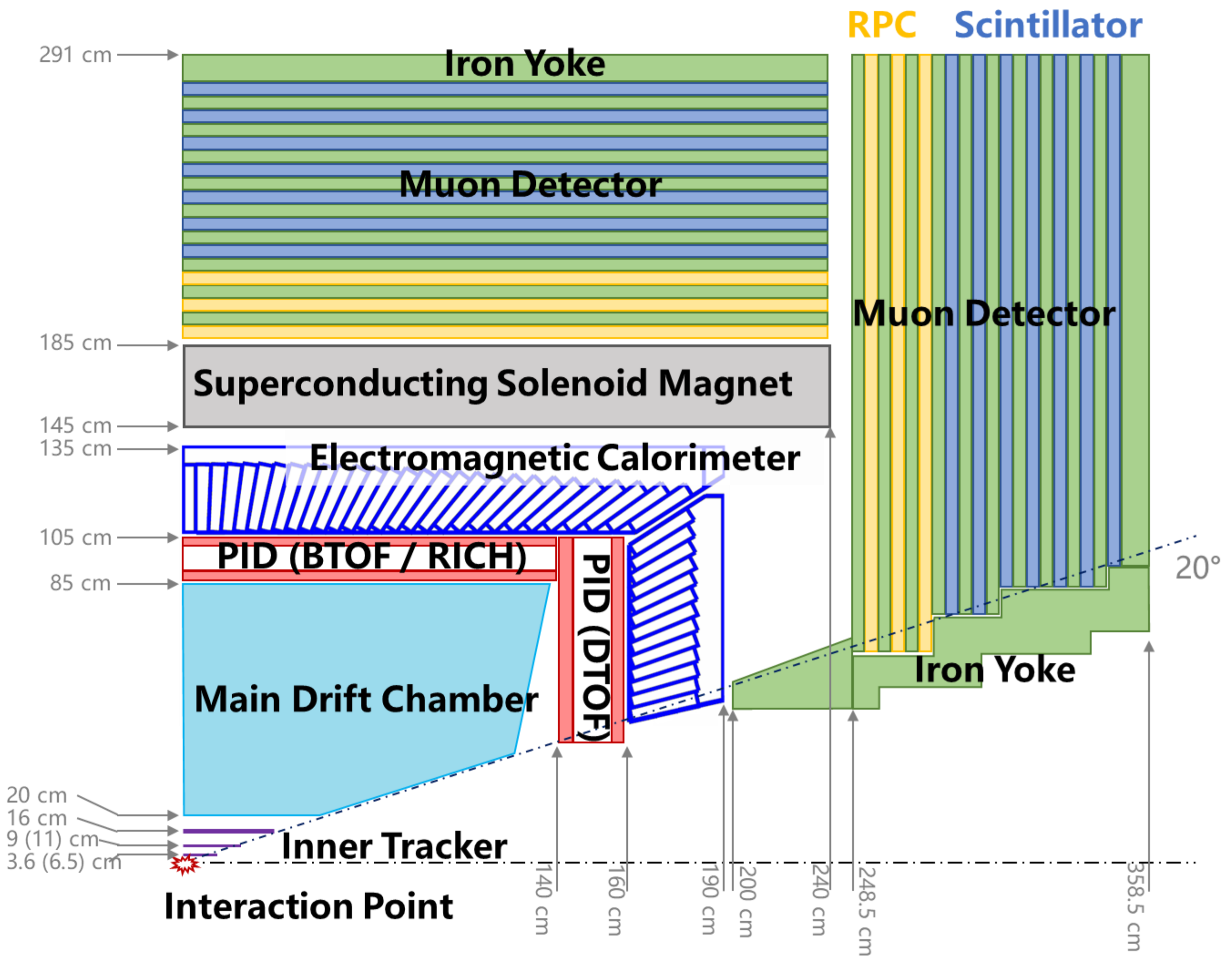
一、STCF探测器介绍

二、全局粒子鉴别

- Simple combined likelihood
- Combined likelihood with weight
- Boosting Decision Tree(BDT)

三、小结与展望

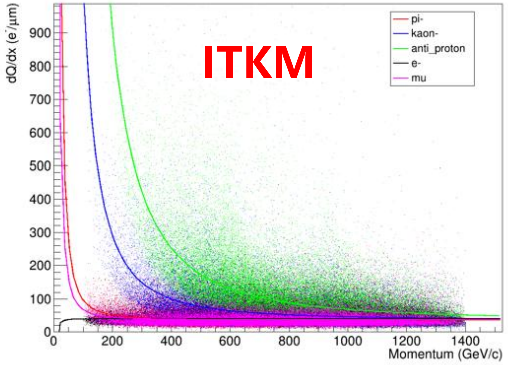
STCF探测器

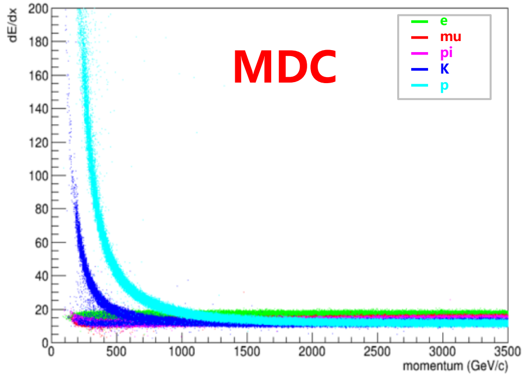


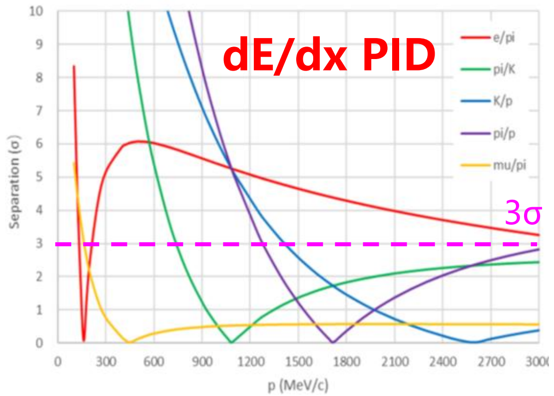
- **Inner tracker (ITK, two options)**
  - MPGD: cylindrical MPGD
  - Silicon: CMOS MAPS (dE/dx)
- **Central tracker (MDC)**
  - Main drift chamber (dE/dx)
- **Particle identification** (Cherenkov)
  - Barrel: **RICH** with CsI-MPGD or barrel DTOF (**BTOF**)
  - Endcaps: DIRC-like TOF (**DTOF**)
- **Electromagnetic calorimeter (EMC)**
  - pure CsI + APD (electron ID)
- **Muon detector (MUD)**
  - RPC + scintillator strips (muon ID)
- **Magnet**
  - Super-conducting solenoid, 1 T

探测器PID性能

- **dE/dx测量**



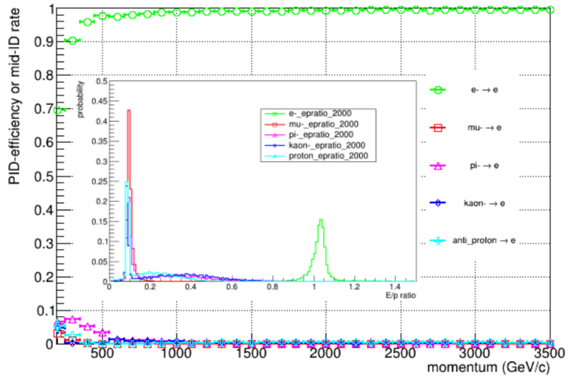




➢ pi/K：>3σ at p<750 MeV/c, >2σ at p>2 GeV/c

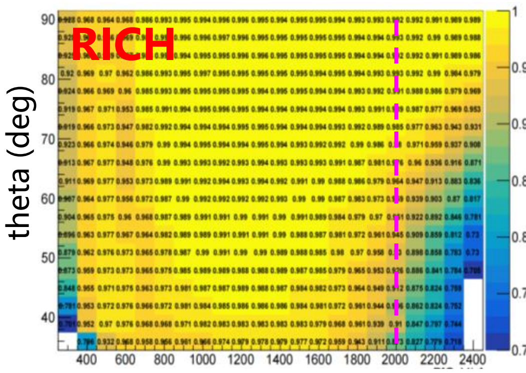
➢ mu/pi：>3σ at p<150 MeV/c; e/pi：>3σ except 120-250 MeV/c

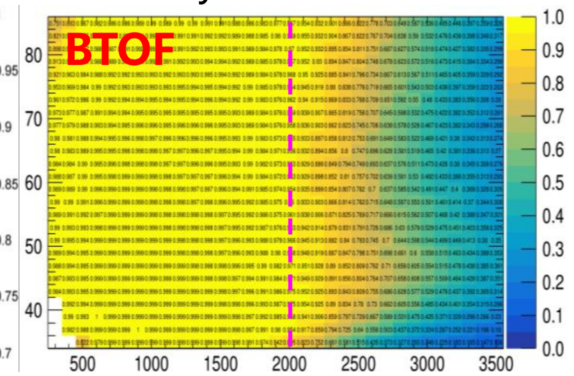
- **电子鉴别（ECAL）**

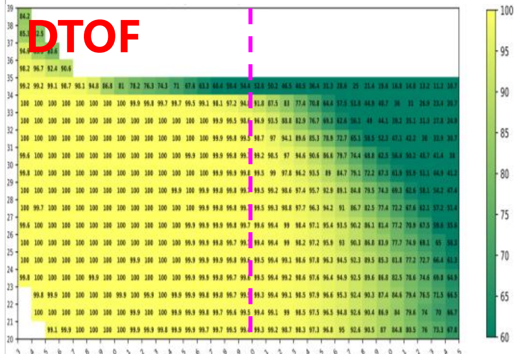


➢ >98% electron efficiency at p>500 MeV/c

- **PID探测器** Pion efficiency with kaon mis-ID=2%



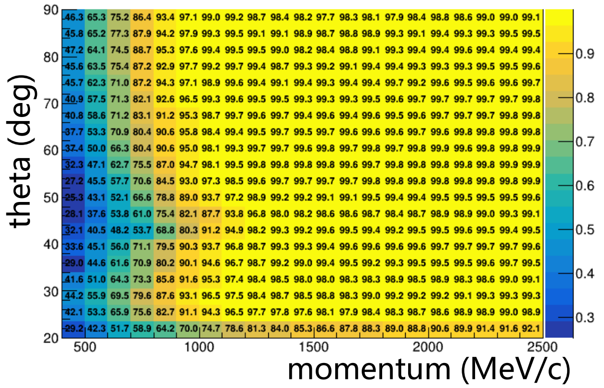




momentum (MeV/c)

➢ >97% pion efficiency with kaon mis-ID=2% at p<2 GeV/c

- **缪子鉴别（MUD&ECAL）**



➢ >95% muon efficiency with pion mis-ID=3% at p>1 GeV/c

Combined likelihood

- **dE/dx测量**
  - 五种粒子假设下的卡方值： $\chi_h^2 = \frac{\left(\frac{\mathrm{d}E}{\mathrm{d}x} - \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_h\right)^2}{\sigma_h^2}$
  - Likelihood： $\mathcal{L}_{\mathrm{dEdx}} = e^{-\frac{\chi^2}{2}}$
- **PID探测器**
  - 五种粒子假设下的likelihood： $\mathcal{L}_{\mathrm{PID}}$
- **ECAL鉴别电子**
  - BDT模型，鉴别为电子概率： P_e
  - Likelihood： $\mathcal{L}_{\mathrm{ECAL}} = \begin{cases} P_e, \text{for electron} \\ 1 - P_e, \text{for others} \end{cases}$
- **MUD鉴别缪子**
  - BDT模型，鉴别为缪子概率： $P_{mu} = \frac{\mathrm{BDTExp}+1}{2}$
  - Likelihood： $\mathcal{L}_{\mathrm{MUD}} = \begin{cases} P_{mu}, \text{for muon} \\ 1 - P_{mu}, \text{for others} \end{cases}$

- **Combined likelihood**

$$\mathcal{L}_h = \prod_{\mathrm{det}} \mathcal{L}_{\mathrm{det},h}$$

- **Likelihood ratio**

$$p(h) = \frac{P_{\mathrm{prior},h}\mathcal{L}_h}{\sum_h P_{\mathrm{prior},h}\mathcal{L}_h}$$

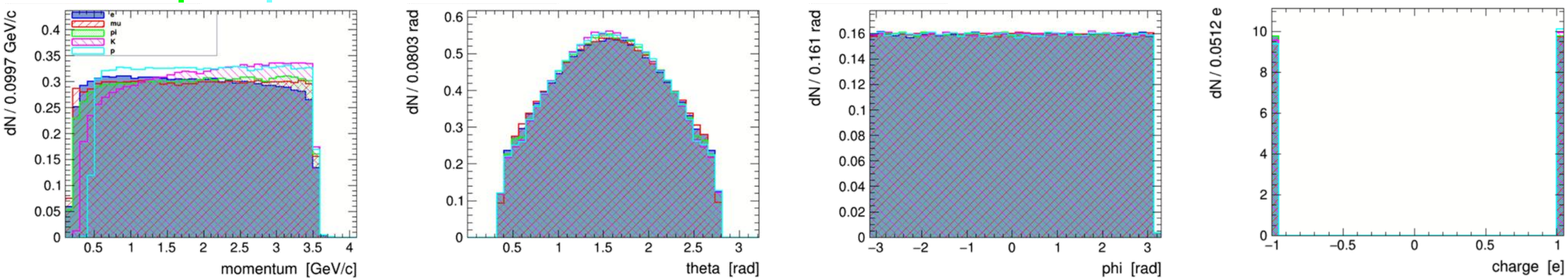
先验概率，即待测样本不同粒子比例

当$\boldsymbol{p(a)} = \max\{\boldsymbol{p(h)}\}$时，粒子鉴别为a

样本选择

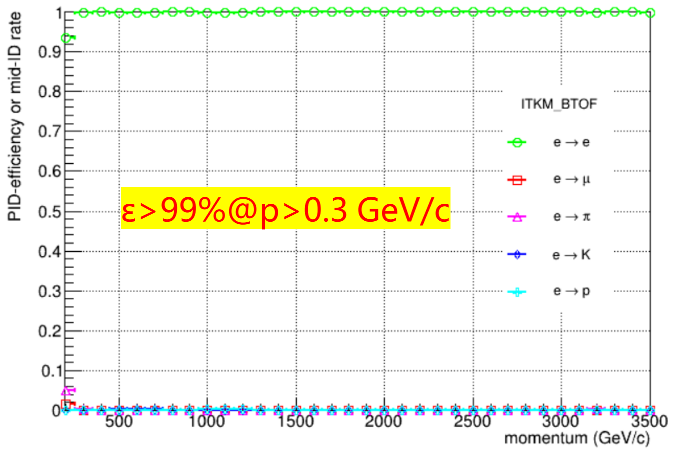
- **样本产生**
  - e/mu/pi/K/p单粒子，动量为0.2-3.5 GeV/c，角度θ为20°-160°，φ为0°-360°
- **样本筛选**
  - 来自顶点：abs(D0)<0.5 cm && abs(Z0)<0.5 cm
  - 重建径迹与原初MC粒子匹配：$\mathrm{TrackID}_{MC,matched} = \mathrm{trackID}_{MC,primary}$
  - 穿过PID探测器：MC粒子endpoint要求 (|R|>1050 || |Z|>1600) && |cos θ|<0.94;

e/mu/pi/K/p

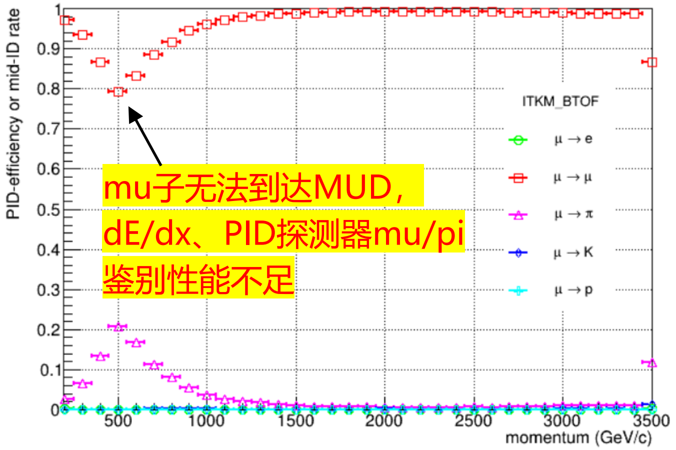


Global PID性能

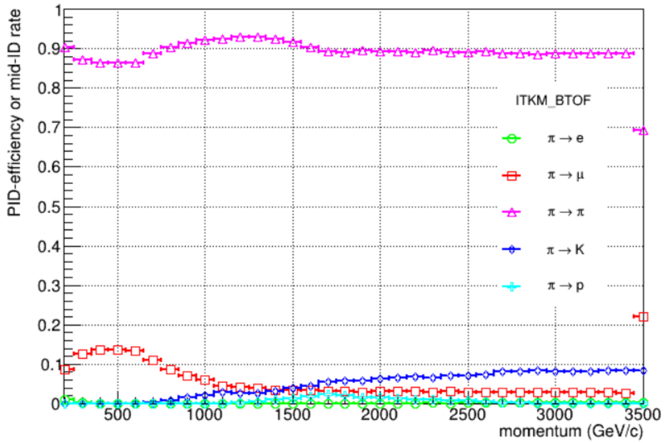
e sample



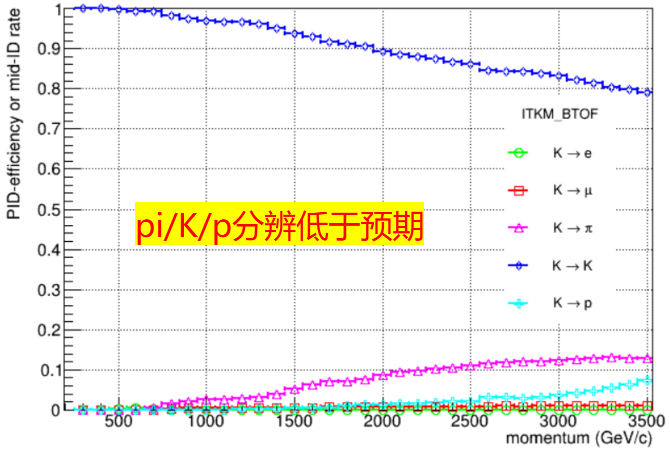
mu sample



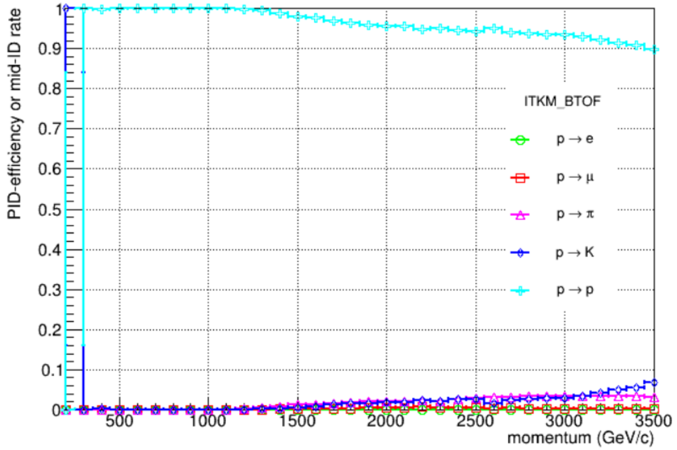
pi sample



K sample



p sample



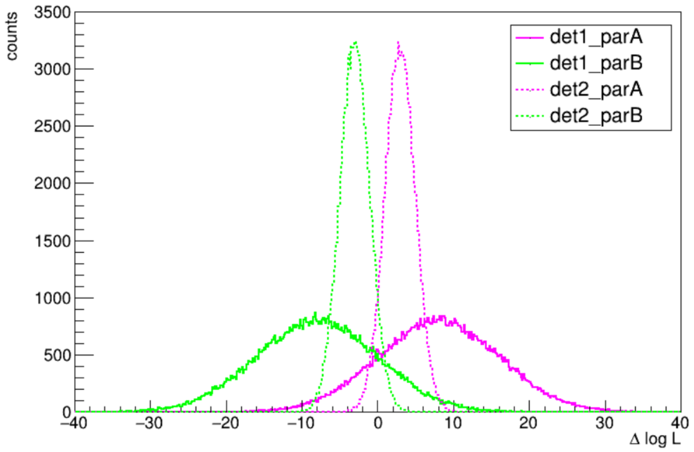
Combined likelihood with weight

- 两个探测器**det1**和**det2**，对**A**、**B**粒子分辨能力分别为**2σ**和**3σ**
  - **Simple**：$\mathcal{L}_h = \mathcal{L}_{\mathrm{det1},h} * \mathcal{L}_{\mathrm{det2},h}$，$N_\sigma = \frac{\bar{\mu}_{\mathrm{det1}}+\bar{\mu}_{\mathrm{det2}}}{\sqrt{\bar{\sigma}^2_{\mathrm{det1}}+\bar{\sigma}^2_{\mathrm{det2}}}} \times 2$
  - **Weight**：$\mathcal{L}_h = \mathcal{L}^{w_1}_{\mathrm{det1},h} * \mathcal{L}^{w_2}_{\mathrm{det2},h}$，$w_i \propto \frac{\bar{\mu}_{\mathrm{det}i}}{\bar{\sigma}^2_{\mathrm{det}i}} \propto \frac{N_{\sigma i}}{\bar{\sigma}_{\mathrm{det}i}}$，$N_\sigma = \sqrt{N^2_{\sigma 1} + N^2_{\sigma 2}}$

受ΔlogL分布影响

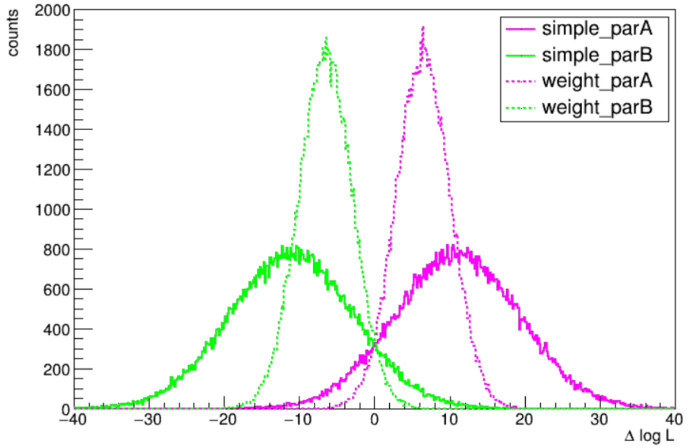
排除ΔlogL分布影响，

结果最优化



----------- **det1, Nσ=2**

———— **det2, Nσ=3**

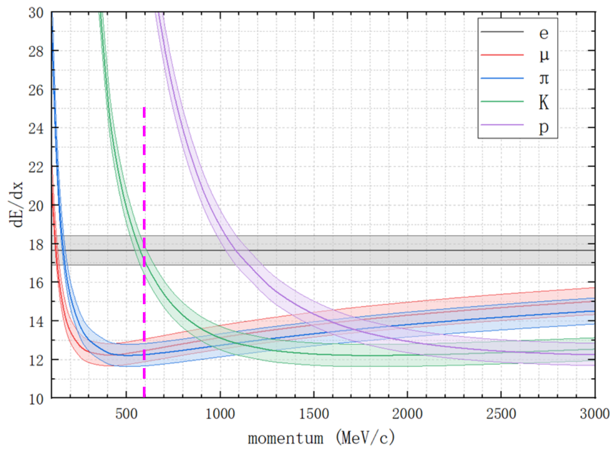


----------- **simple, Nσ=2.67**

———— **weight, Nσ=3.59**

Combined likelihood with weight

- 存在**多个独立**的likelihood，针对五种粒子类型的**两两比较**，需要根据**不同探测器**在**不同动量**区间的鉴别能力，对各likelihood分量**分别赋予权重**
  - 例如，**dE/dx**测量在**0.6GeV/c**动量下对e/K鉴别能力差，但是对e/pi与pi/K鉴别能力较好，应当设置不同权重



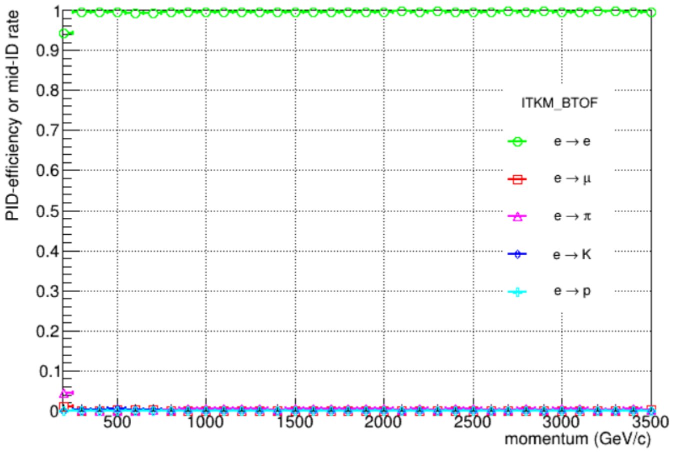
- 五种粒子两两比较，一共有**10×2**种概率值
- 假设五种粒子假设概率分别为$\boldsymbol{P_i}$
  - 构造函数$\chi^2 = \Sigma_{i\neq j}(\frac{P_{ij}(i)}{P_{ij}(i)+P_{ij}(j)} - \frac{P_i}{P_i+P_j})^2$，求**极小值**得到$P_i$（使用Timuit收敛目标函数）
  - $P_{ij}(i) = \prod_{det} P_{prior,i}\mathcal{L}(det, i)^{w_{ij}(det)}$，表示i、j粒子在两两比较情况下的联合likelihood

> 如果ΔlogL为**非对称高斯**分布，则通过简单调整权重**无法实现最优鉴别效果**
> 需要借助**BDT或其他机器学习**方法来进一步提升性能。

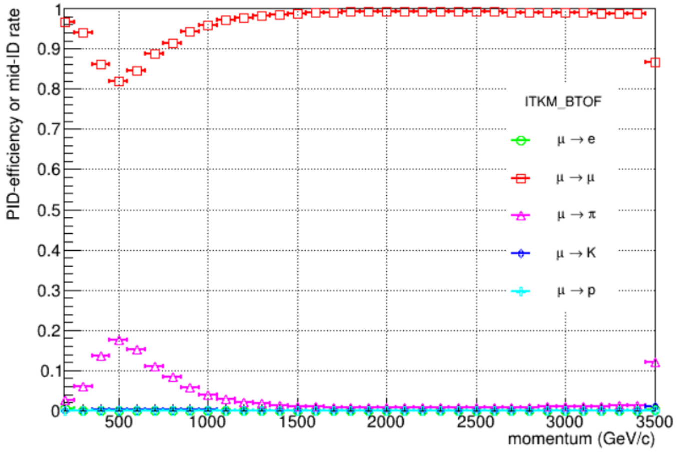
Weight likelihood 性能检验

- e、mu鉴别效率无明显变化
- pi、K、p的鉴别效率提升明显

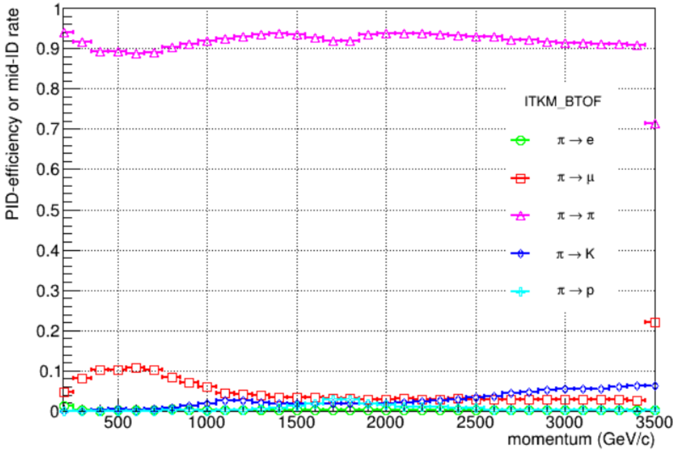
e sample



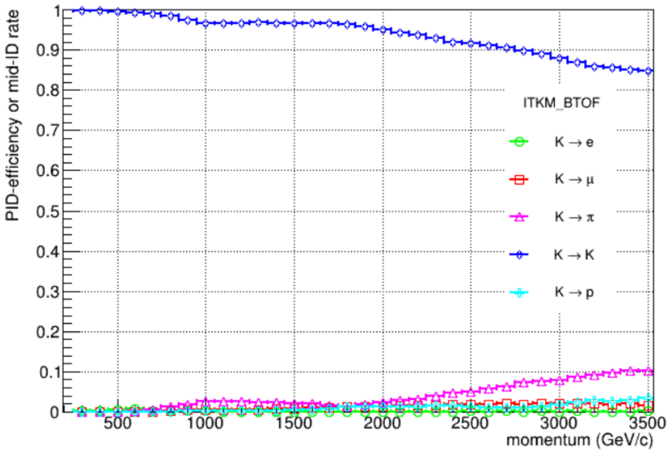
mu sample



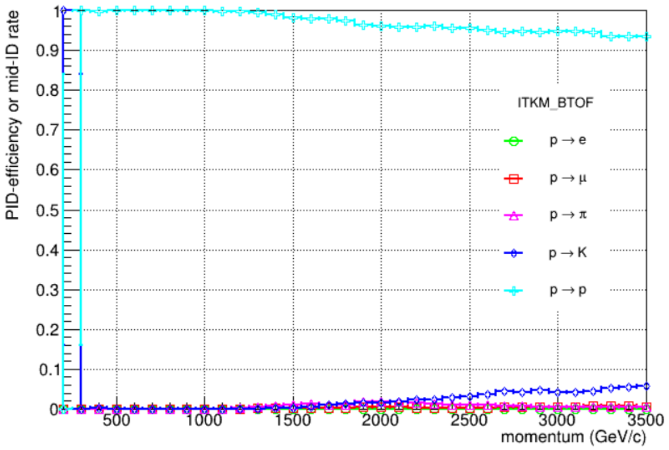
pi sample



K sample

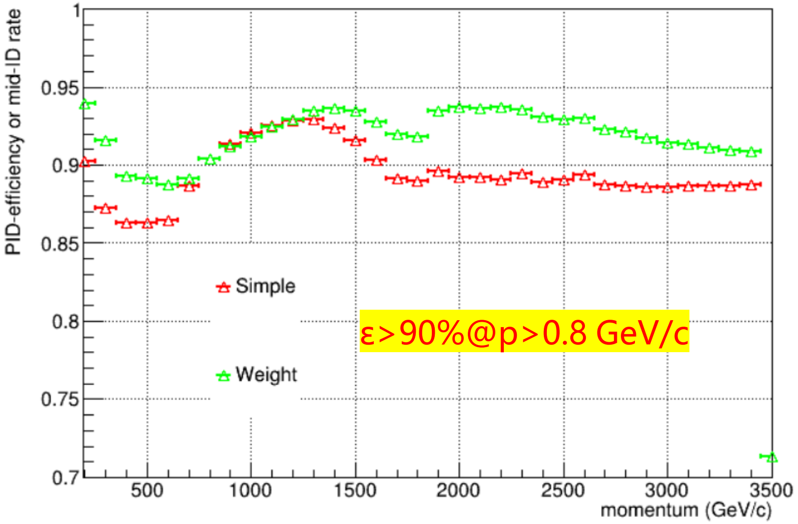


p sample

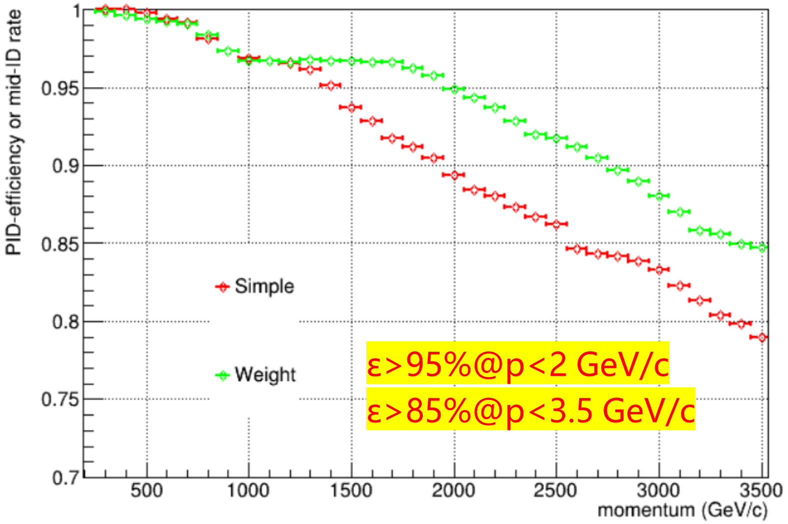


Weight likelihood 性能检验

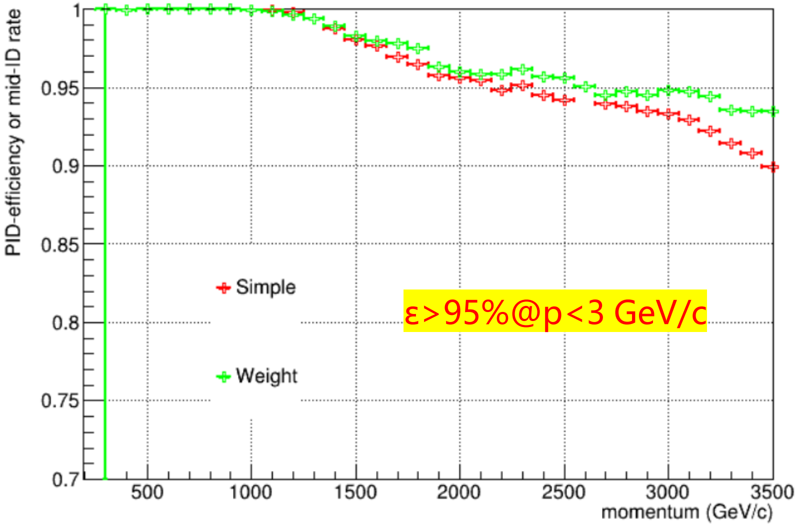
pi sample



K sample



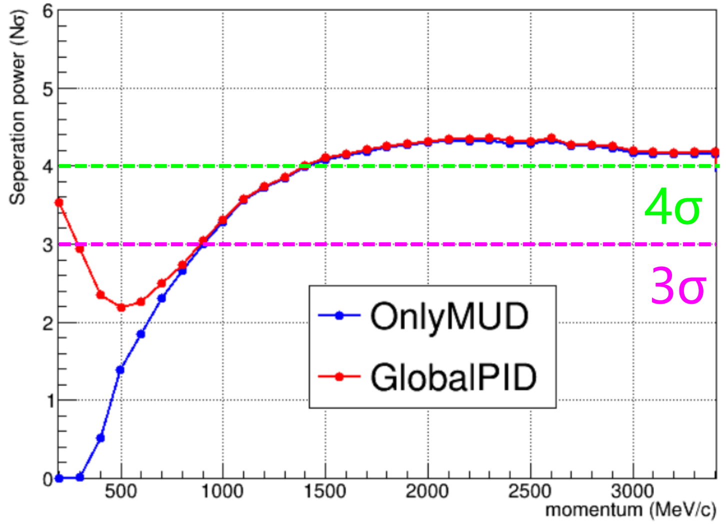
p sample



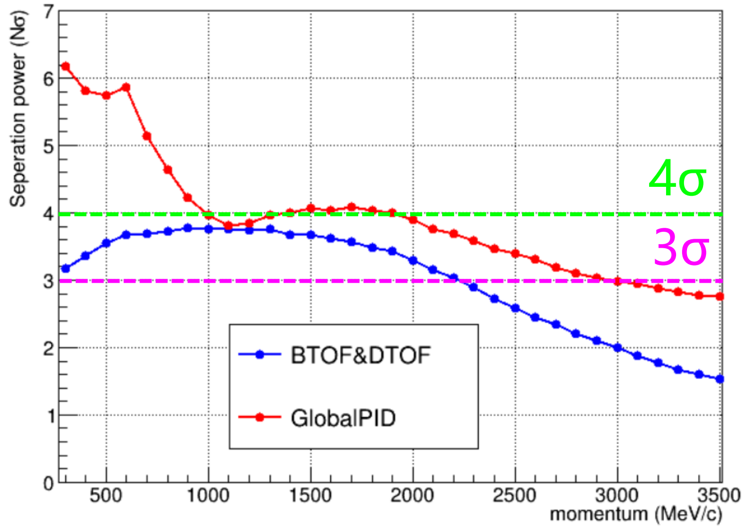
- 通过给likelihood增加权重，可以有效联合**dE/dx测量**与**PID探测器**信息，提升强子鉴别能力

粒子鉴别能力

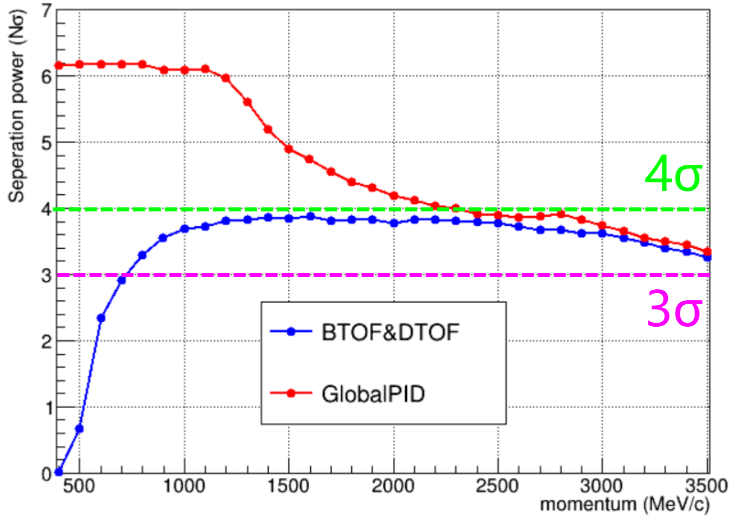
μ/π鉴别



π/K鉴别



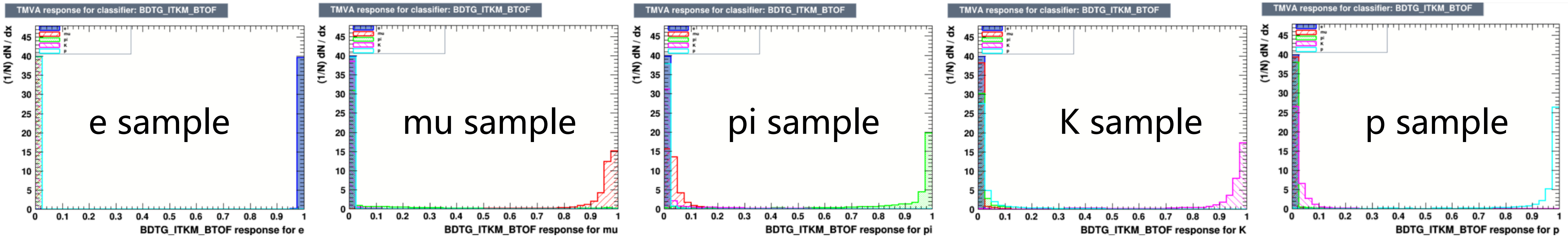
K/p鉴别



- **μ/π separation**: Nσ>3 @ p>0.8 GeV/c
- **π/K separation**: Nσ≈4 @ p=2 GeV/c; Nσ≈3 @ p=3 GeV/c
- **K/p separation**: Nσ≈4 @ p=2.8 GeV/c; Nσ>3 @ p<3.5 GeV/c

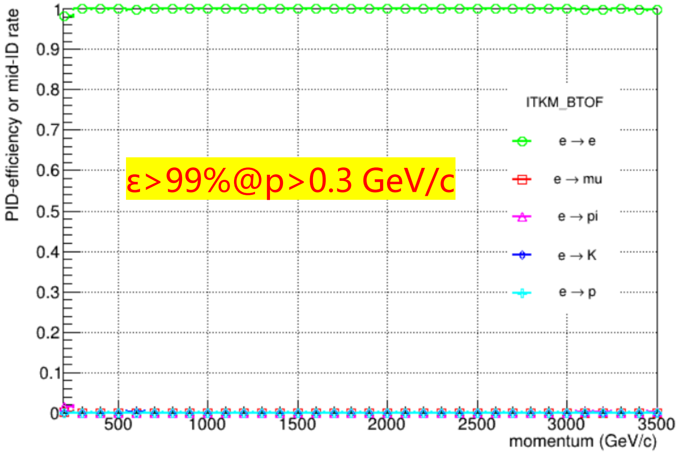
BDT训练

- 基于TMVAMultiClass的BDT训练，训练参数如下：
  - **重建径迹**：charge、momentum、theta、phi
  - **ECAL和MUD**：ECAL_e、MUD_mu
  - **ITK、MDC、PIDB、PIBE**：五种粒子假设下likelihood，归一化
- BDT参数设置：
  - NTrees=1500:BoostType=Grad:Shrinkage=0.05:UseBaggedBoost:BaggedSampleFraction=0.7:nCuts=100:MaxDepth=5

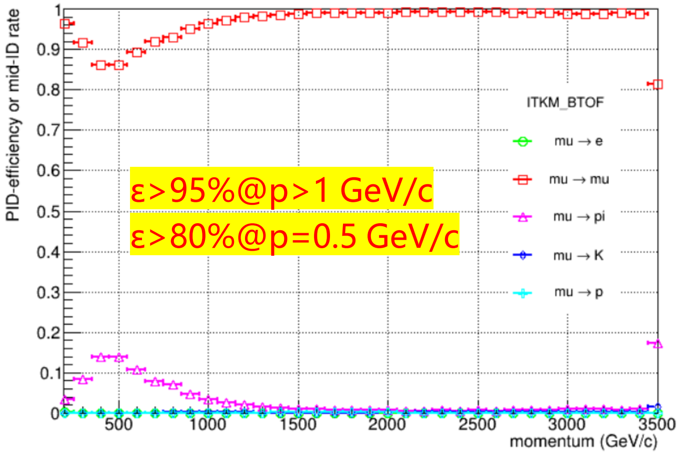


BDT训练结果

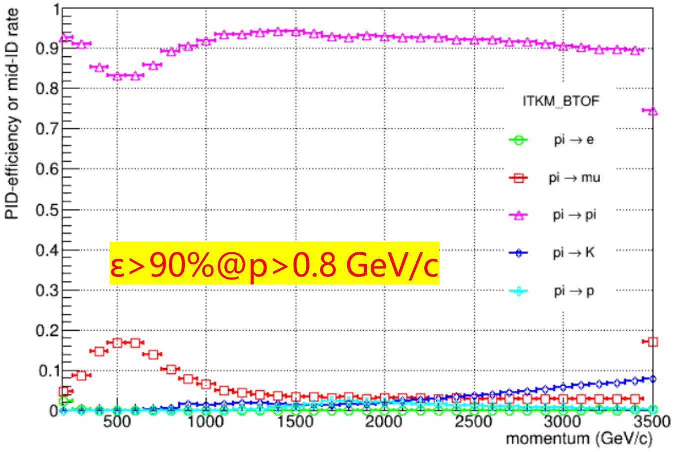
e sample



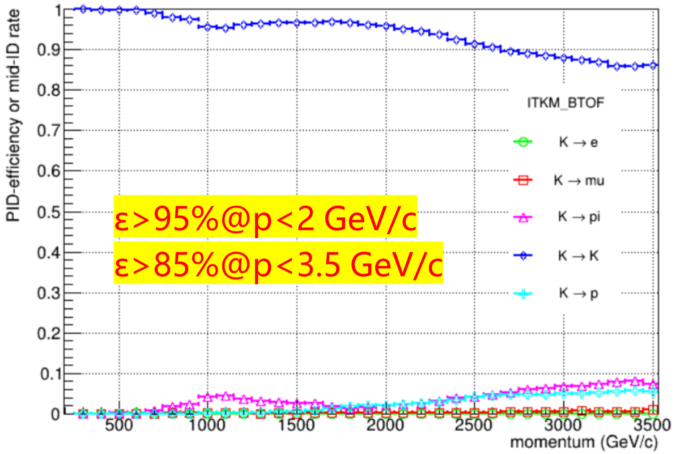
mu sample



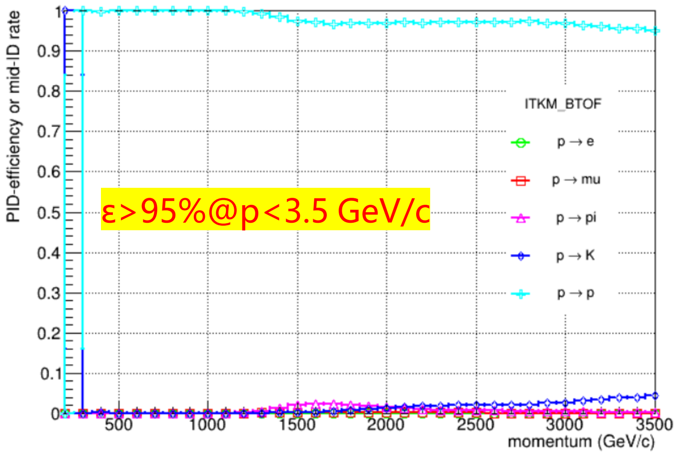
pi sample



K sample

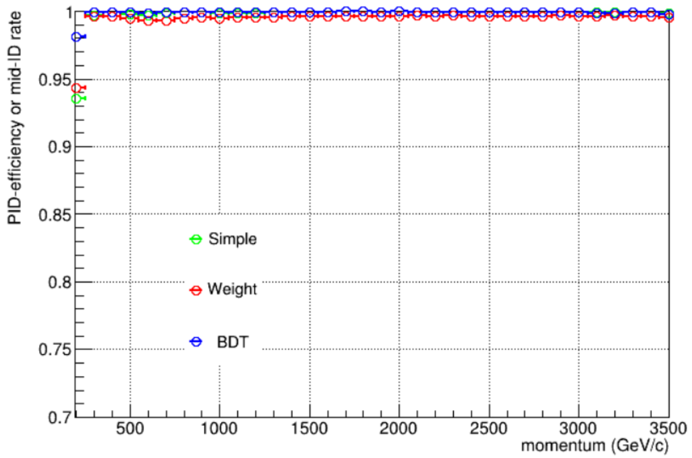


p sample

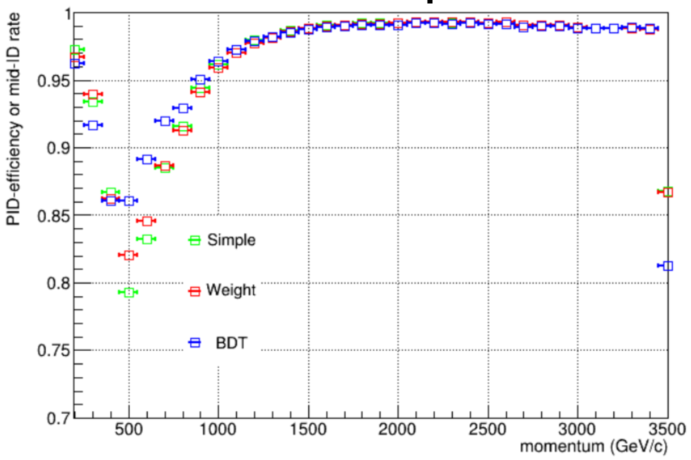


PID性能比较

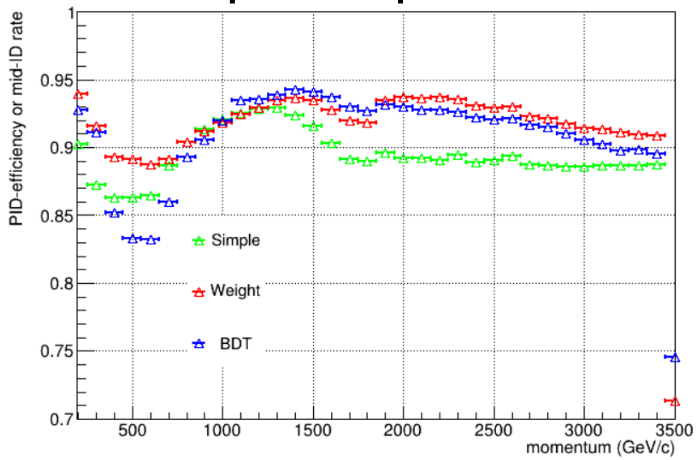
e sample



mu sample

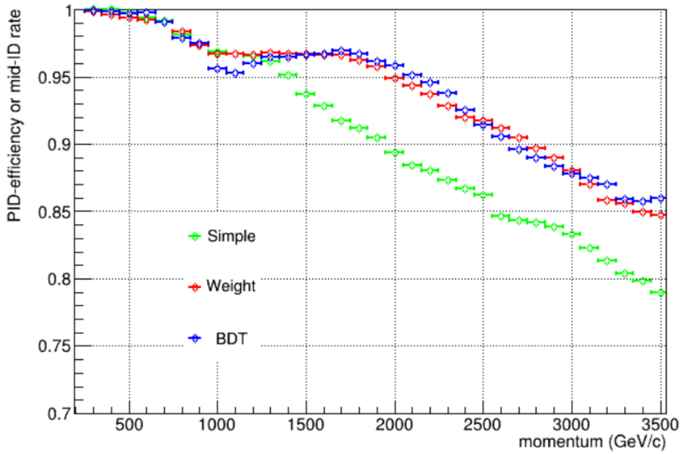


pi sample

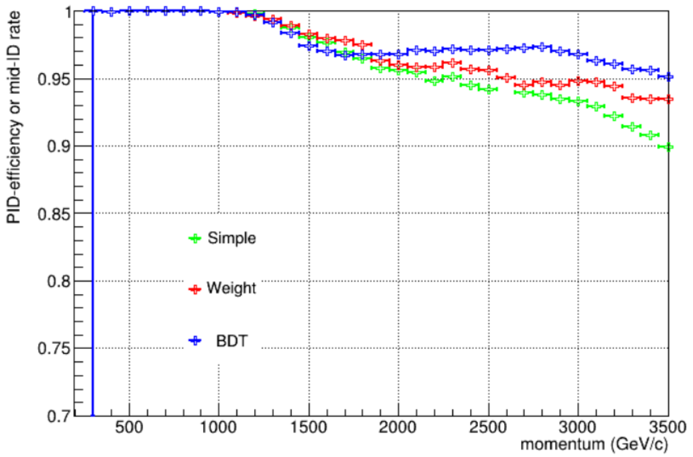


- BDT训练结果与weight likelihood结果基本一致

K sample



p sample



小结与展望

- 基于**多种方法，**联合各个探测器的PID信息，开发**全局五分类**粒子鉴别算法
  - Simple combined likelihood、combined likelihood with weight、BDT
- 基于**单粒子**样本，评估算法性能：
  - **e: ε>99%@p>0.3 GeV/c; mu: ε>95%@p>1 GeV/c, ε>80%@p=0.5 GeV/c**
  - **pi: ε>90% @p>0.8 GeV/c; K: ε>95% @p<2 GeV/c; p: ε>95%@p<3.5 GeV/c**

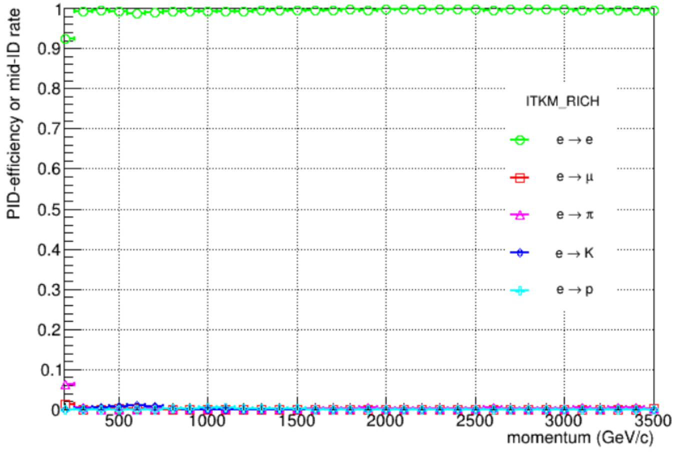
- **下一步计划**
  - BDT算法尝试**增加更多重建信息**，厘清当前算法与ML算法性能存在差异的原因
  - 基于**物理过程**选择控制样本，评估不同物理过程中的粒子鉴别性能

STCF global PID with ML · Zhai Yuncong

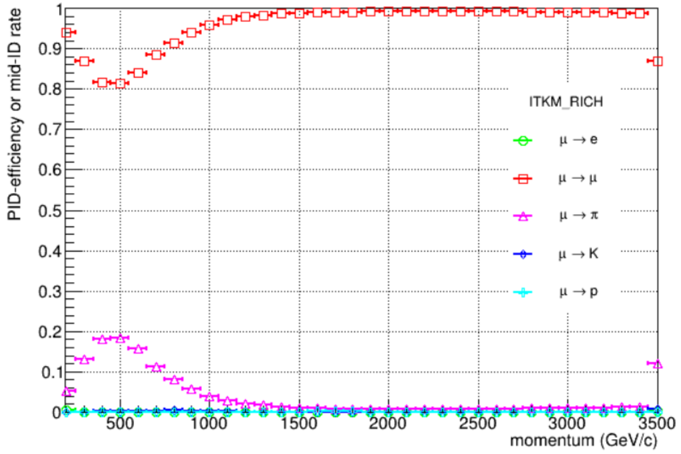
BACKUP

ITKM+RICH (weighted likelihood)

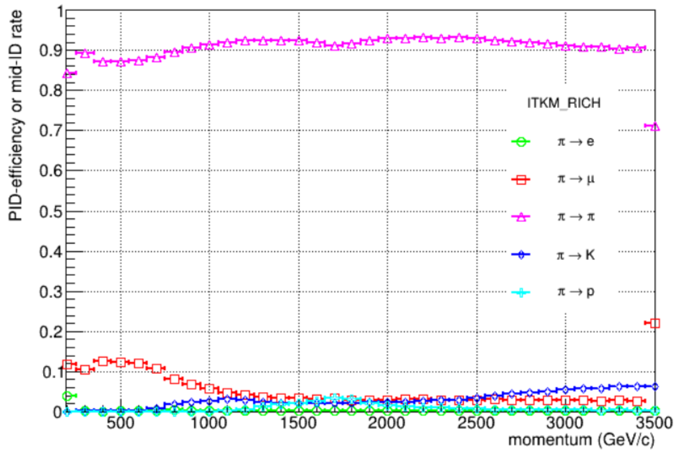
e sample



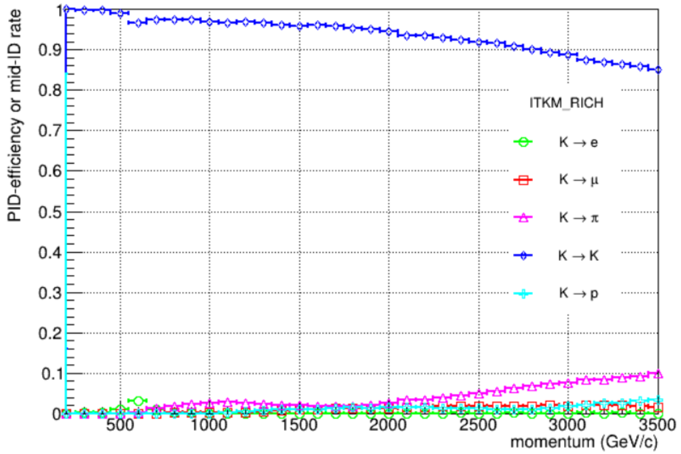
mu sample



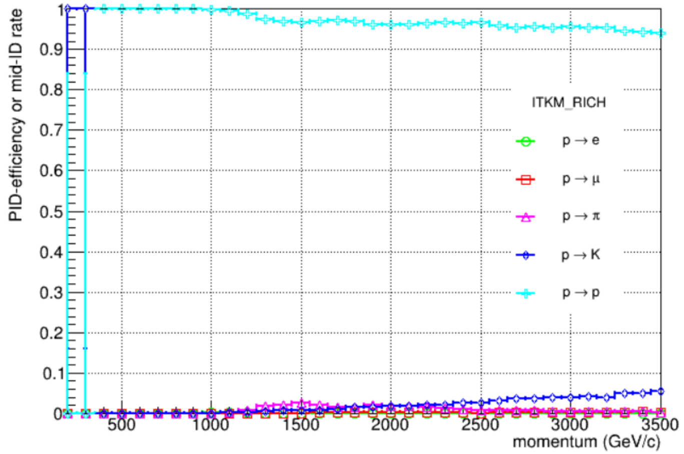
pi sample



K sample

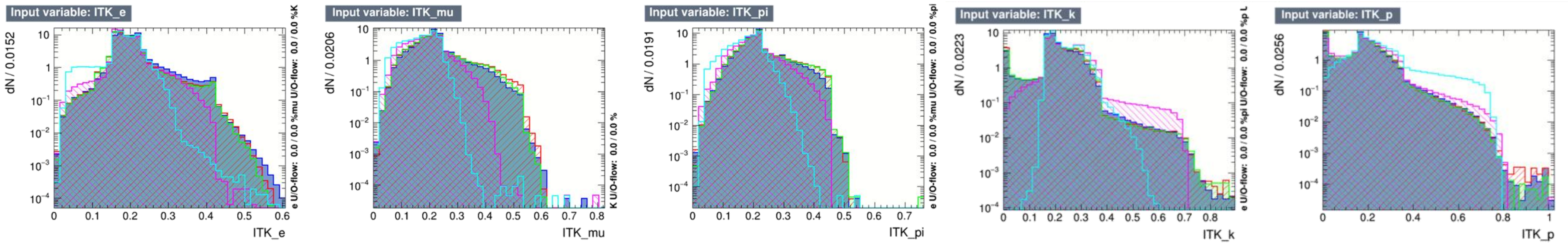


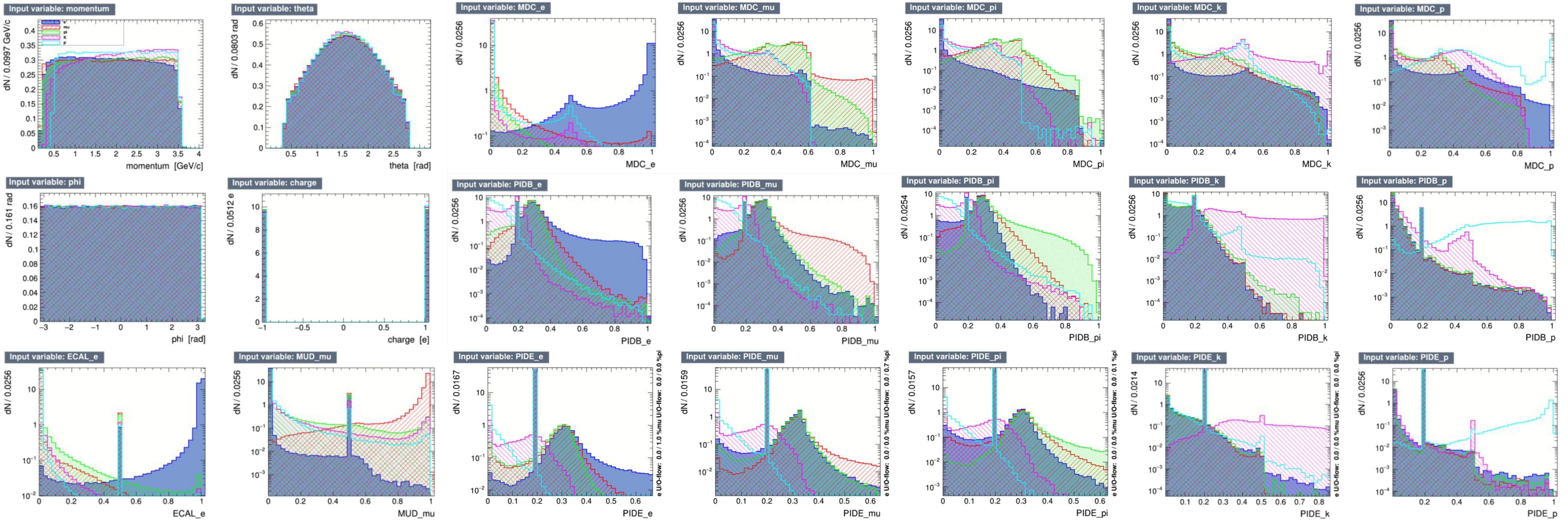
p sample



BDT训练参数分布

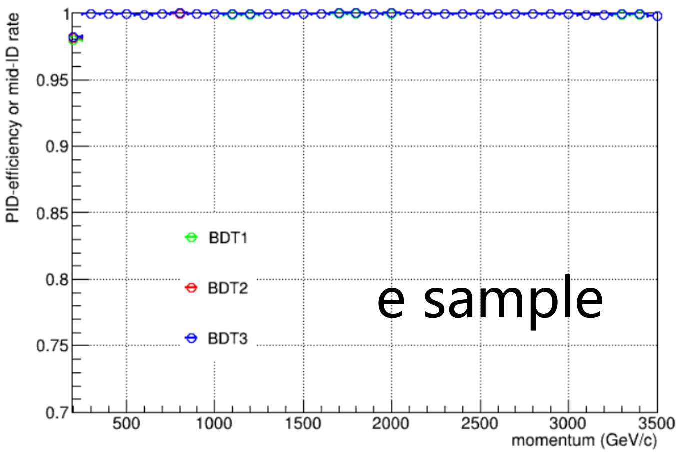
e/mu/pi/K/p

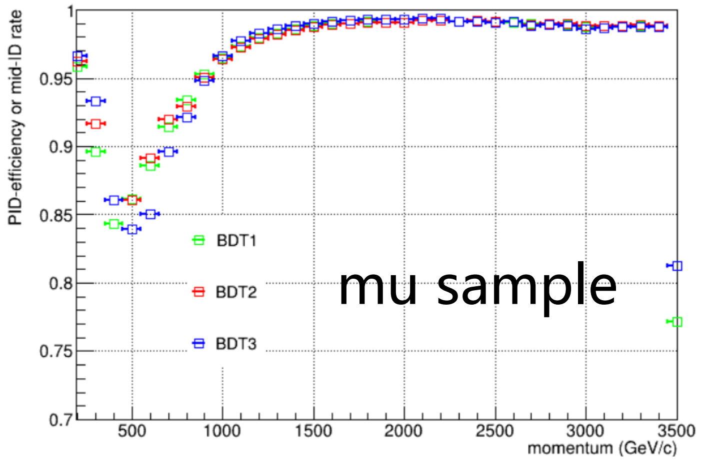


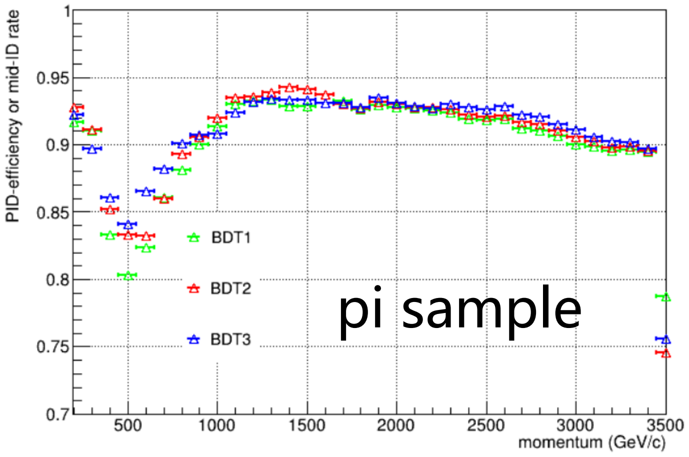


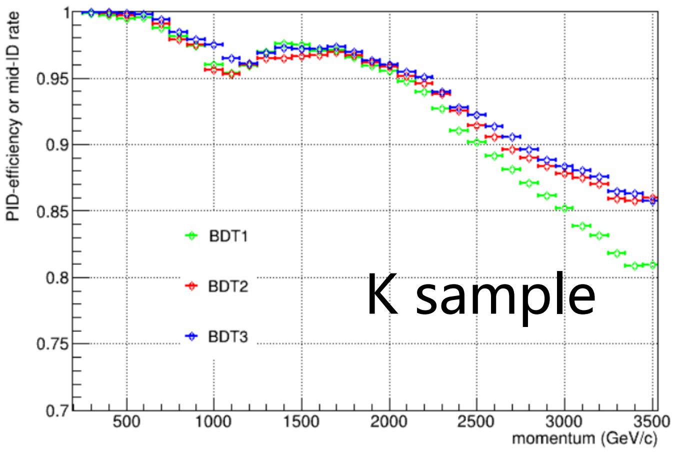
不同BDT参数设置训练结果比较

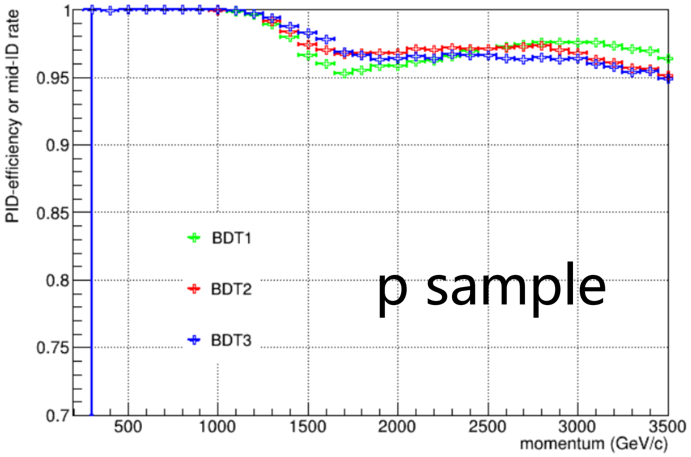
- **BDT1**：NTrees=**1000**:BoostType=Grad:Shrinkage=0.05:UseBaggedBoost:BaggedSampleFraction=0.7:nCuts=**50**:MaxDepth=**4**
- **BDT2**：NTrees=**1500**:BoostType=Grad:Shrinkage=0.05:UseBaggedBoost:BaggedSampleFraction=0.7:nCuts=**100**:MaxDepth=**5**
- **BDT3**：NTrees=**2000**:BoostType=Grad:Shrinkage=0.05:UseBaggedBoost:BaggedSampleFraction=0.7:nCuts=**100**:MaxDepth=**6**











BDT训练参数比较

GlobalPIDSvc

- **Track**
  - charge, MomFirst(x, y, z, theta, phi), D0, Z0, phi0, Kap, Tanlamda
- **MDC dE/dx**
  - Chis(e, mu, pi, k, p)
- **PIDB**
  - LogL(e, mu, pi, k, p)
- **PIDE**
  - LogL(e, mu, pi, k, p)
- **ECAL**
  - Nhits, Energy, Eseed, E3×3, E5×5, position(x, y, z), SecondMoment, A20Moment, A42Moment
- **MUD**
  - Theta, Phi, Nhits, RPCNhits, PSNhits , Maxhit, IsMuon , BDTExp, MaxhitLayer

BDT

BDT: more info. not shown here

GlobalWLPIDAlg

- **Track**
  - charge, MomPOCA(mag, theta, phi)
- **ITKM dE/dx**
  - P(e, mu, pi, k, p)
- **MDC dE/dx**
  - P(e, mu, pi, k, p)
- **PIDB**
  - P(e, mu, pi, k, p)
- **PIDE**
  - P(e, mu, pi, k, p)
- **ECAL**
  - Pe
- **MUD**
  - Pmu