

物理分析II

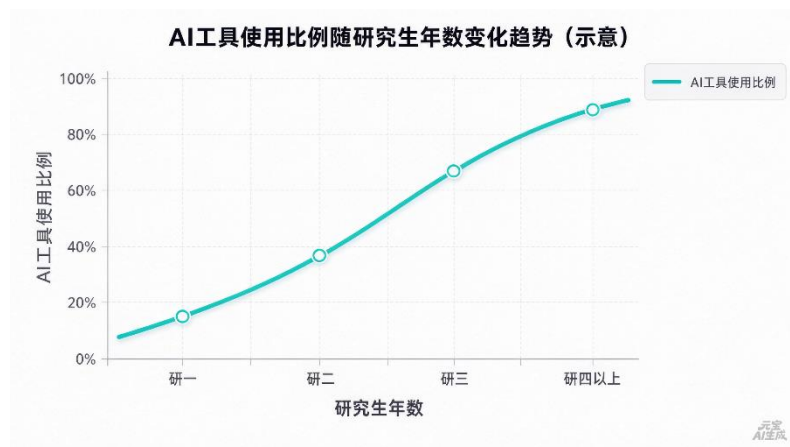
高能物理研究所 朱凯

zhuk@ihep.ac.cn

(第八届) 北京谱仪十一科学研讨会
山东理工大学 淄博 2026年10月1日-10月8日

在人工智能时代，还需要“手搓”物理分析吗？

- 在分析结果可以“一键生成”的时代，对分析者的要求不是变低，而是变高了
 - 研究生毕业，学了几个“cut”不够
 - 除了知道怎么做，还需要理解它的底层逻辑
 - 不仅知道怎么做，还需要知道做什么
- 实验物理研究生该如何培养？



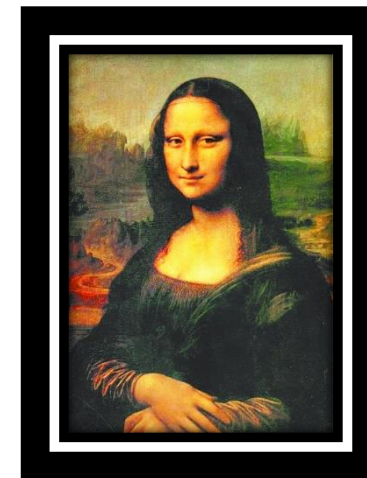
全新一页

主要内容与目标

- 事例选择与本底分析
 - 目标：掌握事例选择的三大原则
- 信号数
 - 目标：了解三种常用提取信号数的方法
- 效率
 - 目标：熟悉与效率相关的三个问题



本次课



将来

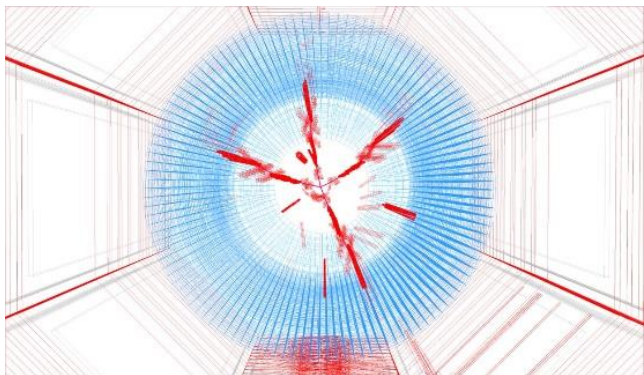
以李刚老师在BESIII上的工作 $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$ 为例

过渡

- 接着赵明刚老师的课谈谈分支比测量
 1. 测量 $B(\psi(2S) \rightarrow \pi^+\pi^-J/\psi)$, $B(J/\psi \rightarrow l^+l^-)$ 时, 过程的末态是什么?
 2. 为什么要测量分支比, 即物理动机是什么?
 3. 分支比的测量实质上测量的什么? 但又转换成了什么的测量? 分支比的计算公式?

末态

- BESIII 上能探测到的末态粒子
 - 中性：光子 (γ)、正反中子 (n and $\bar{n}$)
 - 带电粒子：正反电子 ($e^\pm$)、正反缪子 ($\mu^\pm$)、正负 π 介子 ($\pi^\pm$)，正负K介子 ($K^\pm$)、正反质子 (p and $\bar{p}$)
- $\psi(2S) \rightarrow \pi^+\pi^-J/\psi, J/\psi \rightarrow anything$ 【半单举 (semi-inclusive)】
 - $\pi^+\pi^- + anything$
- $\psi(2S) \rightarrow \pi^+\pi^-J/\psi, J/\psi \rightarrow l^+l^-$ 【遍举 (exclusive)】
 - $\pi^+\pi^-e^+e^-$ 或者 $\pi^+\pi^-\mu^+\mu^-$



术语园地

inclusive MC, generic MC
exclusive MC, signal MC

过渡

- 接着赵明刚老师的课以分支比测量为例
 1. 测量 $B(\psi(2S) \rightarrow \pi^+ \pi^- J/\psi)$, $B(J/\psi \rightarrow l^+ l^-)$ 时, 过程的末态是什么?
 2. 为什么要测量分支比, 即物理动机是什么?
 3. 分支比的测量实质上测量的什么? 但又转换成了什么的测量? 分支比的计算公式?

测量 $B(\psi(2S) \rightarrow \pi^+ \pi^- J/\psi)$, $B(J/\psi \rightarrow l^+ l^-)$ 的动机

- 实验上
 - 其他测量的输入，例如：
 - 暗光子的寻找 $J/\psi \rightarrow \gamma A$
 - LHCb 上 J/ψ 相关测量
- 理论上
 - 末态相互作用（FSI: final state interaction）
 - QCD 多极展开、QCD 手征对称性、势模型
 - 轻子普适性
 -

衔接

- 接着赵明刚老师的课以分支比测量为例
 1. 测量 $B(\psi(2S) \rightarrow \pi^+\pi^-J/\psi)$, $B(J/\psi \rightarrow l^+l^-)$ 时, 过程的末态是什么?
 2. 为什么要测量分支比, 即物理动机是什么?
 3. 分支比的测量实质上测量的什么? 但又转换成了什么的测量? 分支比的计算公式?

分支比测量的实质是反应截面的测量

- 转换为信号事例数的测量，因为 $N_{sig} = \sigma \times L \times \epsilon$ （截面x亮度x效率）

$$B_{\pi\pi J/\psi} = \frac{\sigma_{\pi\pi J/\psi}}{\sigma_{tot}} = \frac{\frac{N_{\pi\pi J/\psi}}{L \times \epsilon_{\pi\pi J/\psi}}}{\frac{N_{tot}}{L}}$$



$$B_{\pi\pi J/\psi} = \frac{N_{\pi\pi J/\psi}}{\epsilon_{\pi\pi J/\psi} \times N_{tot}}$$

$$\begin{aligned} \mathcal{B}_{ll} &= \frac{\mathcal{B}_{\pi\pi J/\psi} \times \mathcal{B}_{ll}}{\mathcal{B}_{\pi\pi J/\psi}} \\ &= \frac{N_{ll}/(\epsilon_{ll} \times N_{tot})}{N_{\pi\pi J/\psi}/(\epsilon_{\pi\pi J/\psi} \times N_{tot})} \\ &= \frac{N_{ll}/\epsilon_{ll}}{N_{\pi\pi J/\psi}/\epsilon_{\pi\pi J/\psi}} \end{aligned}$$

分支比测量的通用公式

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$

- B: 待测分支比，无量纲
- N^{sig} : 信号事例数
- N^X : 总事例数
- ϵ : 事例选择效率，通常由蒙卡模拟确定，需特别关注

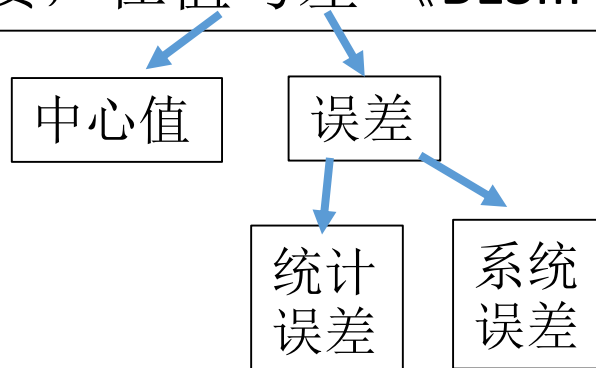
接下来干嘛？当然是挑选信号事例！



事例选择

目的：提高信噪比，降低统计误差

国之大事，在祀与戎 《左传 成公十三年》
分析之要，在值与差 《BESIII 十一研讨会》



事例选择 ABC

- 事例选择益简、松、少
 - 有利于减少系统误差
- 利用统计原理减少误差
 - 相对测量，抵消（系统）误差
 - 同时测量，约束（统计）误差
- 利用运动学、动力学特征
 - 动量分布
 - 角分布
 - 中间窄共振

好的选择应该考虑到系统误差

例一

- Track level cuts for soft pion candidates

- $|V_z| < 10\text{cm}$

- $|V_r| < 1\text{cm}$

- $|\cos\theta| < 0.80$

- $|\vec{p}| < 0.45\text{GeV}/c$

- all charged tracks satisfy above selection criteria are assumed to be pion.

$$\psi' \rightarrow \pi^+ \pi^- J/\psi$$

- $\text{nGood}+ \geq 1 \ \&\& \ \text{nGood}- \geq 1$

- $\cos\theta_{\pi^+\pi^-} < 0.95$ to remove the photon conversion

- $3.05\text{GeV}/c^2 \leq M_{\pi^+\pi^-}^{\text{rec.}} \leq 3.15\text{GeV}/c^2$

assuming all charge tracks are pions, $M_{\pi^+\pi^-}^{\text{rec.}}$ is defined as:

$$M_{\pi^+\pi^-}^{\text{rec.}} = \sqrt{(p_{\text{ecm}} - p_{\pi^+} - p_{\pi^-})^2},$$

where p is 4 momentum of pion.

术语园地

soft, slow, low p

hard, fast, high p

长短相形，高下相倾 《道德经》

术语园地

soft, slow, low p
hard, fast, high p



软硬、快慢、高低相对
物理是一门精确定量的科学

“多”与“少”、“大”与“小”等
表达仅适用于不会产生误解的场合，
所以基本用不上或需要补语

硬度	矿物	替代物品
1	滑石(Talc)	
2	石膏(Gypsum)	指甲 2.5,
3	方解石(Calcite)	
4	萤石(Fluorite)	铜币 3.5-4
5	磷灰石(Apatite)	钢刀 5.5,
6	正长石(Orthoclase)	玻璃 5.5 -6,
7	石英(Quartz)	
8	黄玉(Topaz)	钢锉 6.5
9	刚玉(Corundum)	
10	金刚石(Diamond)	

关于挑选 π 的更多解释

$$\psi' \rightarrow \pi^+ \pi^- J/\psi$$

- 这是一个三体衰变，母粒子和子粒子的质量等非常清楚，那么末态粒子的动量范围就很容易知道：[运动学知识](#)。
- PDG 上也可以查到
 - 末态 π 动量不超过401 MeV，考虑了分辨后 450 MeV 依然非常安全：无需考虑这个 cut 的系统误差；
 - π 的动量不高，故其动量的绝对误差也就很小；
 - $\pi^+ \pi^-$ 的反冲侧是非常窄的 J/ψ ，可以对 $\pi^+ \pi^-$ 系统进行很好的约束，故无需粒子鉴别。

术语园地
 $\psi(2S)$ 、 ψ' 、 $\psi(3686)$

粒子物理学家有乱起名字的传统

- 英文字母与希腊字母混用
 - 轻子: $e \mu \tau$
 - 强子: $\pi K p \Lambda$
 - 粲偶素: $\eta_c, J/\psi, \chi_{cJ}, h_c, \psi(3686), \psi(3770)$
- 打补丁
 - 升级: $\eta \rightarrow \eta', J/\psi \rightarrow \psi'$
 - 高级战士: $f_0(980), f_1(1285), f_2(1270)$
 - 重装战士: $T_{c\bar{c}1}(3900), T_{c\bar{c}s1}(4000)$
- 那些消失的名字: D 介子与 $E(D^*), F(D_s)$
- 名字的变化反映了认识的进步: $\psi', \psi(2S), \psi(3686)$

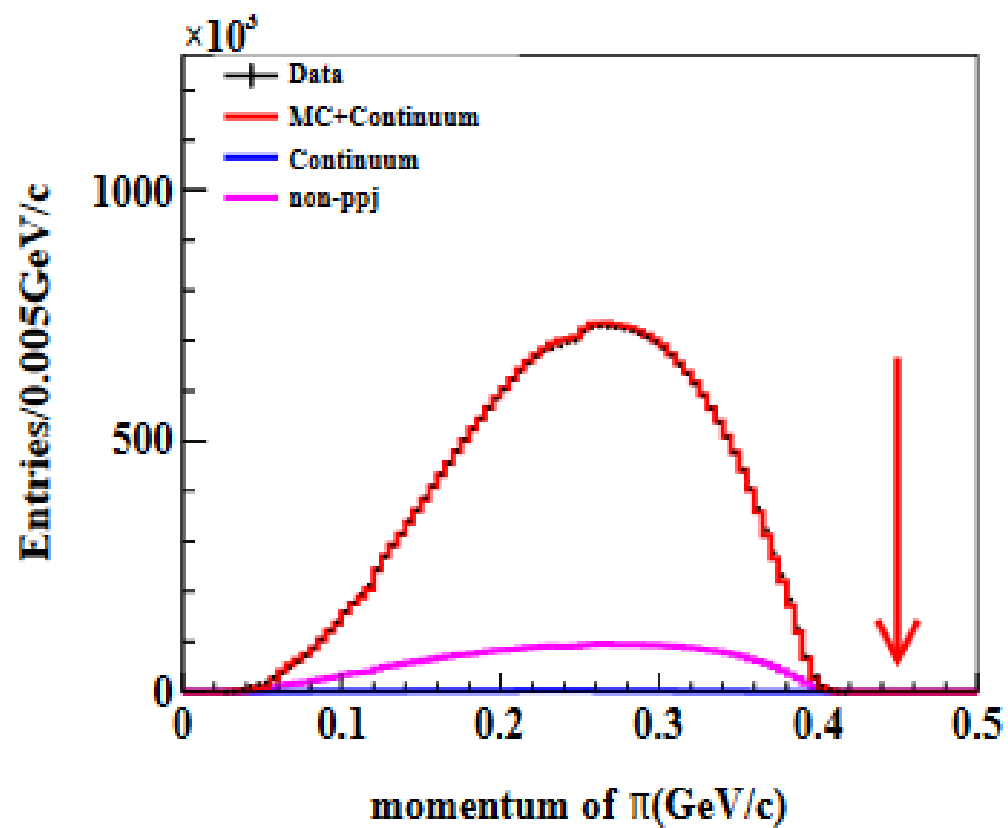
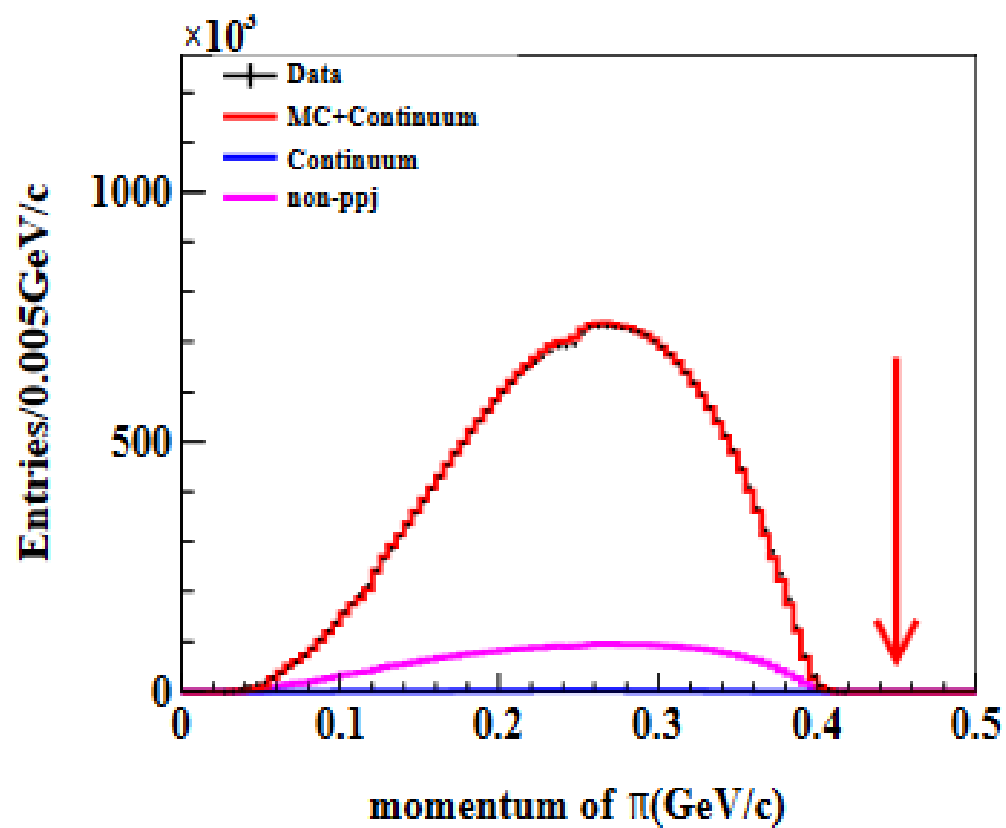


FIG. 2. The momenta of π^+ (left) and π^- (right)

$$p_{\pi} < 0.45 \text{ GeV}$$

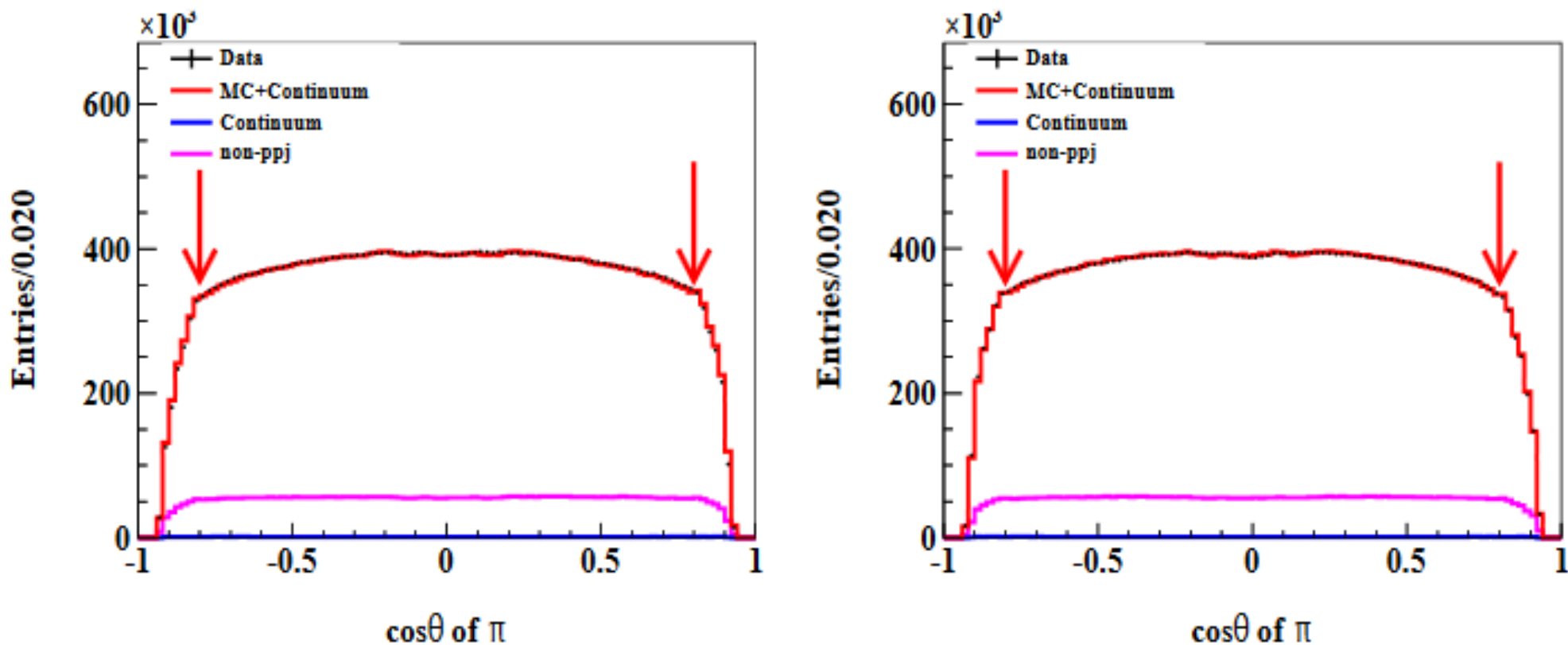
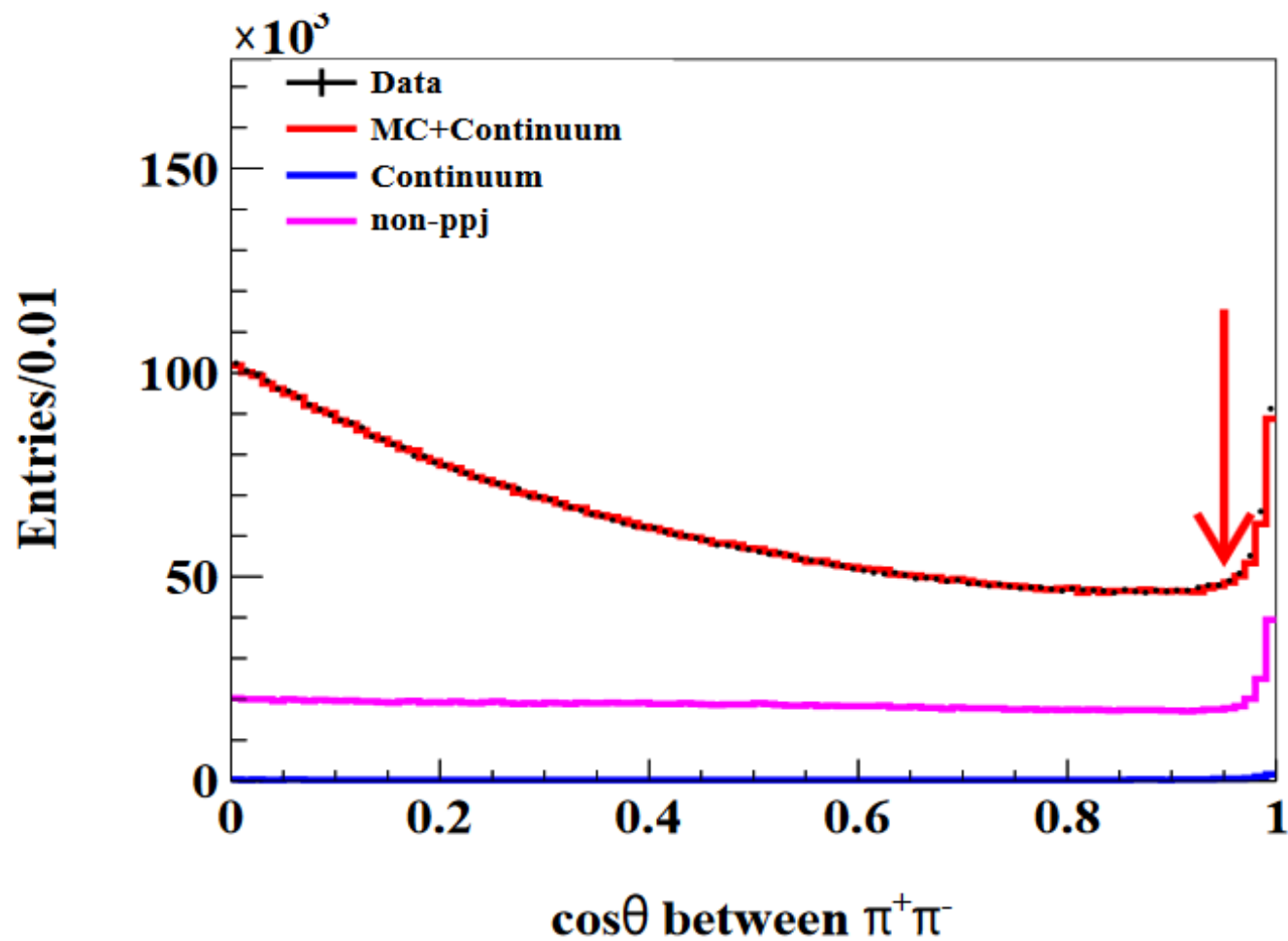


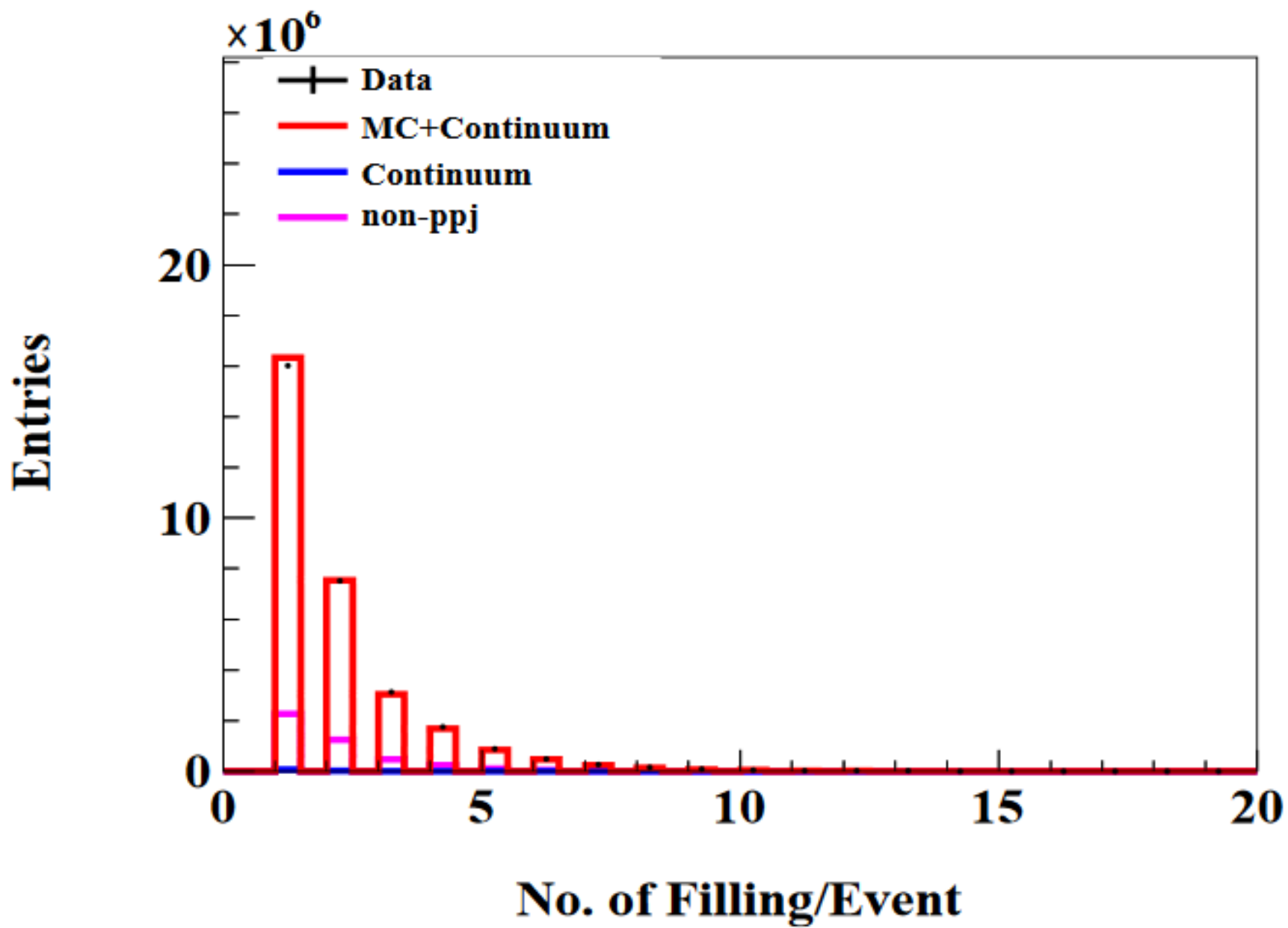
FIG. 3. The cosine of polar angle of π^+ (left) and π^- (right) in lab frame

$\cos(\theta) < 0.8$ 更好的径迹质量? 更可靠的径迹效率?



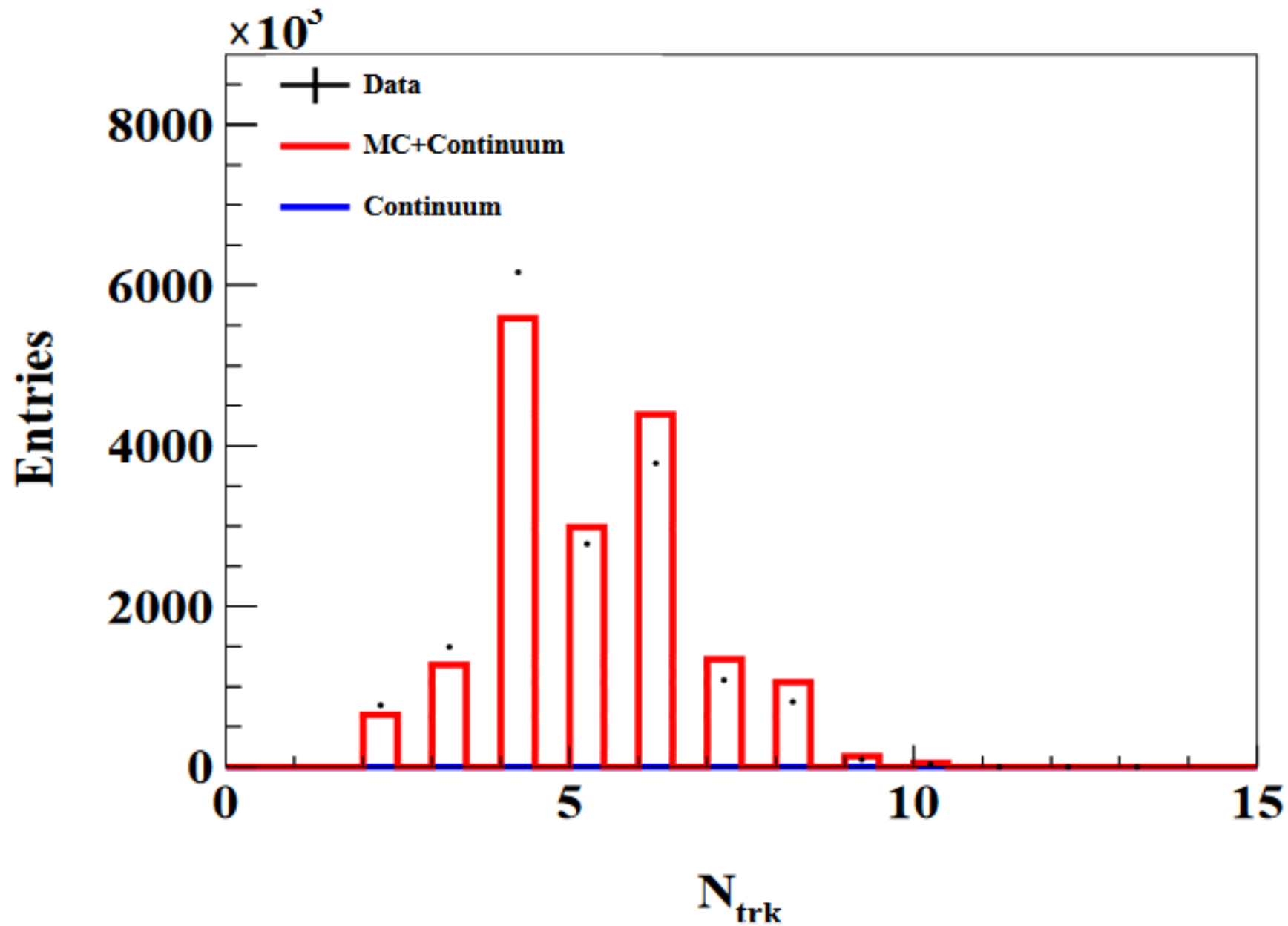
$\cos\theta_{\pi^+\pi^-} < 0.95$
去除来自 γ conversion 的本底

FIG. 4. Cosine distributions of the angle between two pions in $\pi^+\pi^-J/\psi$



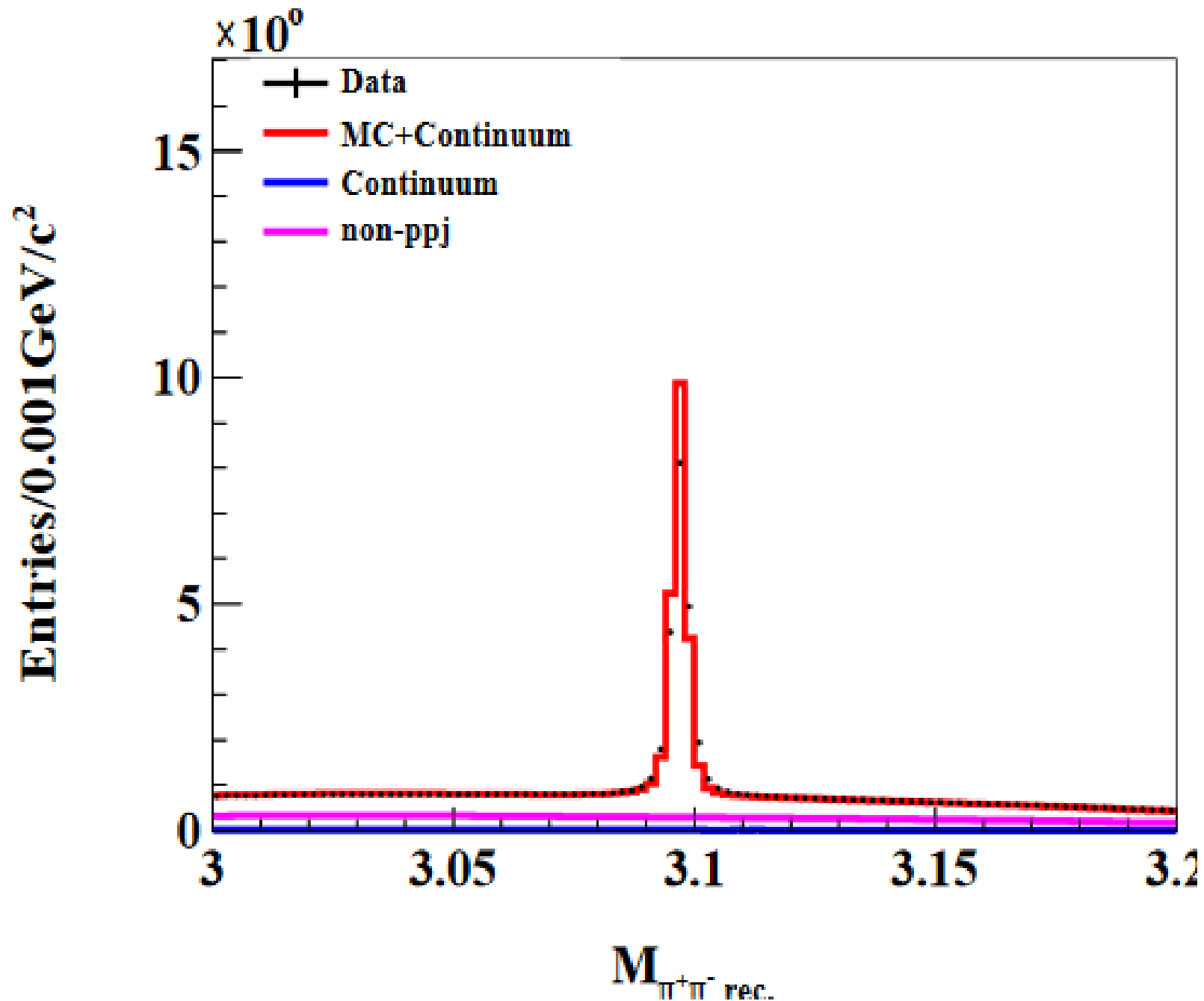
术语园地
event 事例
entry 记录

FIG. 5. The number of filling time of each event



术语园地
multiplicity,
prong

FIG. 6. The number of good tracks of each event

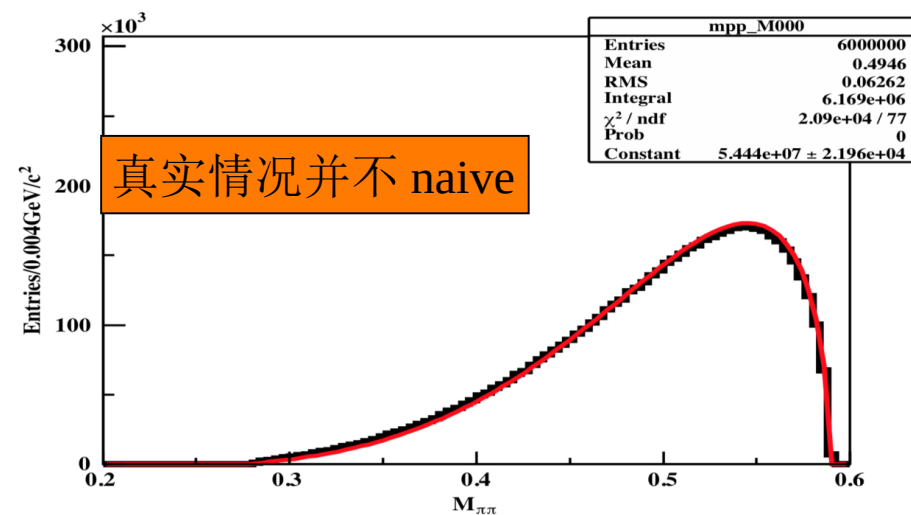
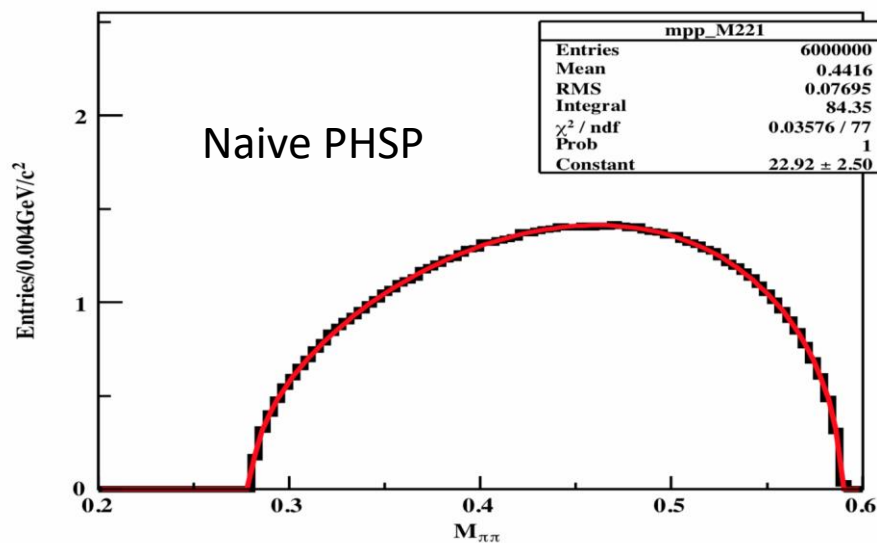


术语园地
信号 signal (SIG)
本底 backgrounds (BG)

- 在物理分析中，一船用信号MC来代表信号的分布；数据里常常是信号与本底同时存在
- 为了方便，常常（不严谨地）将信号中的典型“峰”状分布称为信号

$M_{\pi\pi}$ spectrum

BesEvtGen: JPIPI model



两 π 倾向于背对背, $M_{\pi\pi}$ 对效率有显著影响。

$$\frac{d\sigma}{dm_{\pi\pi}} \propto (s - 2m_{\pi\pi}^2)$$

例二

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

$$J/\psi \rightarrow l^+ l^-$$

- Track level cuts
 - $|V_z| < 10\text{cm}$
 - $|V_r| < 1\text{cm}$
 - $|\cos \theta| < 0.80$
 - $|\vec{p}| < 2.0\text{GeV}/c$
- $n_{\text{Good}} \geq 4$
- The charged tracks with $|\vec{p}| < 0.45 \text{ GeV}/c$ are assumed to be pion, select the $\pi^+ \pi^-$ pair candidates by minimizing $|M_{\pi^+ \pi^-}^{\text{rec.}} - M_{J/\psi}|$
- $\cos \theta_{\pi^+ \pi^-} < 0.95$
- $3.05\text{GeV}/c^2 \leq M_{\pi^+ \pi^-}^{\text{rec.}} \leq 3.15\text{GeV}/c^2$
- Take the two fastest positive and negative tracks as lepton candidates, identify the e/μ pair
 - $\mu^+ \mu^-: [E/p]^+ < 0.26$ and $[E/p]^- < 0.26$
 - $e^+ e^-: [E/p]^+ > 0.80$ or $[E/p]^- > 0.80$ or $\sqrt{([E/p]^+ - 1)^2 + ([E/p]^- - 1)^2} < 0.4$
- $\cos \theta_{l+l-} < -0.95$ in lab frame.
- $2.7\text{GeV}/c^2 < m_{l+l-} < 3.2\text{GeV}/c^2$ for $\pi^+ \pi^- e^+ e^-$ channel and $3.0\text{GeV}/c^2 < m_{l+l-} < 3.2\text{GeV}/c^2$ for $\pi^+ \pi^- \mu^+ \mu^-$ channel

轻子选择的一些说明

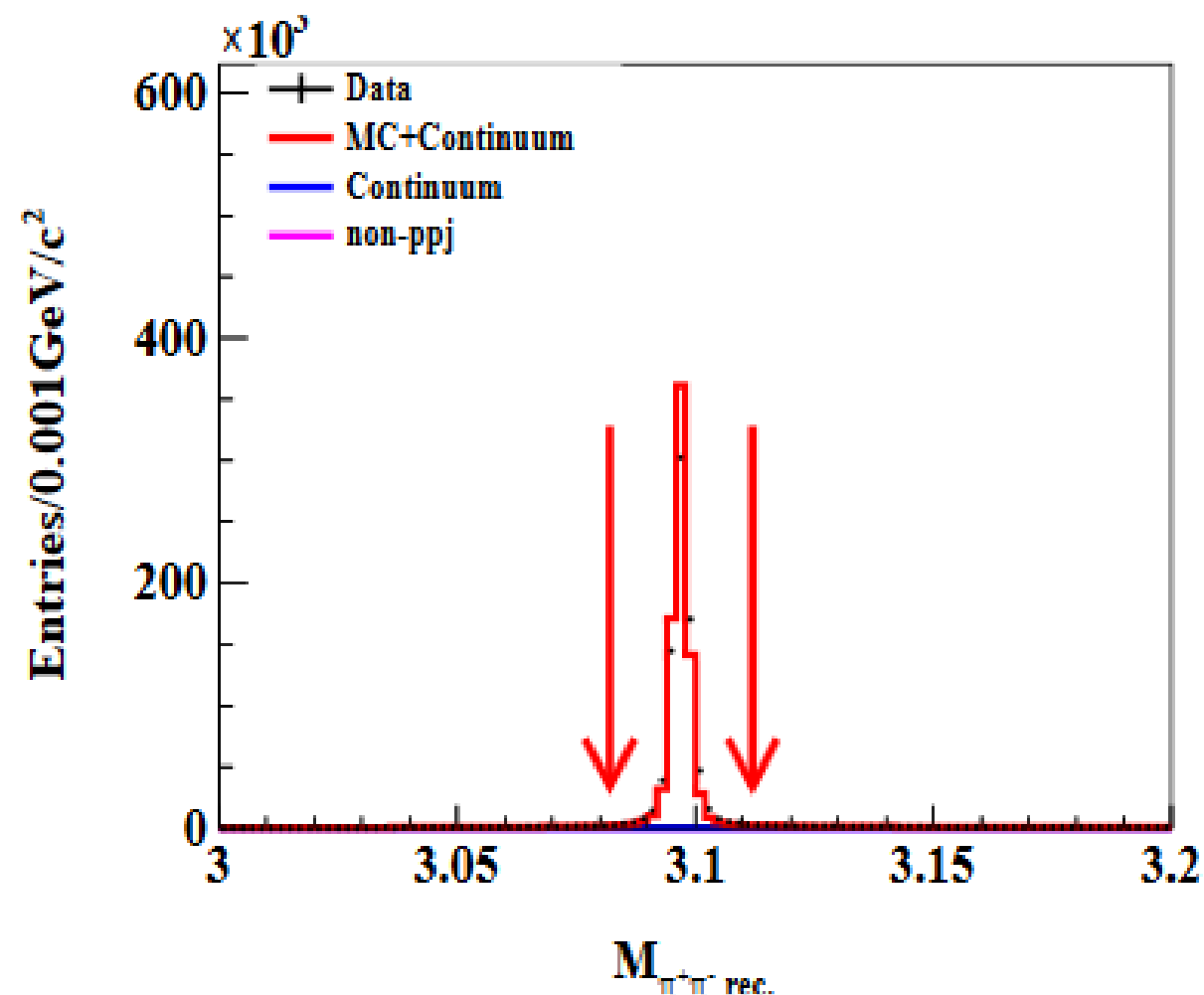
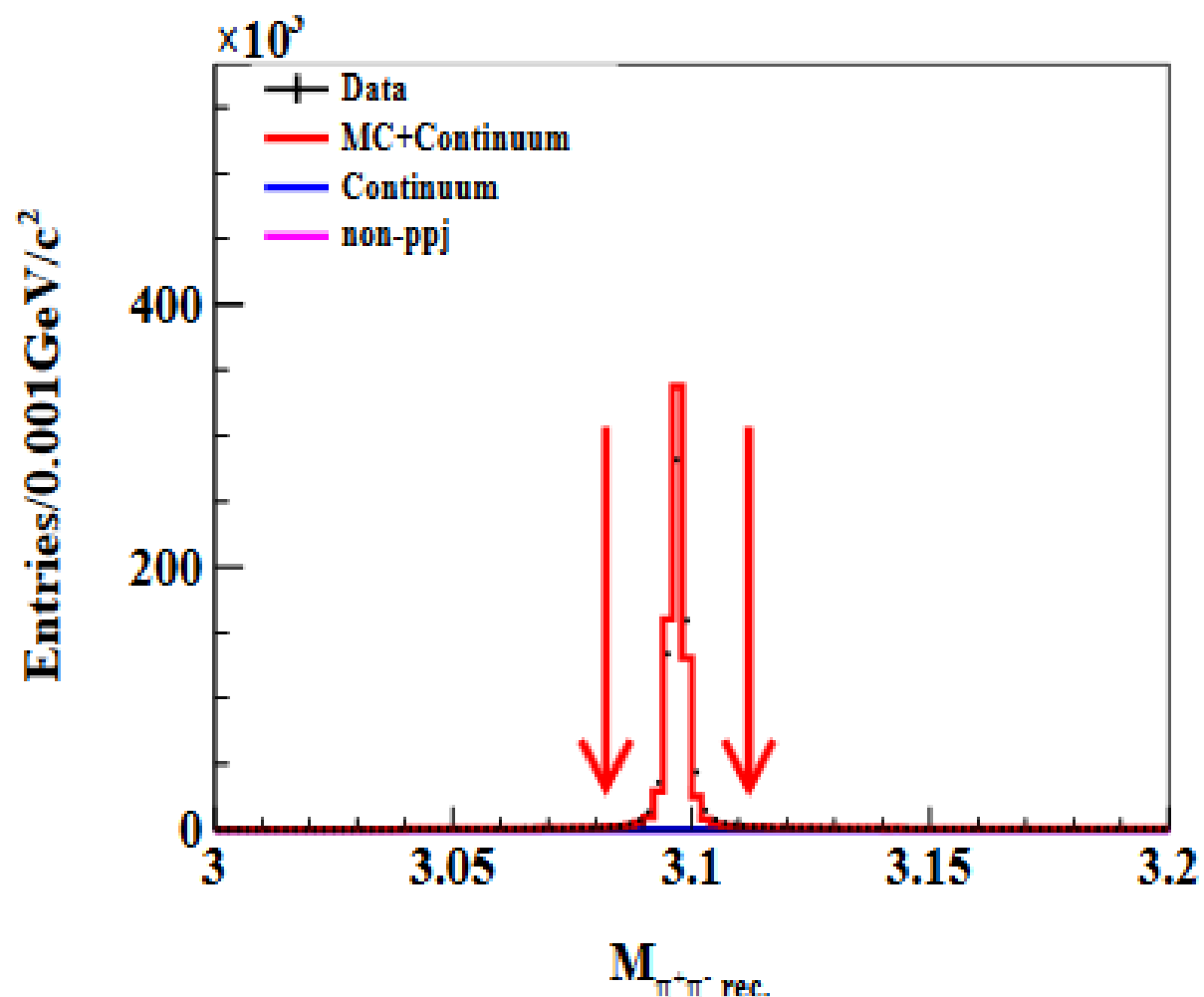
- 两体衰变，在母粒子质心系是单（高）能的：大约1.55 GeV
- 即使 J/ψ 运动导致一定的动量展宽，仍然和 π 的动量区间完全无重叠：无需粒子鉴别，只需区分 e、 μ 即可
- 因此这个分析无需用到 dE/dx 和 TOF 的 PID，也就不需要分析这项系统误差

术语园地

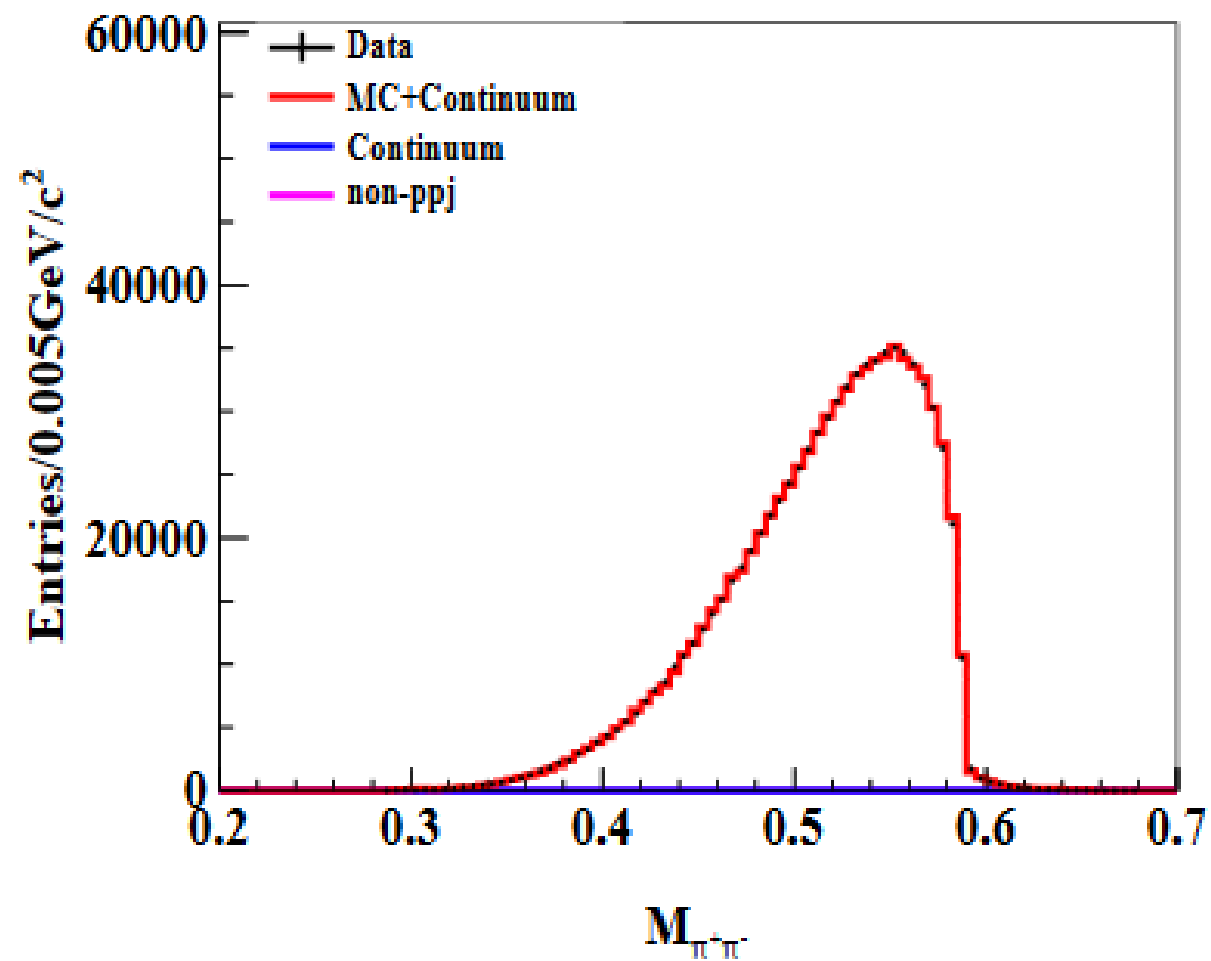
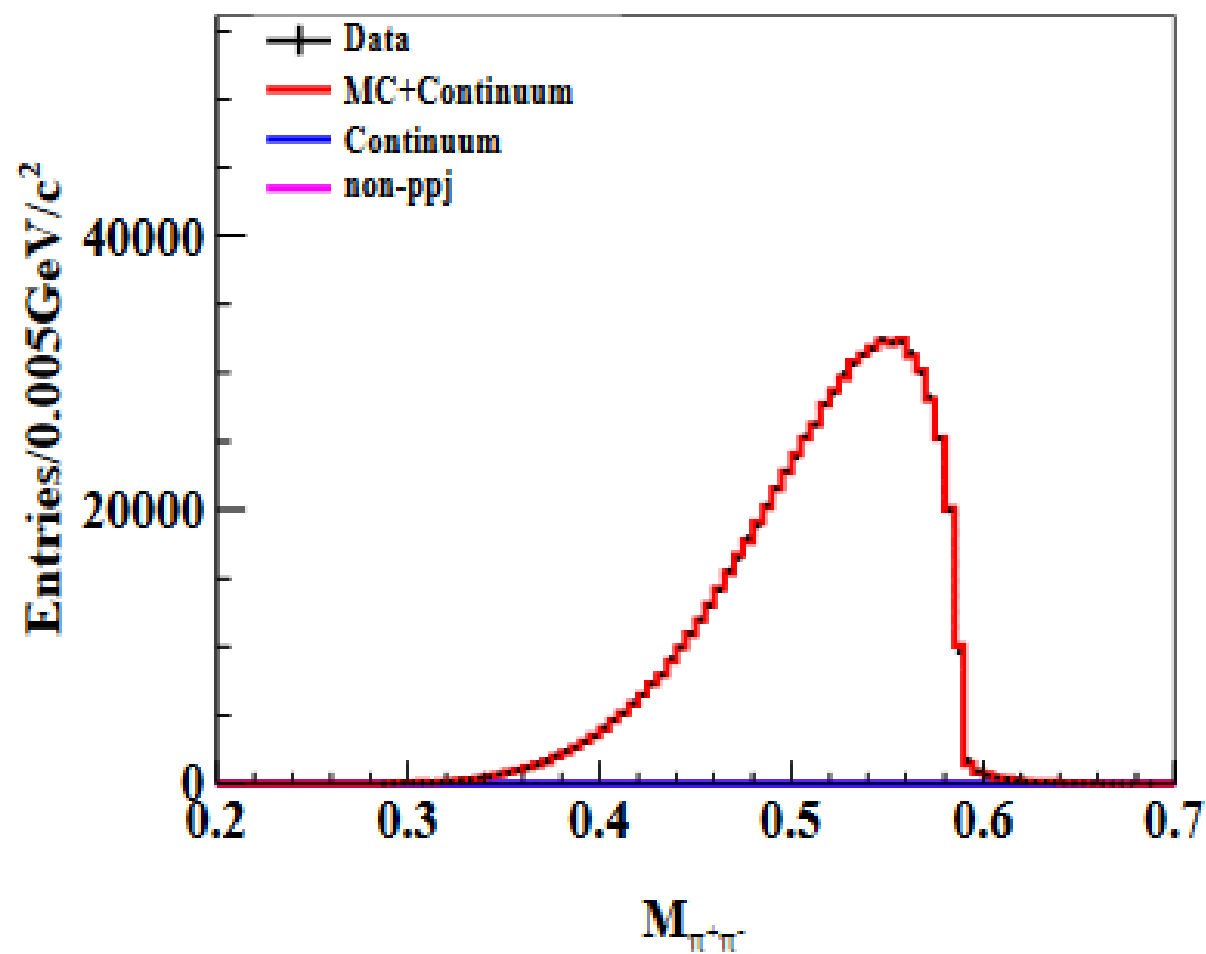
dE/dx 单位距离能损

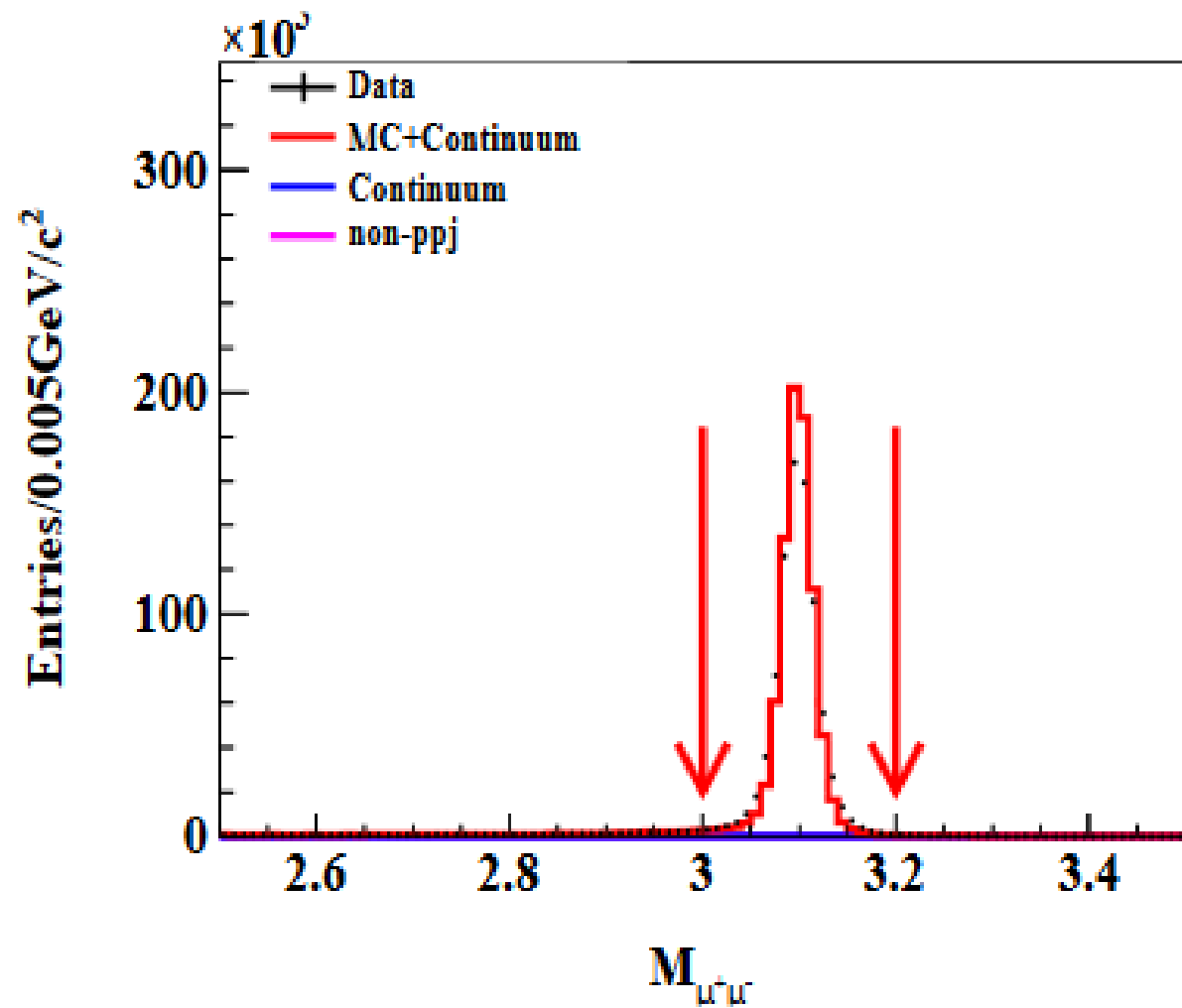
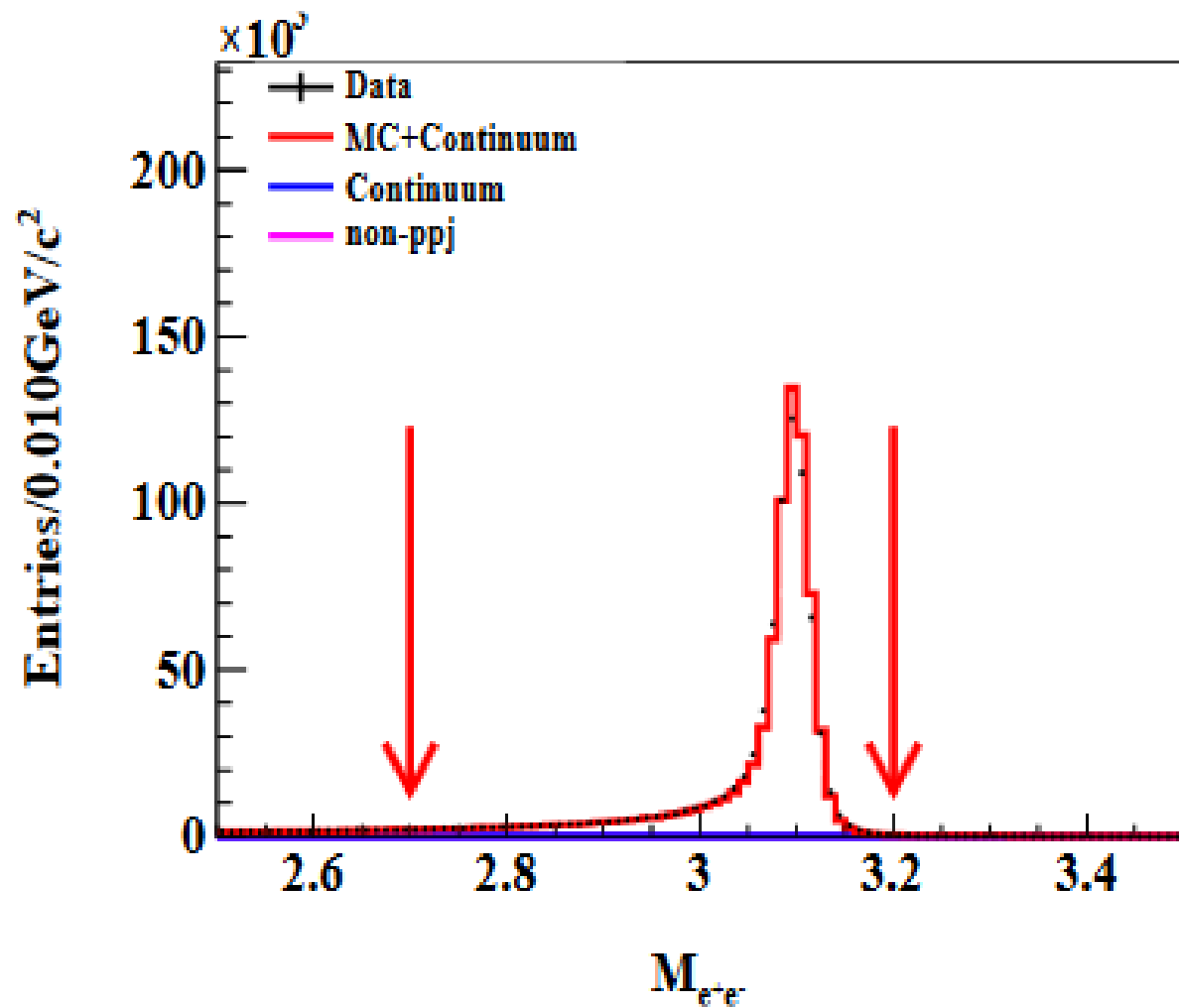
TOF(time of flight) 粒子飞行时间

PID (particle identification) 粒子鉴别

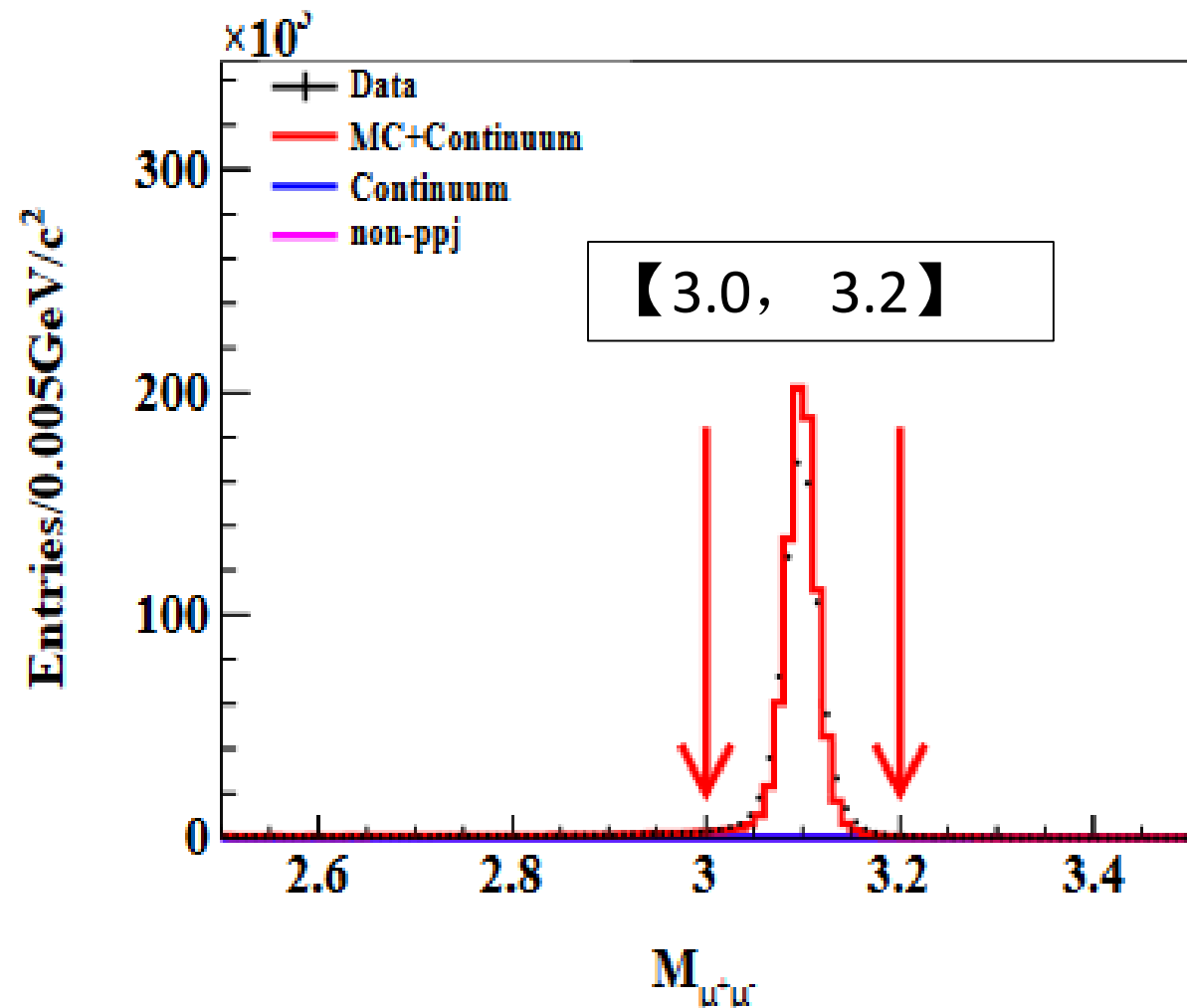
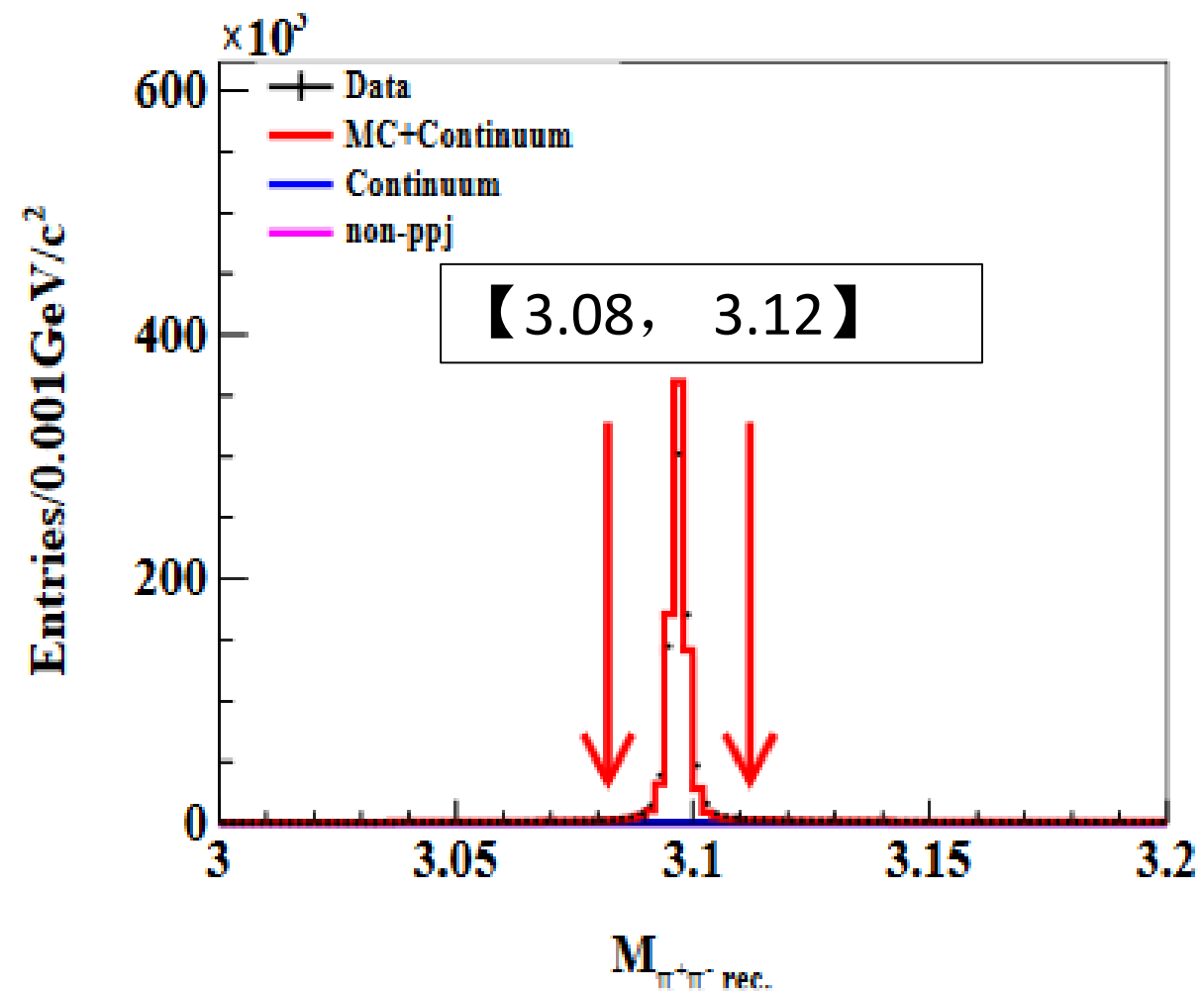


$$3.08 \text{ GeV} < M_{\pi^+\pi^- \text{ rec.}} < 3.12 \text{ GeV}$$

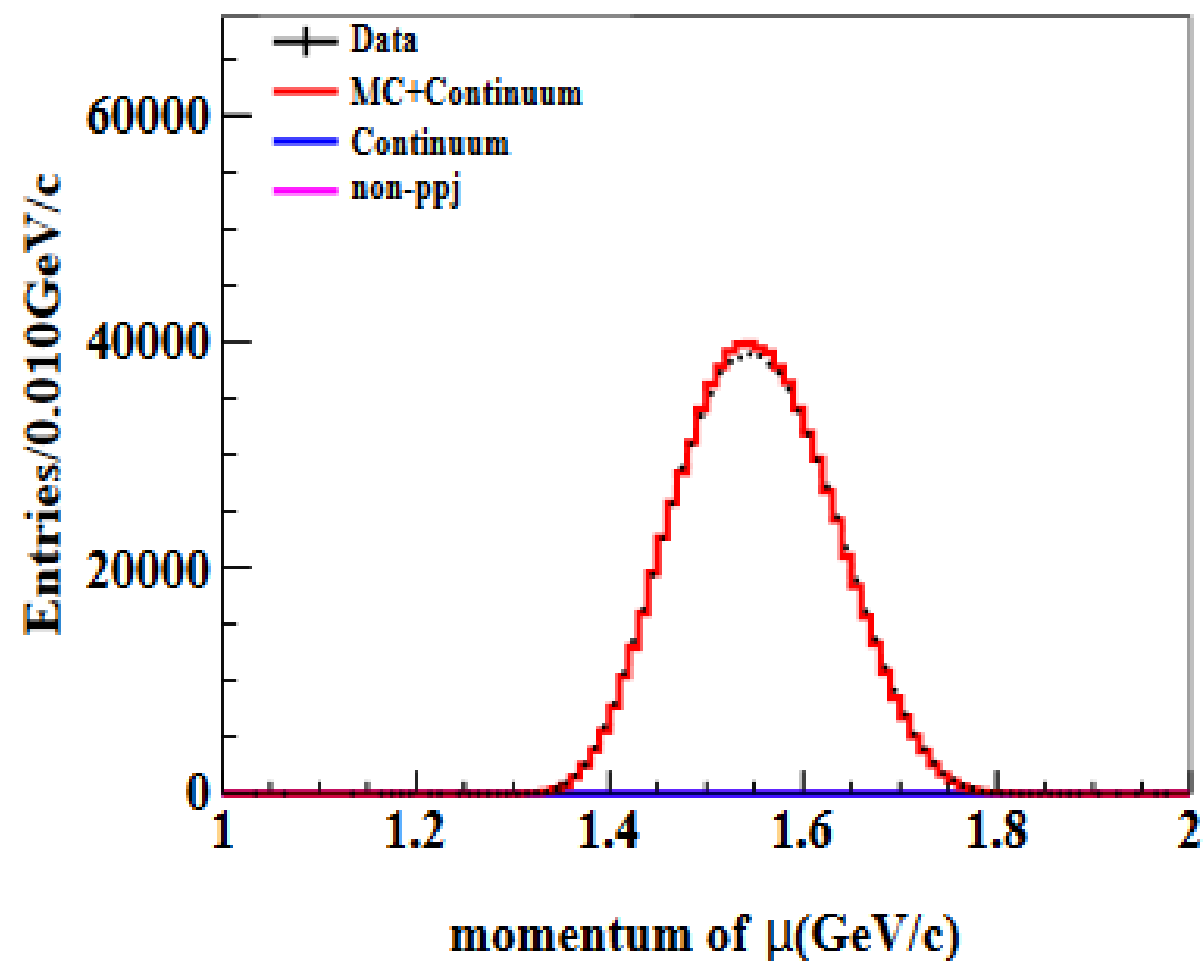
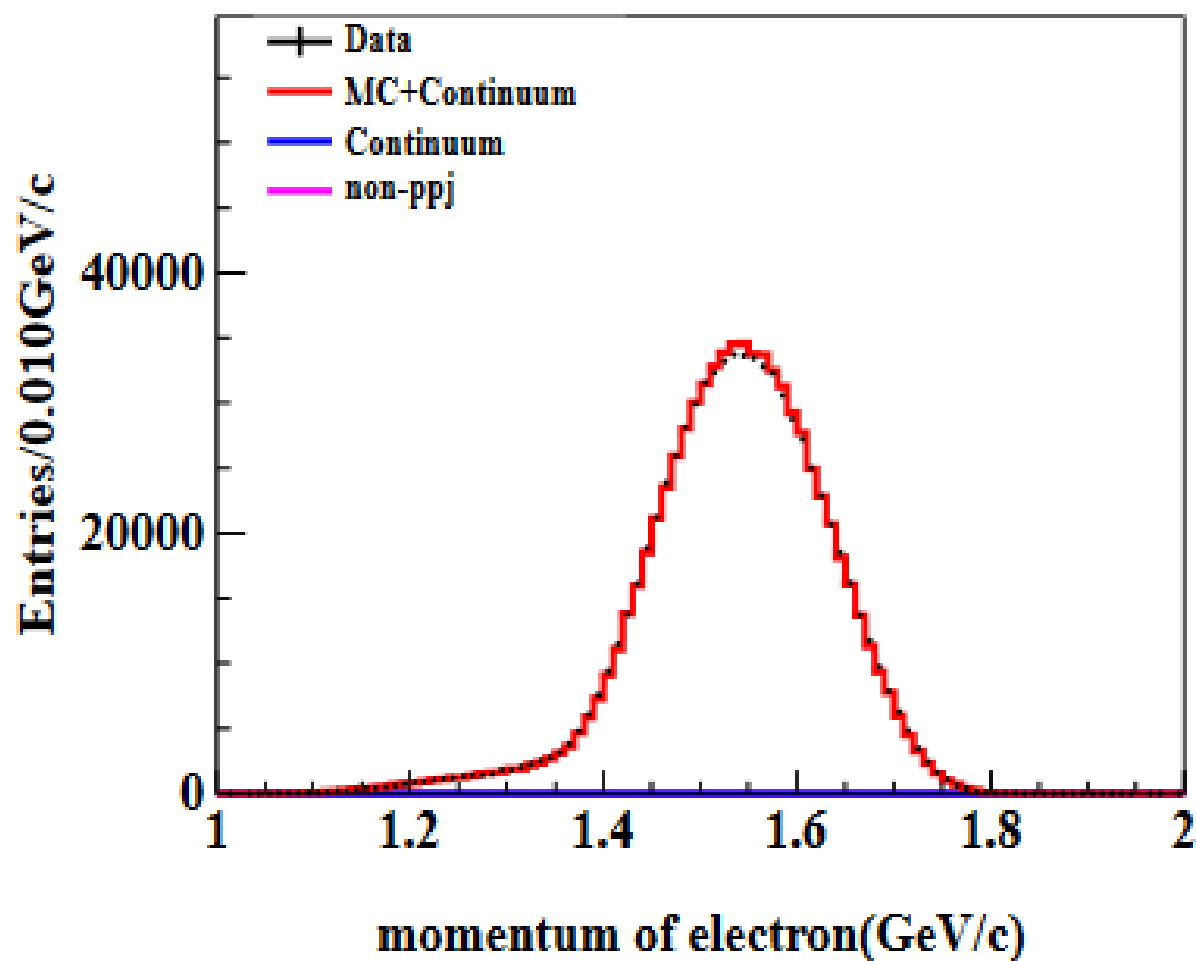


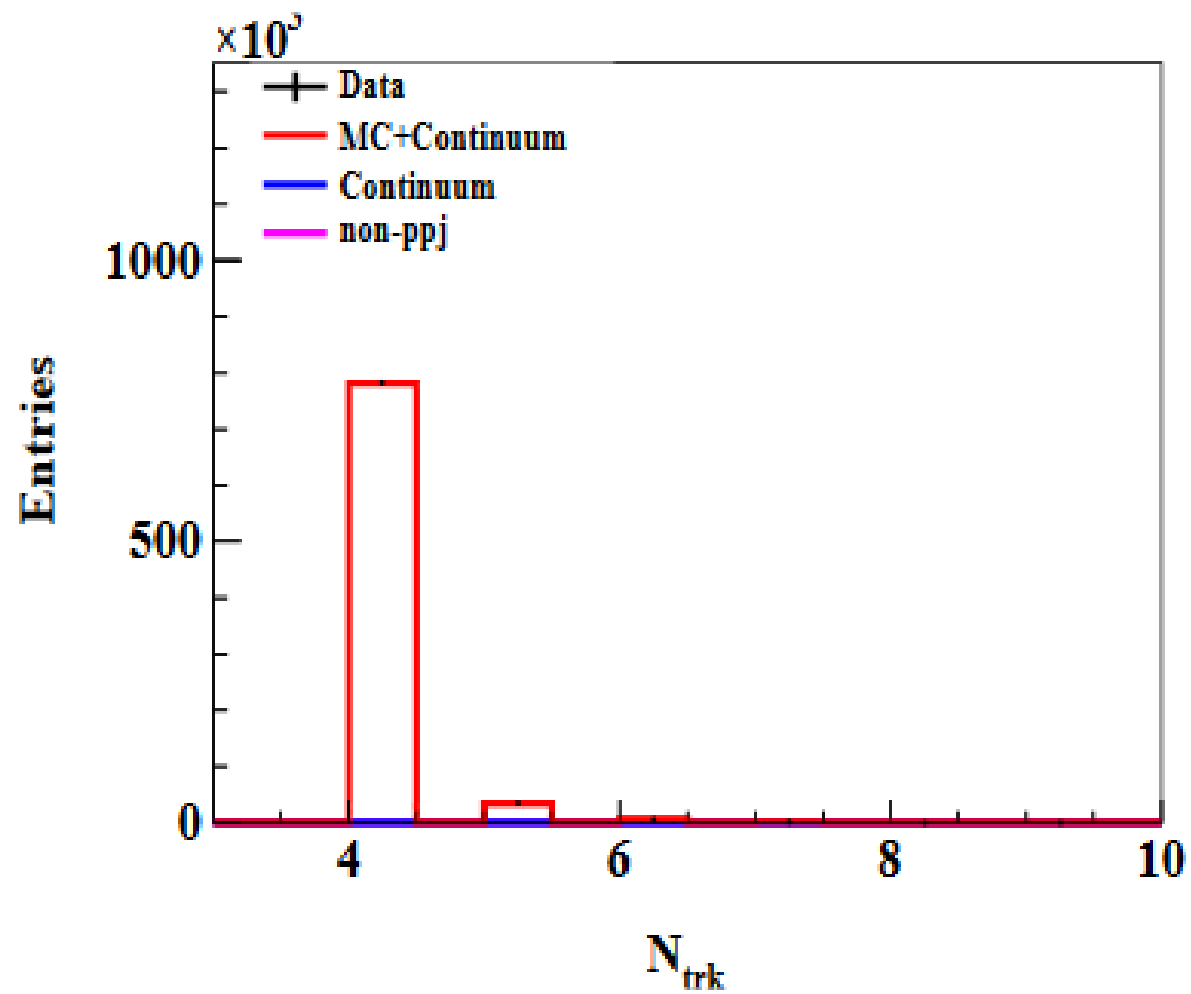
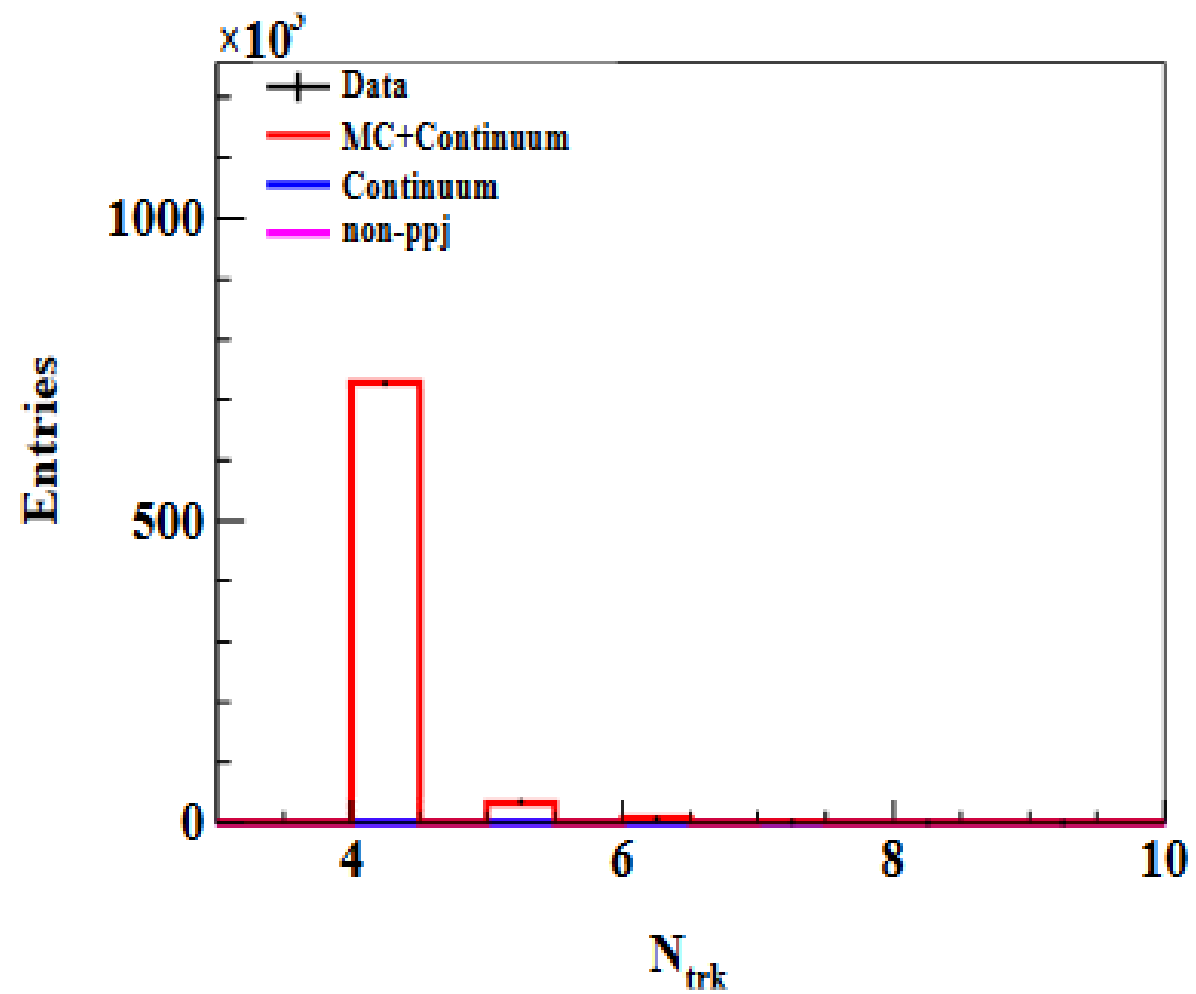


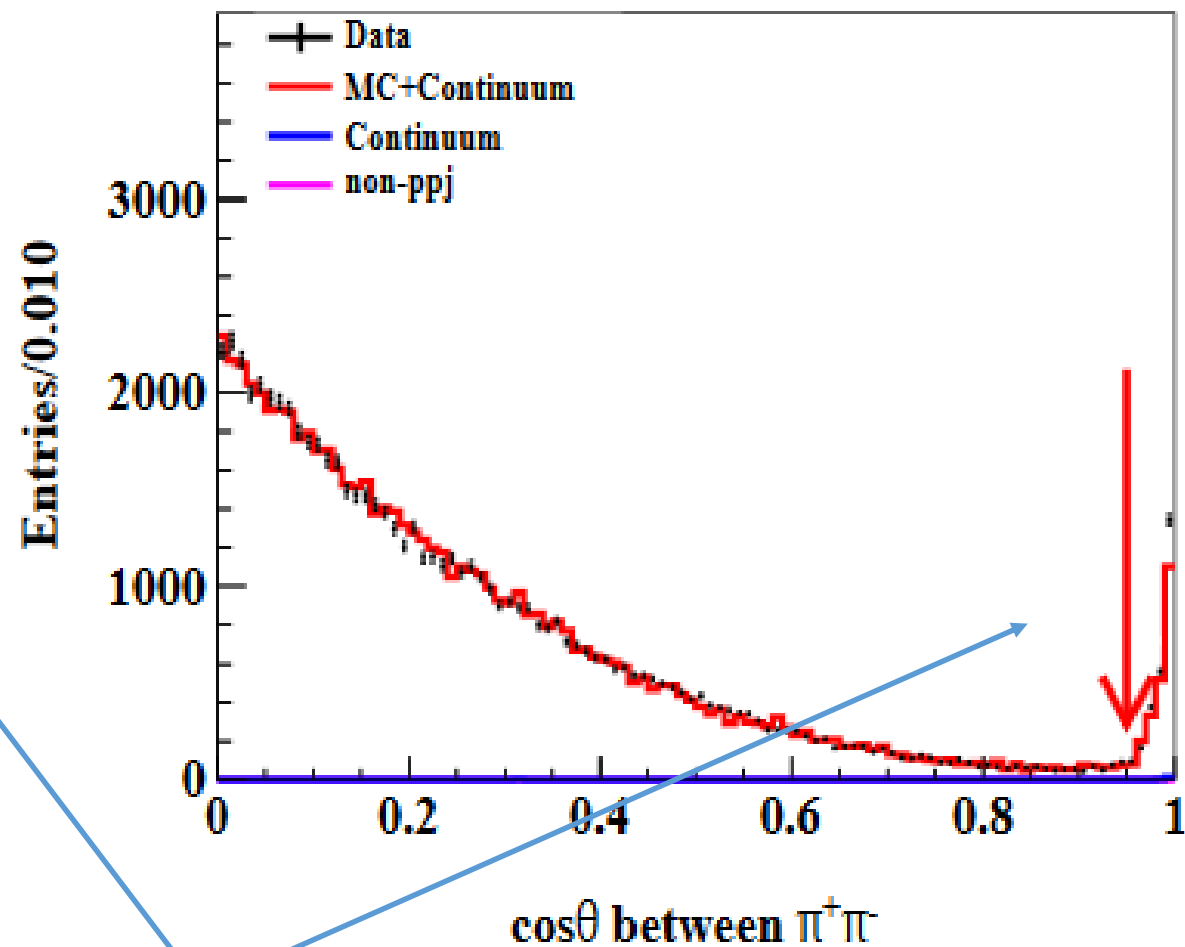
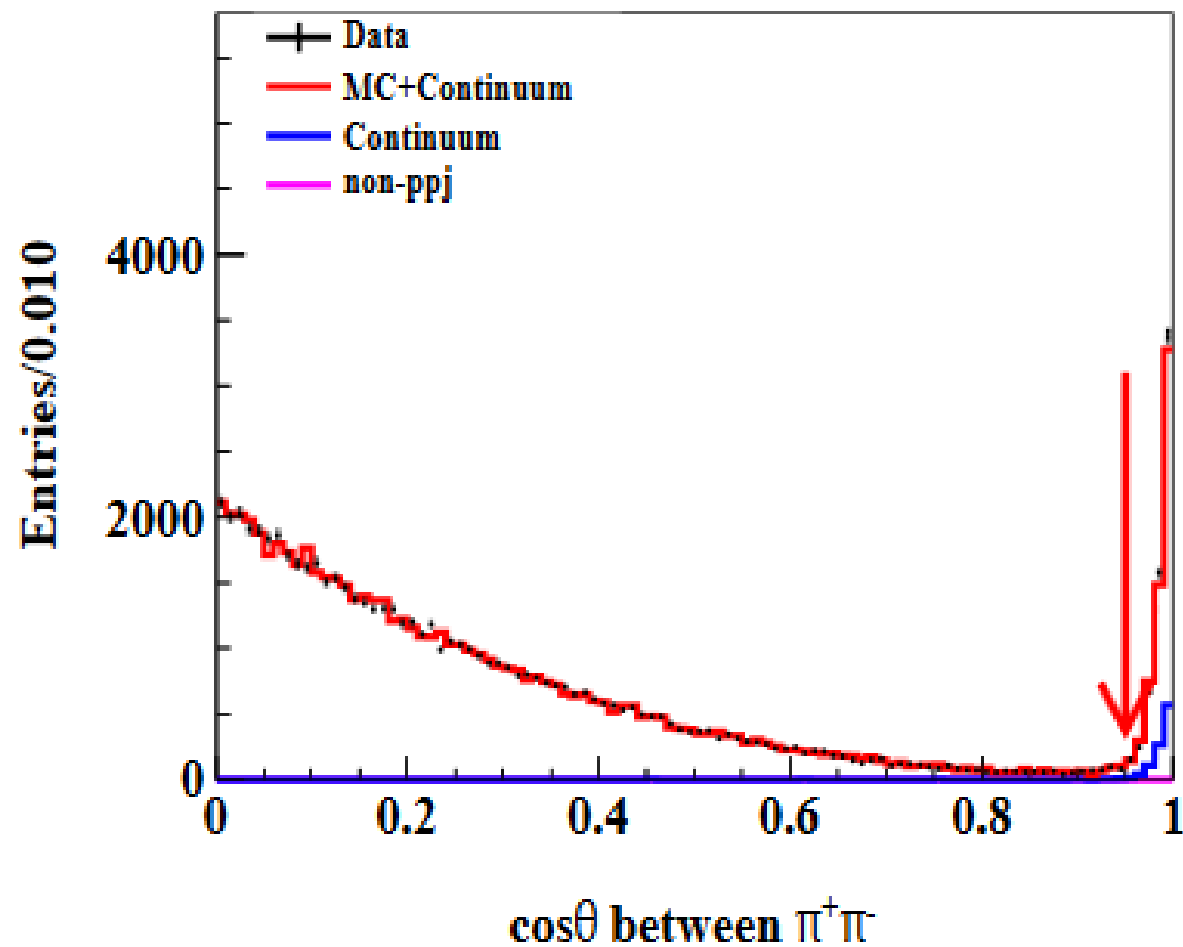
e 与 μ 的不变质量分布有什么差别？为什么？



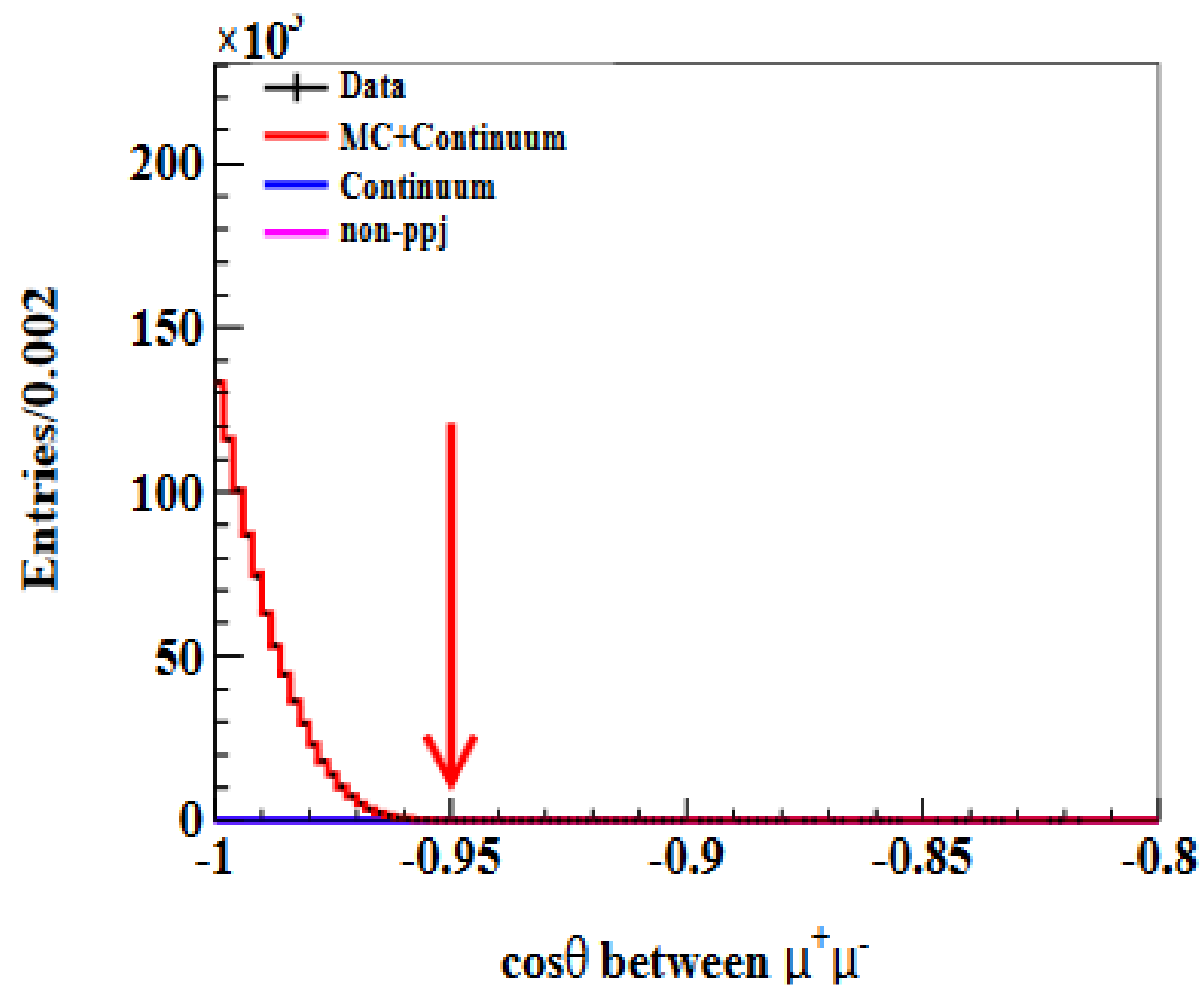
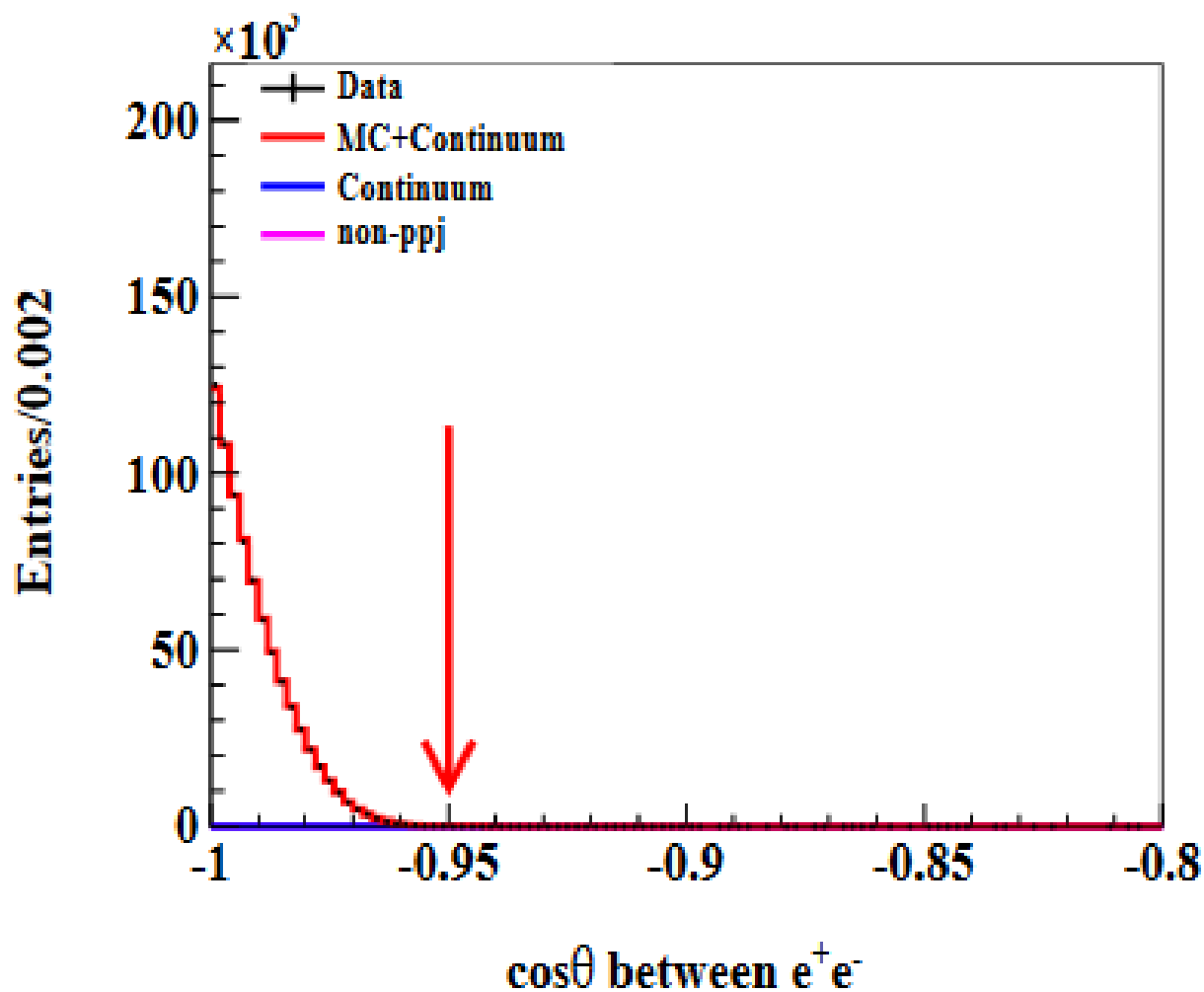
与 $\pi\pi$ 反冲比较，轻子对的不变质量分布有什么不同？



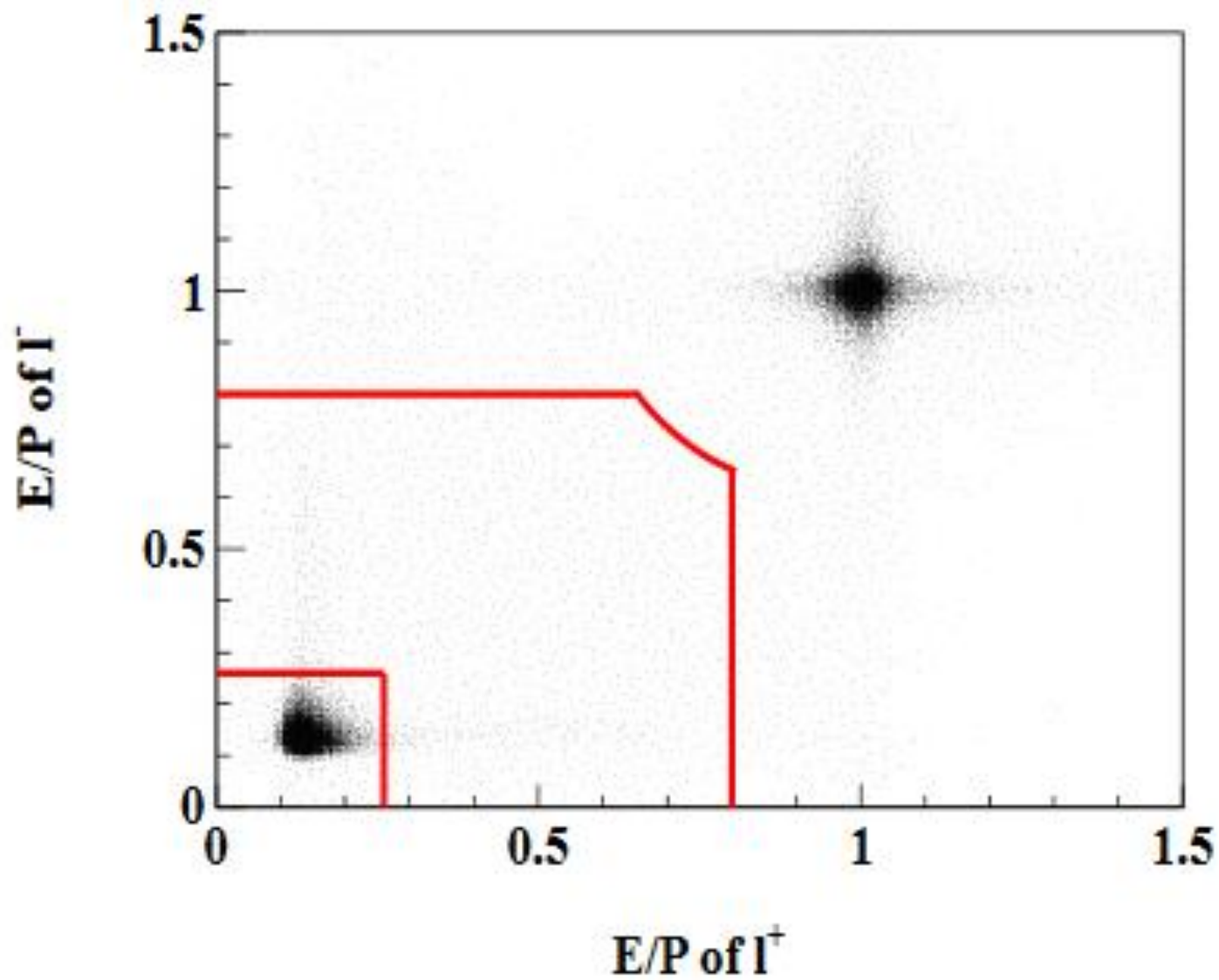




去除来自 γ conversion 的本底



轻子对之间的夹角与 $\pi^+\pi^-$ 之间的夹角有什么明显不同？为什么？



利用 E/p 区分 e 与 μ

术语园地

EMC

(ElectroMagnetic Calorimeter)

电磁量能器

事例选择进阶：本底分析



谁是信号？谁是本底？
峰耶？树耶？云耶？天耶？
有无相生、难易相成

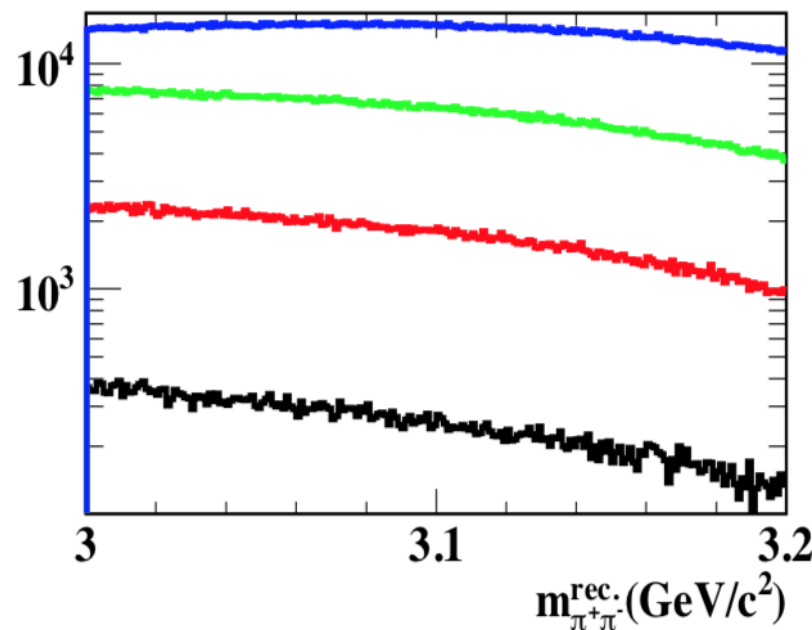
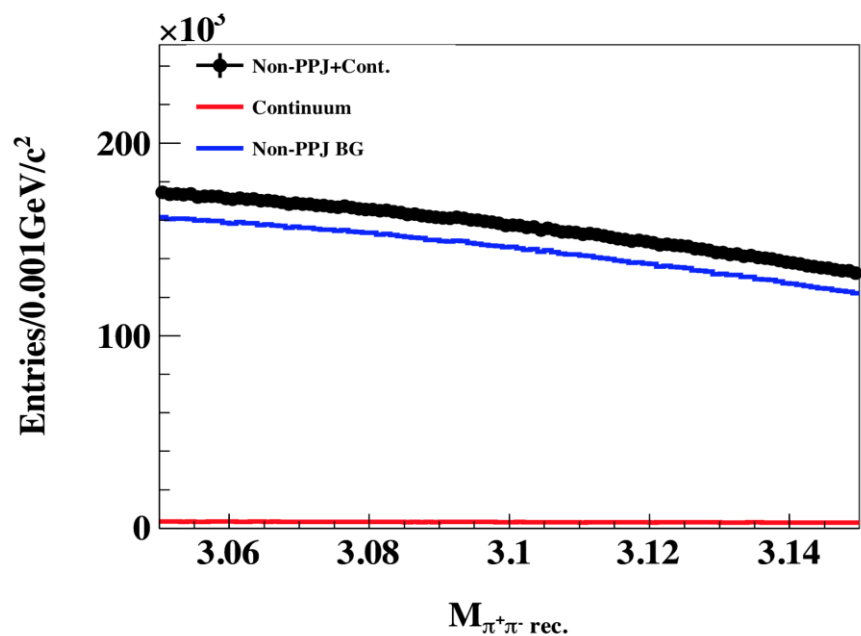
本底分析

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X} = \frac{N^{obs} - N^{bg}}{\epsilon \cdot N^X}$$

- 本底分析与事例选择常常是一个迭代的过程
 - 针对特定的本底，有时可以用简单有效的方法去除（veto）
- 本底须知：
 - 形状是否平滑？有没有类似信号的起伏（峰状本底）？
 - 本底数量大小？
- 确定本底的主要方法：
 - Inclusive MC: **topology**（a fantastic tool at BESIII）
 - 数据驱动（data driven）：其他数据、边带法（sideband）、检查分布
 - 猜测（经验依赖）

$\pi^+\pi^-J/\psi$ 过程的本底

- 平滑，没有 peak
- 受益于事例选择中：一个事例（event）可以有多条记录（entries）
 - 虽然错误组合更多了，但是避免了人为的偏差（bias）
- 进一步检查不同的 J/ψ 衰变的误组合，仍然平滑



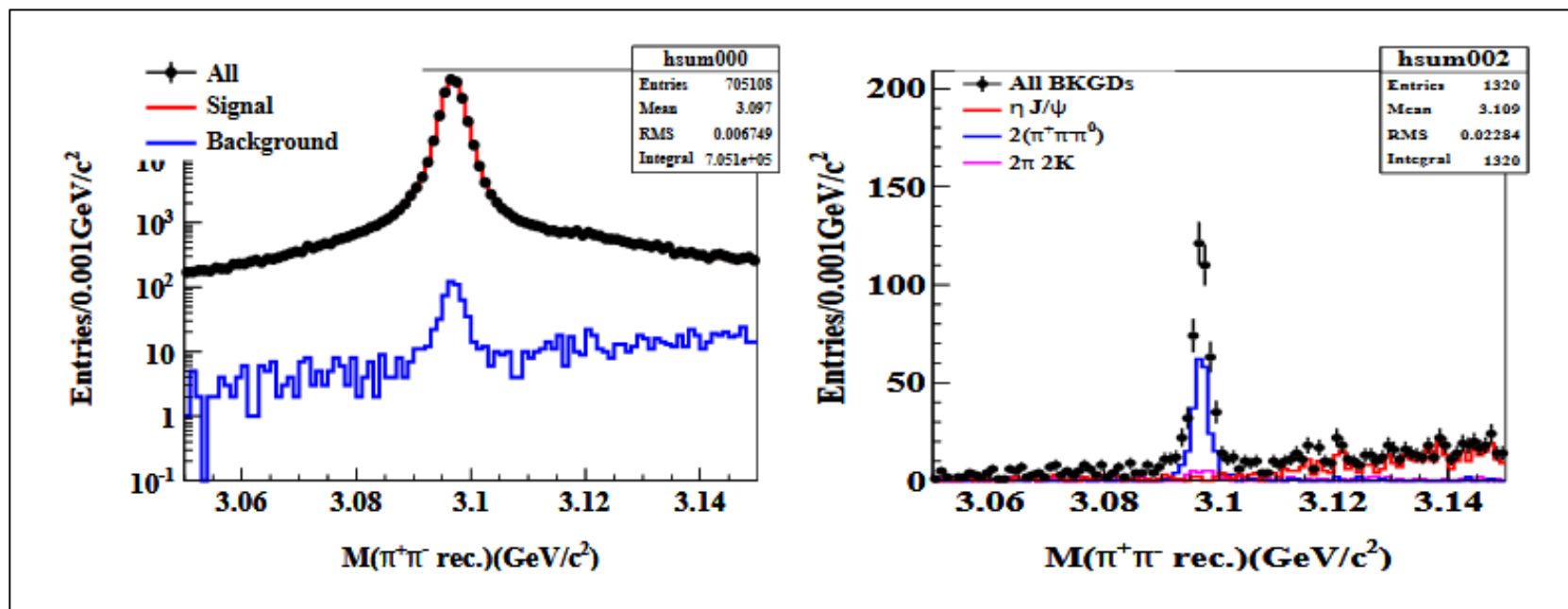
$J/\psi \rightarrow e^+ e^-$ 测量的本底

☑ MC topology 分析工具（首创自杜书先，由李刚等改进）

No.	decay chain	iTopo	N_{evt}
$e^- \pi^- e^+ \pi^+$ (final state)			
0	$\psi' \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow e^- e^+$	0	566983
1	$\psi' \rightarrow \pi^- J/\psi \pi^+, J/\psi \rightarrow e^+ \gamma_{FSR} e^-$	1	118425
2	$\psi' \rightarrow \pi^- J/\psi \pi^+, J/\psi \rightarrow \gamma_{FSR} e^- e^+ \gamma_{FSR}$	2	16079
3	$\psi' \rightarrow \pi^+ J/\psi \pi^- \gamma_{FSR}, J/\psi \rightarrow e^- e^+$	3	1802
4	$\psi' \rightarrow J/\psi \pi^+ \pi^- \gamma_{FSR}, J/\psi \rightarrow e^- e^+ \gamma_{FSR}$	9	447
11	$\psi' \rightarrow J/\psi \gamma_{FSR} \pi^- \pi^+, J/\psi \rightarrow e^- \gamma_{FSR} \gamma_{FSR} e^+$	12	51
90	$\psi' \rightarrow J/\psi \gamma_{FSR} \gamma_{FSR} \pi^+ \pi^-, J/\psi \rightarrow e^- e^+ \gamma_{FSR}$	90	1
703788 evts (7 modes)			

No.	decay chain	iTopo	N_{evt}
$e^- \pi^- e^+ \pi^+ \gamma$ (final state)			
5	$\psi' \rightarrow \eta J/\psi, \eta \rightarrow \gamma \pi^+ \pi^-, J/\psi \rightarrow e^- e^+$	7	269
10	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow e^- \gamma_{FSR} e^+, \eta \rightarrow \gamma \pi^+ \pi^-$	25	54
45	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow \gamma_{FSR} e^+ \gamma_{FSR} e^-, \eta \rightarrow \pi^- \pi^+ \gamma$	192	3
326 evts (3 modes)			

Peaking backgrounds!
 在本例中无需恐慌
 本底水平很低：0.1%水平



以 $J/\psi \rightarrow \mu^+ \mu^-$ 作为对比看 FSR

比 ee 末态含有更少的末态辐射效应

No.	decay chain	iTopo	N_{evt}
	$\mu^- \pi^- \mu^+ \pi^+$ (final state)		
0	$\psi' \rightarrow \pi^- J/\psi \pi^+, J/\psi \rightarrow \mu^+ \mu^-$	0	717031
1	$\psi' \rightarrow J/\psi \pi^+ \pi^-, J/\psi \rightarrow \gamma_{FSR} \mu^+ \mu^-$	1	<u>33801</u>
2	$\psi' \rightarrow \pi^- \gamma_{FSR} J/\psi \pi^+, J/\psi \rightarrow \mu^- \mu^+$	7	2458
4	$\psi' \rightarrow J/\psi \pi^+ \pi^-, J/\psi \rightarrow \mu^- \gamma_{FSR} \mu^+ \gamma_{FSR}$	6	569
7	$\psi' \rightarrow \gamma_{FSR} J/\psi \pi^+ \pi^-, J/\psi \rightarrow \mu^- \gamma_{FSR} \mu^+$	11	113
30	$\psi' \rightarrow \pi^+ J/\psi \pi^- \gamma_{FSR}, J/\psi \rightarrow \mu^+ \mu^- \gamma_{FSR} \gamma_{FSR}$	82	4
48	$\psi' \rightarrow \gamma_{FSR} \pi^- J/\psi \gamma_{FSR} \pi^+, J/\psi \rightarrow \mu^+ \mu^-$	42	2
753978 evts (7 modes)			

ee: 118425

事例选择小结

- ✓ 事例选择的目的是：提高信噪比
- ✓ 三条原则：
 - ✓ 尽可能简单的事例选择
 - ✓ 尽可能在保证效率的同时压低本底
 - ✓ 尽可能地利用运动学特征
- ✓ 作为小白该如何入手？
 - ✓ 思考
 - ✓ 阅读备忘录（memo）与文章
 - ✓ 积极参加组会、研讨会、合作组会议
 - ✓ 有疑问时向老师、同学、AI请教

思考题

- 分析换成测量 $\psi(2S) \rightarrow \pi^0 \pi^0 J/\psi$ 的分支比

1. 这个过程的末态是什么？

2. 分支比的计算公式？

3. 事例选择？

提示：一个 π^0 衰变成两个光子 ($\pi^0 \rightarrow 2\gamma$)

Tea break

提取信号数

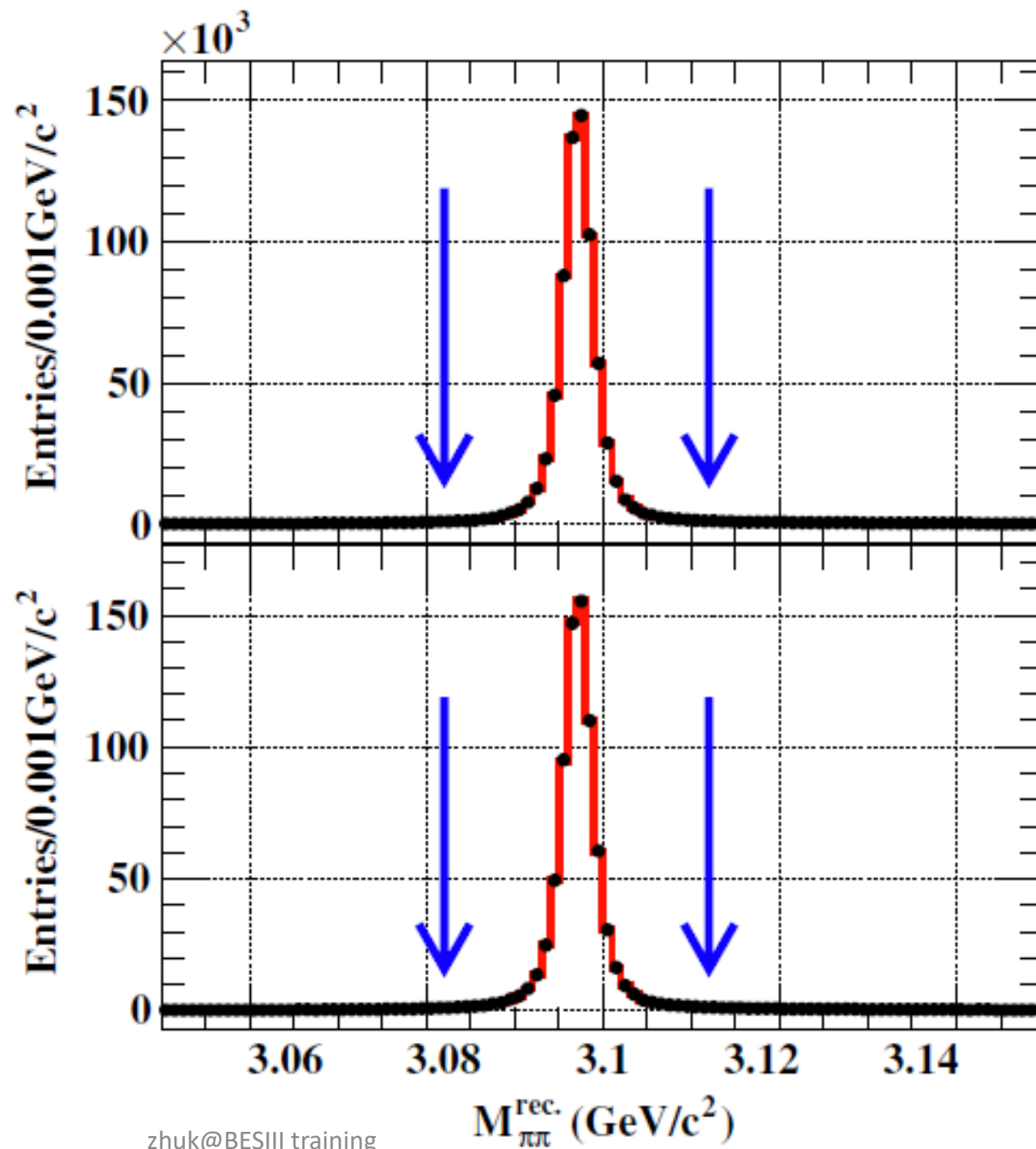
- 观测数减去本底数 ($N_{sig} = N_{obs} - N_{bg}$)
 - 1) Inclusive MC
 - 2) 边带法 (sideband)
- 3) 拟合

利用Inclusive MC 提供的信息

No.	decay chain	iTopo	N_{evt}
$e^- \pi^- e^+ \pi^+ \gamma$ (final state)			
5	$\psi' \rightarrow \eta J/\psi, \eta \rightarrow \gamma \pi^+ \pi^-, J/\psi \rightarrow e^- e^+$	7	269
10	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow e^- \gamma_{FSR} e^+, \eta \rightarrow \gamma \pi^+ \pi^-$	25	54
45	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow \gamma_{FSR} e^+ \gamma_{FSR} e^-, \eta \rightarrow \pi^- \pi^+ \gamma$	192	3
326 evts (3 modes)			

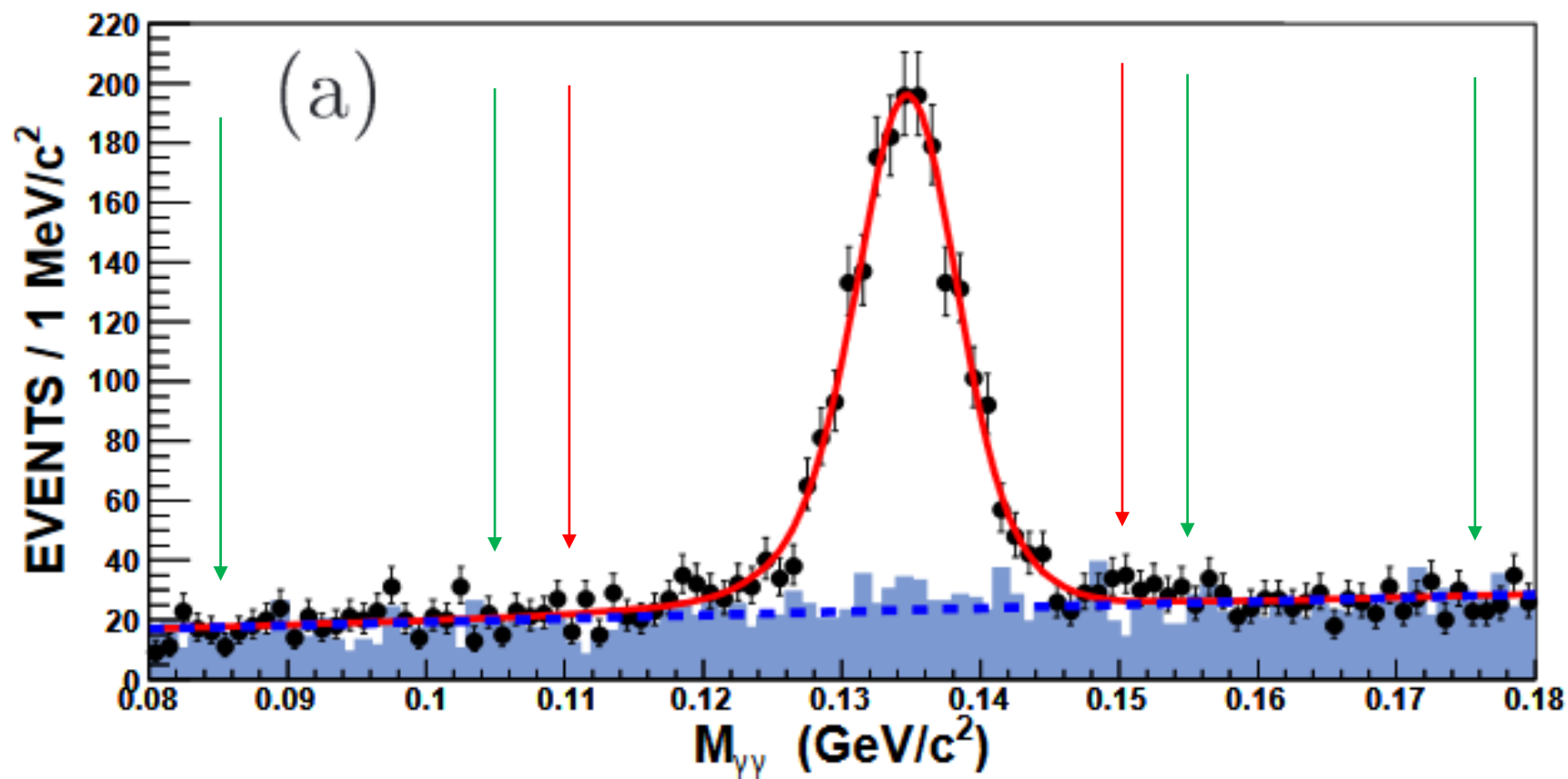
$$J/\psi \rightarrow l^+ l^-:$$

$$N_{obs} - N_{bg}$$

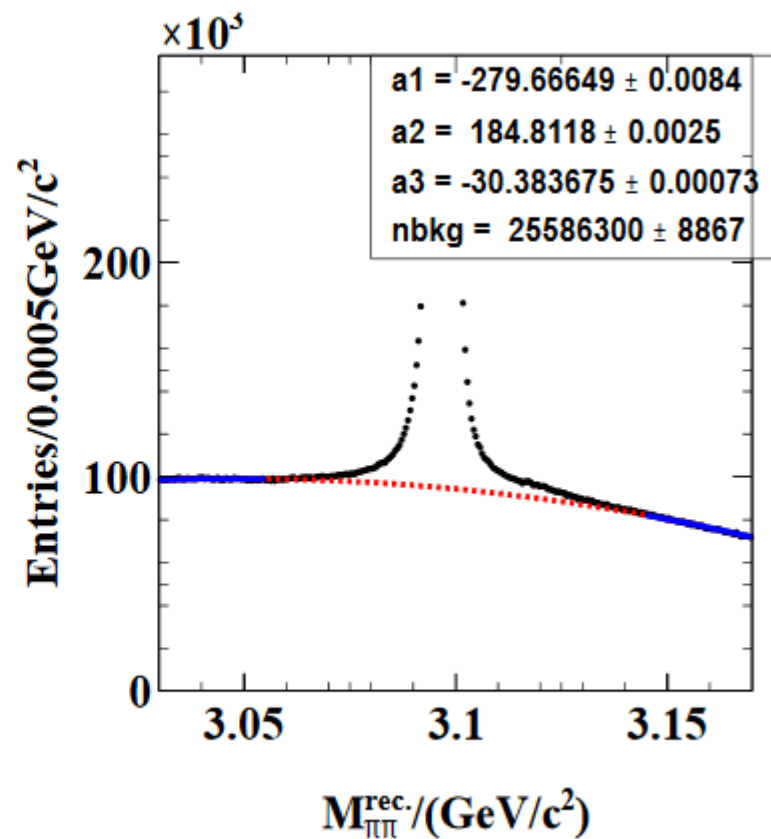
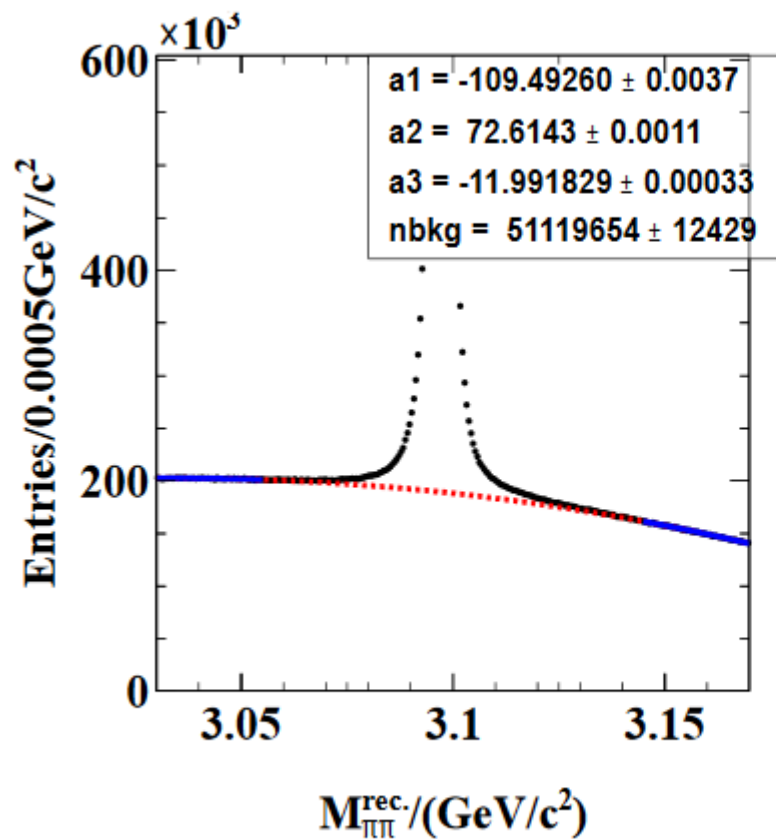


边带法 (sideband)

- 注意:
仅适用于平滑本底



拟合边带（非典型） $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$



二维边带

M_{34}

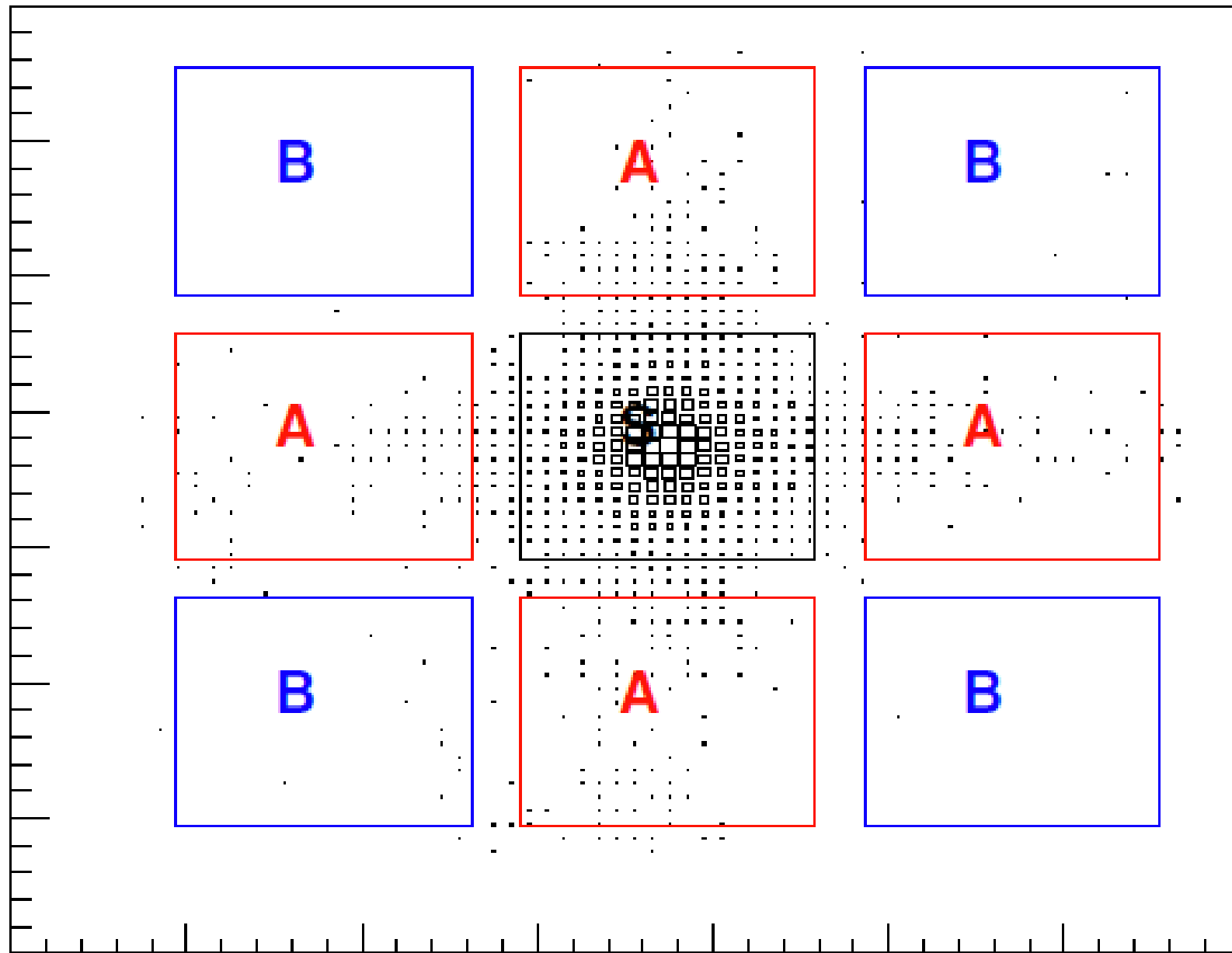
每一维对应一个共振态
的不变质量谱分布

常用公式:

$$S = \frac{A}{2} + \frac{B}{4}$$

(S、A、B面积相等)

为什么是这个公式?



M_{12}

拟合法 $M_E(x) = N_S \cdot S(x) + N_B \cdot B(x)$

- RooFit 几乎能满足一切需求（下午柯老师）
- 主要工作：
确定描述信号与本底的函数
需要对物理与探测器有一定程度的了解

术语园地

PDF（probability density function）

对应于量子力学里的 $|A|^2$

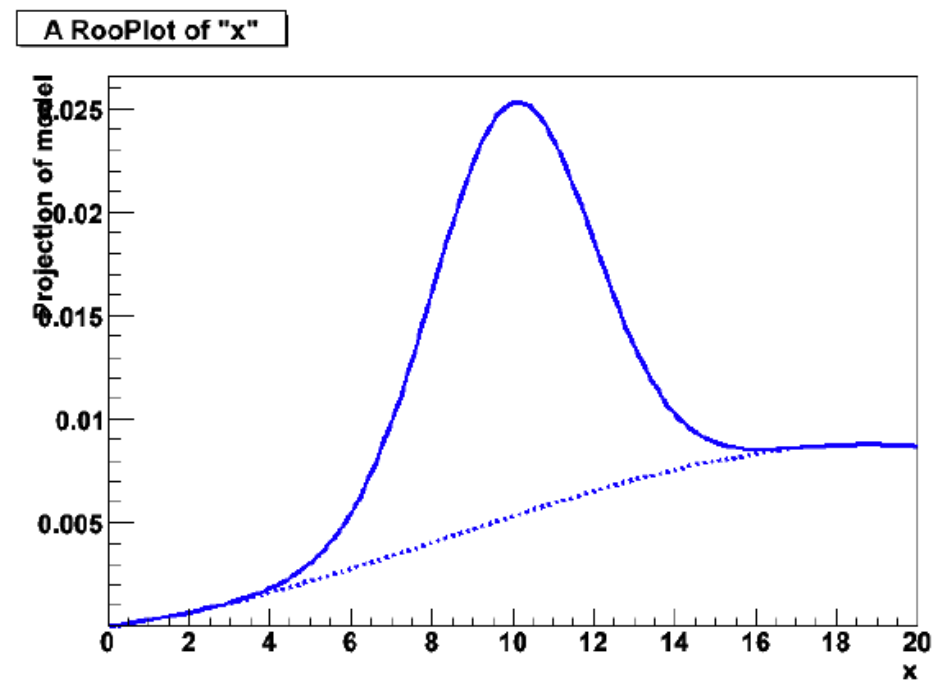


Figure 6 – Drawing of composite model and its components

Name	Functional form	Class name
Gaussian	$\exp\left(-0.5\left(\frac{x-m}{s}\right)^2\right)$	RooGaussian(name,title,x,m,s)
Exponential	$\exp(a \cdot x)$	RooExponential(name,title,x,a)
Polynomial	$1 + \sum_{i=1,n} a_i x^i$	RooPolynomial(name,title,x,alist)
Chebyshev polynomial	$1 + \sum_{i=1,n} a_i T_i(x)$	RooChebyshev(name,title,x,alist)

Table 1 – Basic functions implemented in RooFit

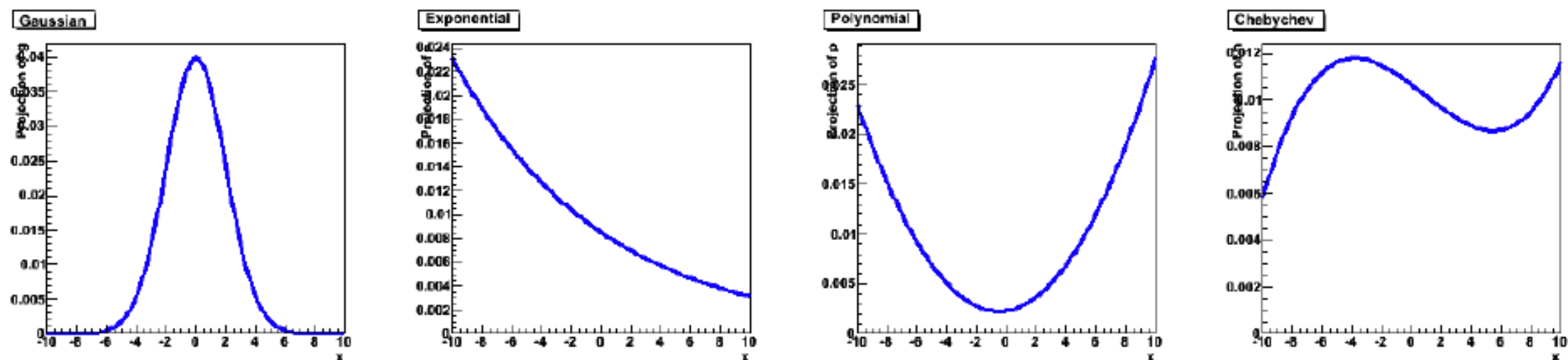
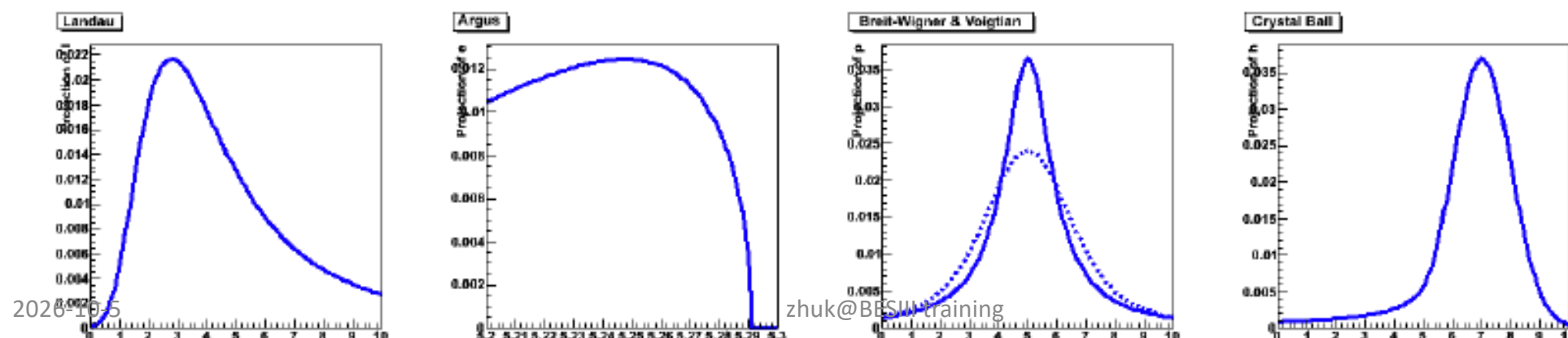


Figure 7 – Basic p.d.f shapes: Gaussian, Exponential, Polynomial and Chebyshev polynomial

Landau	$\text{TMath::Landau}(x, \text{mean}, \text{sigma})$	$\text{RooLandau}(\text{name}, \text{title}, x, \text{mean}, \text{sigma})$
Argus	$x \left(1 - \left(\frac{x}{m}\right)^2\right)^p \cdot \exp\left(c \left(1 - \left(\frac{x}{m}\right)^2\right)\right)$	$\text{RooArgusBG}(\text{name}, \text{title}, x, m, c, p)$
Breit-Wigner	$\frac{1}{(x - m)^2 + \frac{1}{4} g^2}$	$\text{RooBreitWigner}(\text{name}, \text{title}, x, m, g)$
Voigtian	$\frac{1}{(x - m)^2 + \frac{1}{4} g^2} \otimes \exp\left(-\frac{1}{2} \left(\frac{x}{s}\right)^2\right)$	$\text{RooVoigtian}(\text{name}, \text{title}, x, m, g, s)$
Crystal Ball	$\frac{\left(\frac{n}{ a }\right)^n e^{-\frac{1}{2} a^2}}{\left(\frac{n}{ a } - a - x\right)^n} \Big _{x < - a }, \exp\left(-\frac{1}{2} \left(\frac{x - m}{s}\right)^2\right) \Big _{x > - a }$	$\text{RooCBShape}(\text{name}, \text{title}, x, m, s, a, n)$
Decay	$\exp(- x / \tau) \otimes R(x)$	$\text{RooDecay}(\text{name}, \text{title}, x, \text{tau}, R)$



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

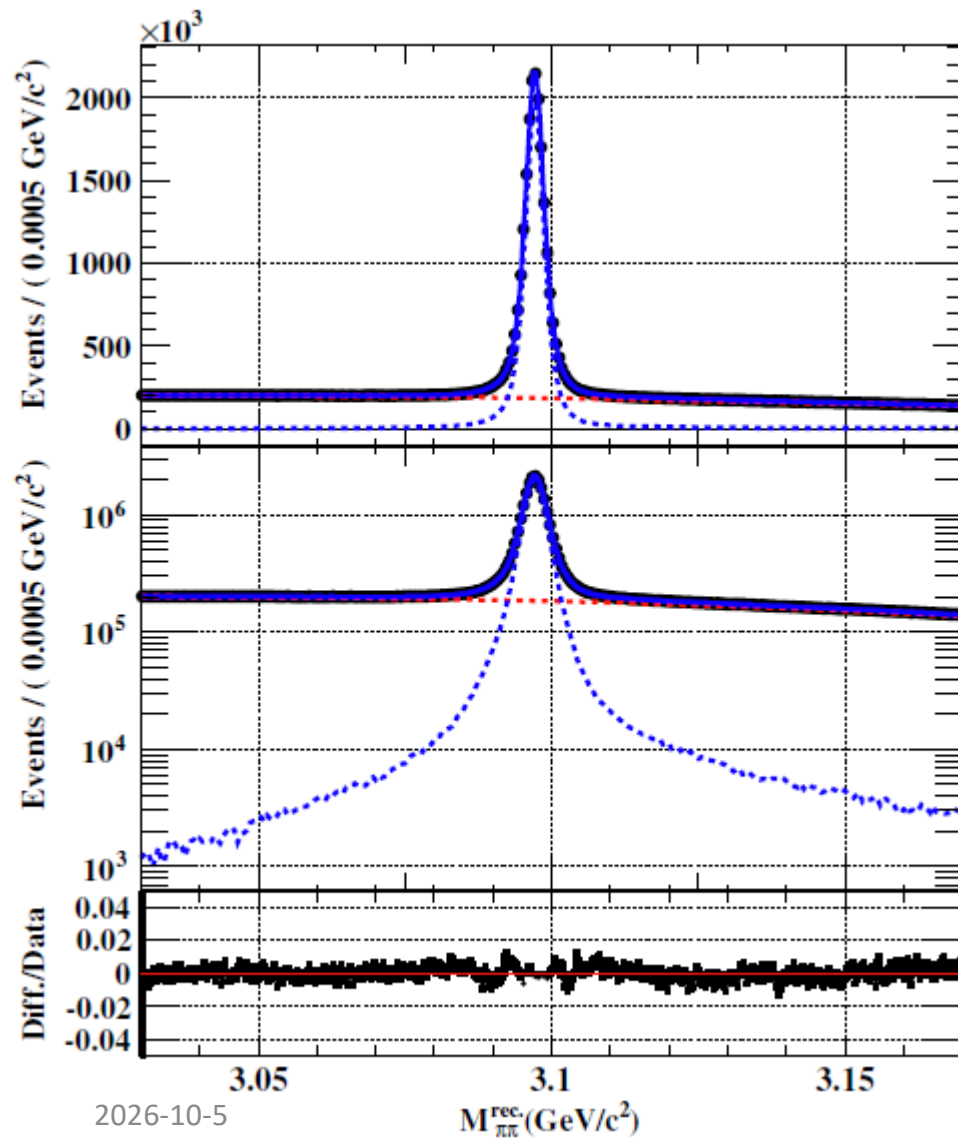


FIG. 3 (color online). The dipion recoil mass spectrum for $\psi(3686) \rightarrow \pi^+ \pi^- J/\psi$, $J/\psi \rightarrow \text{anything}$. Top: Data points (black) overlaid with the fit result (solid blue curve) obtained using the signal shape from $\psi(3686) \rightarrow \pi^+ \pi^- J/\psi$, $J/\psi \rightarrow e^+ e^-$ (blue dashed curve) and a second-order polynomial background shape (red dashed curve). Middle: The same plot as the top but with a log scale. Bottom: The fractional difference between the fit and the data.

- 信号形状采用的是无参数PDF
- 本底为二阶多项式

术语园地

χ^2 与 likelihood 拟合

$$\chi^2: \sum_i \frac{(m_i - f_i)^2}{\sigma_i^2}, \quad \mathcal{L}: \prod_i P_i$$

检查拟合是否成功、结果是否合理

```
*****
** 13 **MIGRAD      1000      1
*****
FIRST CALL TO USER FUNCTION AT NEW START POINT, WITH IFLAG=4.
START MIGRAD MINIMIZATION. STRATEGY 1. CONVERGENCE WHEN EDM .LT. 1.00e-03
FCN=25139.4 FROM MIGRAD  STATUS=INITIATE      10 CALLS      11 TOTAL
EDM= unknown      STRATEGY= 1      NO ERROR MATRIX
EXT PARAMETER      CURRENT GUESS      STEP      FIRST
NO.  NAME      VALUE      ERROR      SIZE      DERIVATIVE
1  mean      -1.00000e+00      1.00000e+00      1.00000e+00      -6.53357e+01
2  sigma      3.00000e+00      1.00000e+00      1.00000e+00      -3.60009e+01
ERR DEF= 0.5
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=25137.2 FROM MIGRAD  STATUS=CONVERGED      33 CALLS      34 TOTAL
EDM=8.3048e-07      STRATEGY= 1      ERROR MATRIX ACCURATE
EXT PARAMETER      STEP      FIRST
NO.  NAME      VALUE      ERROR      SIZE      DERIVATIVE
1  mean      -9.40910e-01      3.03997e-02      3.32893e-03      -2.95416e-02
2  sigma      3.01575e+00      2.22446e-02      2.43807e-03      5.98751e-03
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.      NDIM= 25      NPAR= 2      ERR DEF=0.5
9.241e-04 -1.762e-05
-1.762e-05 4.948e-04
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL      1      2
1  0.02606      1.000 -0.026
2  0.02606      -0.026 1.000
*****
** 18 **HESSE      1000
*****
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=25137.2 FROM HESSE  STATUS=OK      10 CALLS      44 TOTAL
EDM=8.30707e-07      STRATEGY= 1      ERROR MATRIX ACCURATE
EXT PARAMETER      INTERNAL      INTERNAL
NO.  NAME      VALUE      ERROR      STEP SIZE      VALUE
1  mean      -9.40910e-01      3.04002e-02      6.65786e-04      -9.40910e-01
2  sigma      3.01575e+00      2.22449e-02      9.75228e-05      3.01575e+00
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.      NDIM= 25      NPAR= 2      ERR DEF=0.5
9.242e-04 -1.807e-05
-1.807e-05 4.948e-04
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL      1      2
1  0.02672      1.000 -0.027
2  0.02672      -0.027 1.000
```

1. 检查输出信息
2. 估计统计误差
3. 检查拟合优度

术语园地

拟合优度 goodness of fit

对一个好的拟合有 $\chi^2/NDF \approx 1$

自由度: NDF (number of freedom)

$NDF = \text{No. of Bin} - \text{No. of paras}$

检查拟合是否成功、结果是否合理 II

MIGRAD MINIMIZATION HAS CONVERGED.

MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.

COVARIANCE MATRIX CALCULATED SUCCESSFULLY

FCN=25137.2 FROM MIGRAD STATUS=CONVERGED 33 CALLS 34 TOTAL
EDM=8.3048e-07 STRATEGY= 1 ERROR MATRIX ACCURATE

EXT	PARAMETER				STEP	FIRST
NO.	NAME	VALUE	ERROR	SIZE	DERIVATIVE	
1	mean	-9.40910e-01	3.03997e-02	3.32893e-03	-2.95416e-02	
2	sigma	3.01575e+00	2.22446e-02	2.43807e-03	5.98751e-03	

ERR DEF= 0.5

EXTERNAL ERROR MATRIX. NDIM= 25 NPAR= 2 ERR DEF=0.5

9.241e-04 -1.762e-05
-1.762e-05 4.948e-04

PARAMETER	CORRELATION	COEFFICIENTS	
NO.	GLOBAL	1	2
1	0.02606	1.000	-0.026
2	0.02606	-0.026	1.000

猪的浑身都是宝
这些参数皆重要

拟合小结

- 拟合是对前期分析工作已获得信息的集大成，包含了分析者对
 - 信号形状的认识：物理模型、数据样本的大小、探测器分辨、效率曲线、测量的中心值是否有偏等
 - 本底的了解：来源与大小、是否有未知/不理解的本底
- 多关注如何构建模型，这是知识的汇总

术语园地

分波分析（PWA: partial wave analysis）
带着复杂的物理模型、参数可多达上百的拟合

效率

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$

什么是效率

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot B_{int} \cdot N^X}$$

把中间态分支比单列出来的定义

- 分支比公式中的效率为事例的选择效率，包括探测器几何接收度、选择条件的效率、中间态的分支比等。
- 效率的估计一般用Monte Carlo模拟。基本的假设是Monte Carlo可以很好地模拟数据。
- 应当用实际数据检验Monte Carlo模拟的可靠性，必要时对Monte Carlo模拟效率进行修正。

术语园地

接收度 acceptance

效率 efficiency

探测效率 detection efficiency

什么是效率 (II)

- 一般地

$$\epsilon_{DT} = \epsilon_{DT}^{Geom} \cdot \epsilon_{DT}^{Track} \cdot \epsilon_{DT}^{Ptid} \cdot \epsilon_{DT}^{\gamma ID} \dots$$

$$\epsilon_{MC} = \epsilon_{MC}^{Geom} \cdot \epsilon_{MC}^{Track} \cdot \epsilon_{MC}^{Ptid} \cdot \epsilon_{MC}^{\gamma ID} \dots$$

$$\epsilon_{DT}^i = \epsilon_{MC}^i \cdot f^i$$

估一估

- $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow l^+ l^-$ 的效率是多少?

- $f^i - 1$ 是数据与MC的效率的差别, 即为一项系统误差的来源
- 对 f^i 的仔细研究即可对 MC 所得到的效率进行修正, 减少系统误差
 - 例如: $f^i = 1.06 \pm 0.01$

cut flow

检查事例选择程序的好帮手

重点关注效率明显变化的地方

Criteria	$\psi(2s) \rightarrow \Sigma^0 \bar{\Lambda}$	$\psi(2s) \rightarrow \bar{\Sigma}^0 \Lambda$
Total Events	80000	80000
$N_{good} = 4, N_{charge} = 0$	44198	43983
$N_p = 1, N_{\bar{p}} = 1, N_{\pi^+} = 1, N_{\pi^-} = 1$	43856	43643
$1 \leq N_\gamma \leq 10$	38075	38146
$\theta(\gamma, \bar{p}) > 20^\circ$	36491	34538
$\Lambda, \bar{\Lambda}$ reconstruction	34354	32564
Pass 6C	25192	22390
$\chi_{6C}^2(\gamma) < \chi_{6C}^2(\gamma\gamma/0\gamma)$	24681	21850
$\chi_{6C}^2 < 25$	18368	15447
$L_\Lambda/\sigma > 2 (\bar{\Lambda})$	16084	13502
$M_{\Lambda\bar{\Lambda}} > 3.48 \text{ GeV}/c^2$	15891	13498
$M_{\gamma\bar{\Lambda}}(M_{\gamma\Lambda}) \geq 1.15 \text{ GeV}$	15870	13488
Efficiency ($\epsilon\%$)	19.84%	16.86%

Table 2: Cut flow of $\psi(2S) \rightarrow \bar{\Sigma}^0 \Lambda + c.c.$
zhuk@BESIII training

关于效率估计的说明

- 数据与 MC 的不一致永远存在，原则上利用 MC 得到的效率永远需要修正，但对于不同的过程要求也不同
 - 对于小统计量情况，统计误差为主，数据与信号模拟的差别对结果影响很小，但对本底估计时需特别关注（稀有过程、新物理）
 - 对于大统计量情形，期望高精度结果，数据与信号模拟的差别不容忽视

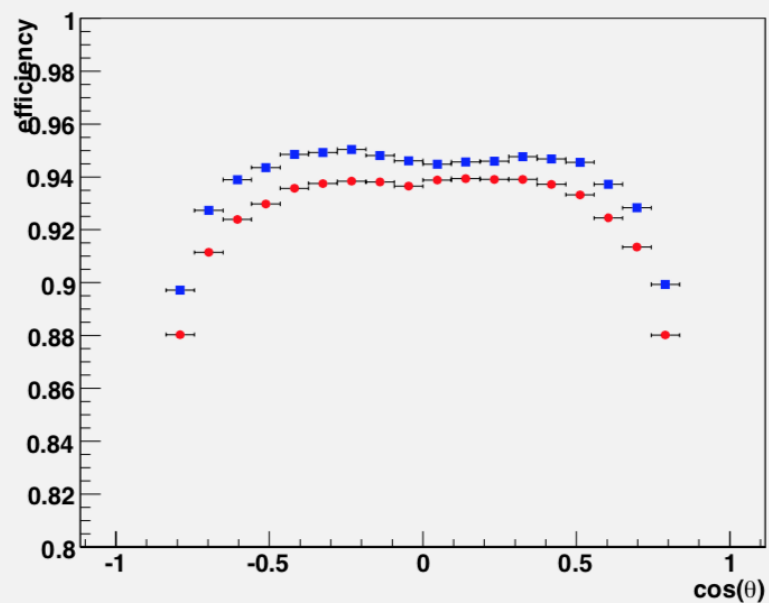
不考虑误差的分析就是瞎胡闹

径迹效率修正

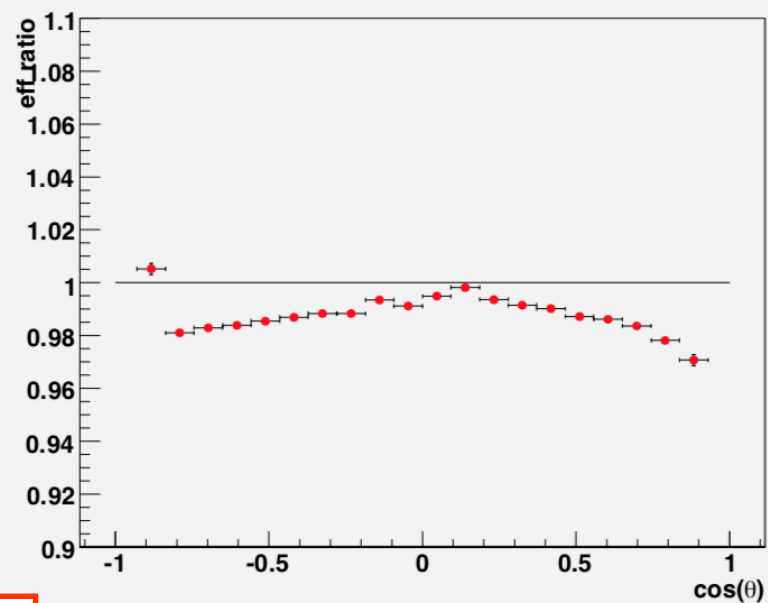
$$B_{\pi\pi J/\psi} = \frac{N_{\pi\pi J/\psi}}{\epsilon_{MC} \cdot N^{tot}} \quad \epsilon_{MC} = \epsilon_{\pi^+} \cdot \epsilon_{\pi^-} \cdot \epsilon_{m_{\pi\pi}} \cdots$$

- ❑ 径迹效率研究
 - ❑ 目的：给出数据与MC径迹效率关于动量与角度的二维分布
- ❑ 在分析中用 data/MC 效率之比 reweight MC （效率修正）
- ❑ 目的：精度控制在 0.1%

Tracking efficiency of pionm_

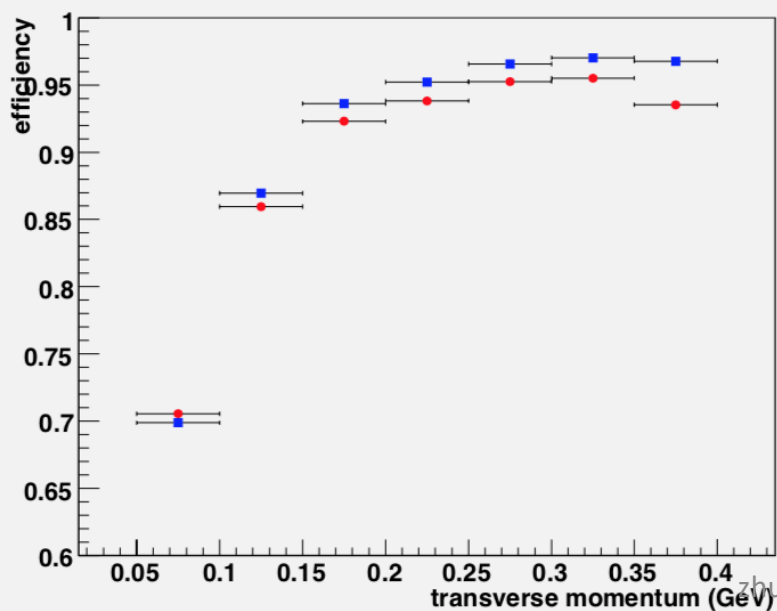


Tracking efficiency difference e_data/e_mc

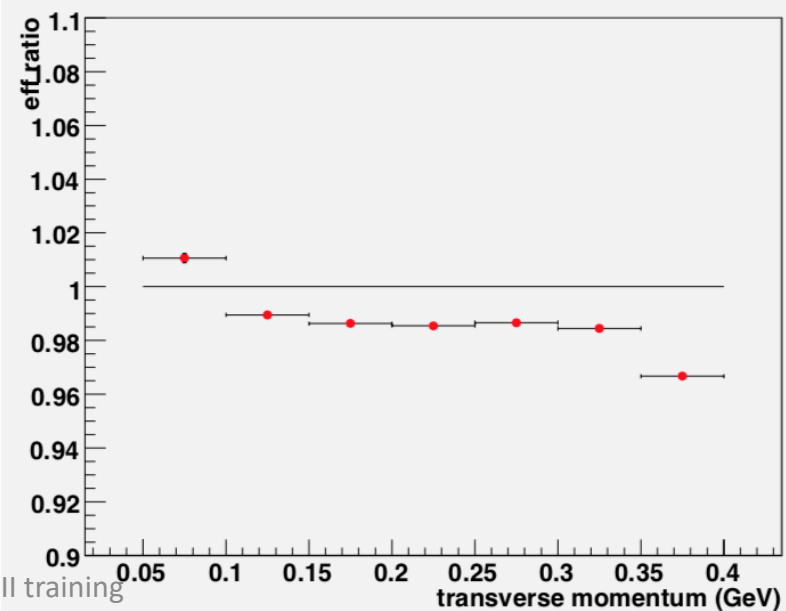


pi+

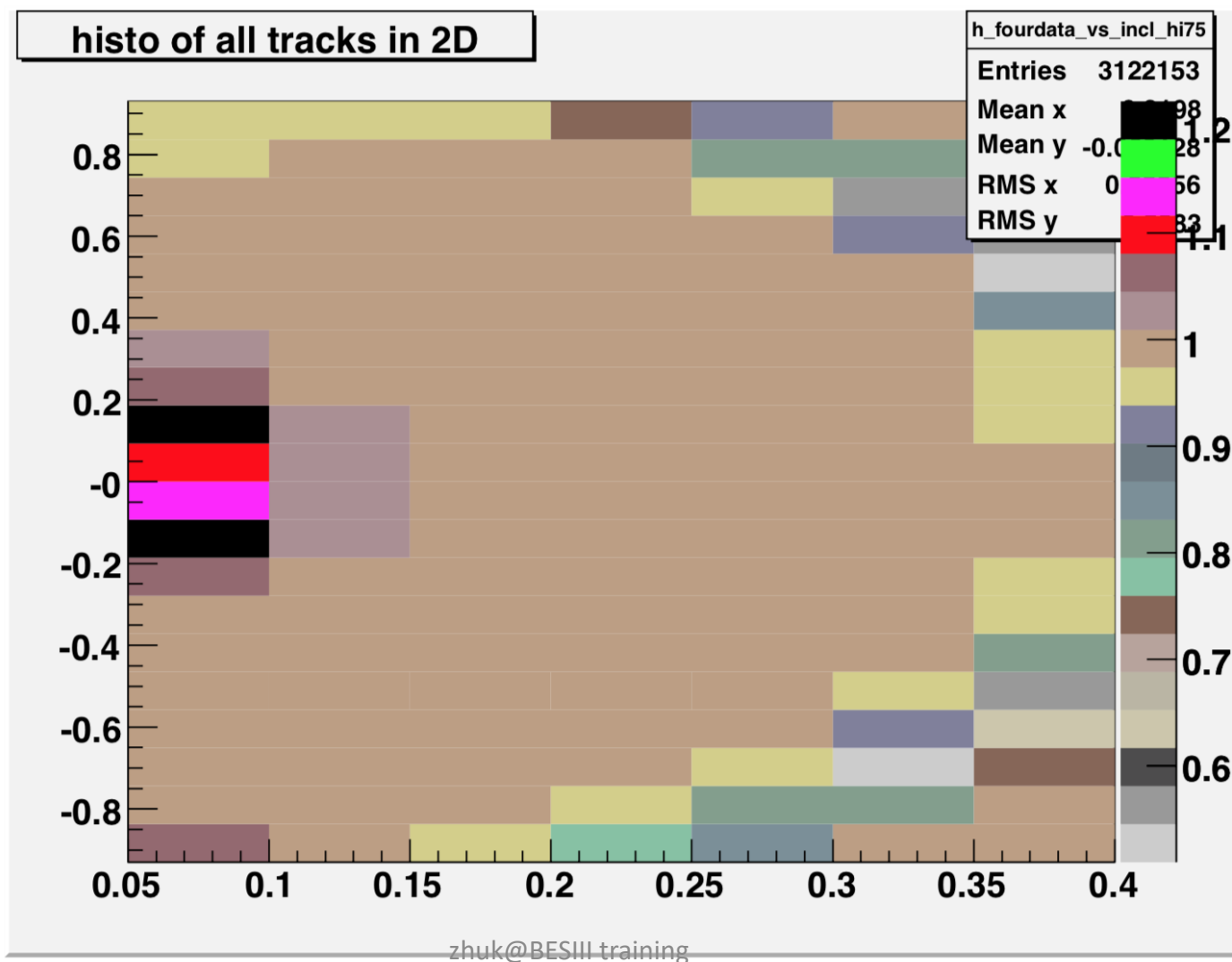
Tracking efficiency of pionm_



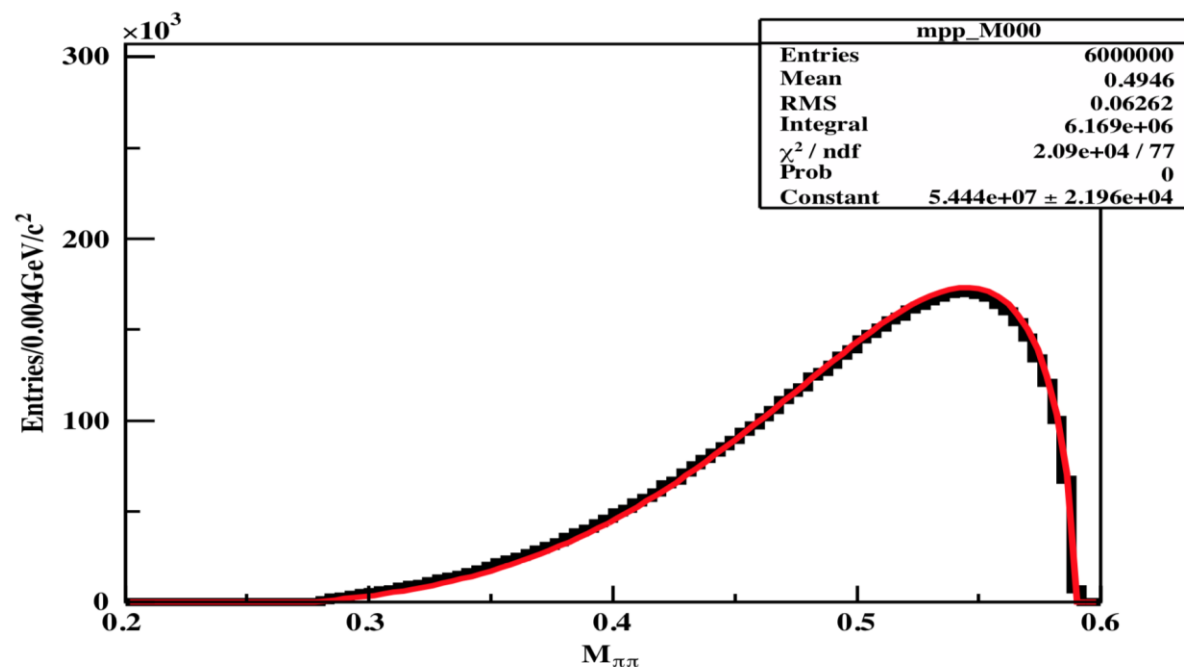
Tracking efficiency difference e_data/e_mc



数据与MC二维径迹效率比（reweight 的依据）



谱形分布的效率修正



- MC 采用了 JPIPI, g 和 g_1 来自 BES1 的测量:

$$g(s - 2m_{\pi\pi}^2) + 2g_1 E_{\pi^+} E_{\pi^-} \epsilon_{\psi'}^*(M) \cdot \epsilon_{\psi}(\Lambda)$$

- 最终采用数据（轻子道的干净样本）来 reweight MC $M_{\pi\pi}$ 分布

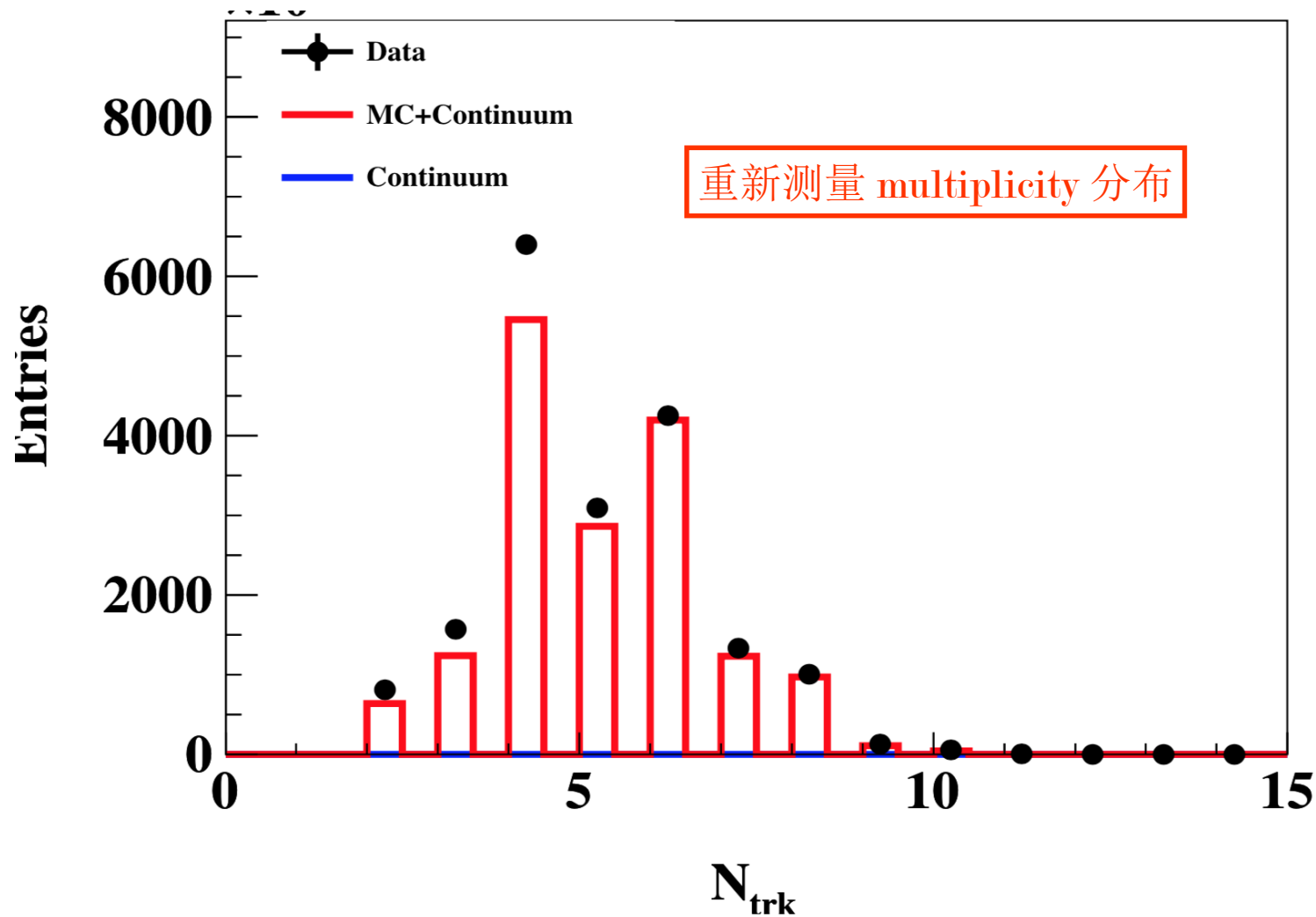
多重数分布的效率修正

- 高统计量的新问题:

探测器里出现 2 tracks、4tracks, 6tracks, ...时, 每种情形的每根 track 的重建效率并不同相同: track 越多, 效率越低, ...

prongs of J/ψ	0	2	4	6	8	10	Total
number of events(n_i or n)	607255	11885458	14889294	6462789	688523	11853	34549656 ^a
w_k from EvtGen	0.0175	0.3440	0.4310	0.1871	0.0199	0.0003	53.505%
w_k from Mehtod-I	0.0225	0.3899	0.4024	0.1667	0.0169	0.0015	53.549%
w_k from Mehtod-II	0.0230	0.3960	0.4010	0.1624	0.0164	0.0013	53.574%
ϵ_k (MC)	55.98%	55.02%	53.20%	51.20%	48.61%	44.70%	

数据、MC 的多重数分布符合不佳



解方程

来自数据

$$\begin{pmatrix} 884138.2 \\ 1691653.1 \\ 6947052.0 \\ 3295541.7 \\ 4557824.2 \\ 1395702.8 \\ 1059873.7 \\ 134480.3 \\ 55084.1 \\ 4179.1 \\ 835.5 \end{pmatrix}$$

来自MC

$$= \begin{pmatrix} 52.8927 & 2.8744 & 0.1369 & 0.0179 & 0.0022 & 0.0000 \\ 1.7753 & 9.5437 & 1.2390 & 0.1800 & 0.0247 & 0.0291 \\ 0.9905 & 41.1825 & 6.3557 & 1.0455 & 0.2037 & 0.0571 \\ 0.0434 & 0.9993 & 18.4289 & 4.1838 & 0.9895 & 0.5728 \\ 0.0104 & 0.3829 & 25.8516 & 11.2145 & 3.4419 & 1.8181 \\ 0.0006 & 0.0148 & 0.9357 & 18.6851 & 8.4610 & 4.7885 \\ 0.0003 & 0.0026 & 0.2430 & 15.2442 & 13.9761 & 9.5418 \\ 0.0001 & 0.0001 & 0.0090 & 0.5556 & 14.2377 & 12.1279 \\ 0.0000 & 0.0000 & 0.0010 & 0.0810 & 6.9984 & 10.1813 \\ 0.0000 & 0.0000 & 0.0001 & 0.0023 & 0.2583 & 4.4639 \\ 0.0000 & 0.0000 & 0.0000 & 0.0003 & 0.0230 & 1.0272 \end{pmatrix}$$

求解的真实
多重数分布

$$\begin{pmatrix} w_0 \\ w_2 \\ w_4 \\ w_6 \\ w_8 \\ w_{10} \end{pmatrix}$$

测量 $B(J/\psi \rightarrow l^+ l^-)$ 时的系统误差相消

- 常规方法: $B = \frac{N^{obs}}{\epsilon \cdot N_{tot}}$
- 这里 $B_{ll} = \frac{N_{\pi\pi ll}}{\epsilon_{\pi\pi ll} \cdot N_{tot}} / \frac{N_{\pi\pi}}{\epsilon_{\pi\pi} \cdot N_{tot}} = \frac{N_{\pi\pi ll}}{N_{\pi\pi}} \frac{\epsilon_{\pi\pi}}{\epsilon_{\pi\pi ll}}$
- 因此
 - 总数 N_{tot} 可以消掉
 - 原则上 π 的径迹效率也可以相消, $\epsilon_{\pi\pi ll} = \epsilon_{\pi\pi} \epsilon_{ll}$
 - 只剩下高动量的轻子需要仔细考虑, 而高动量的粒子效率高 (因此误差小)

Sources	$\pi^+\pi^- J/\psi$	e^+e^-	$\mu^+\mu^-$
Tracking	0.80	0.20	0.20
Multiplicity of J/ψ	0.20	0.20	0.20
$M_{\pi^+\pi^-}$ distribution	0.35	0.01	0.01
Background Shape	0.03	0.03	0.04
Fit/Count Range	0.06	0.14	0.14
Bin Size	0.06	0.06	0.06
E/p	—	0.18	0.09
$\cos \theta_{\pi^+\pi^-}$	0.13	0.07	0.07
$\cos \theta_{l^+l^-}$	—	0.04	0.05
FSR effect of l^+l^-	—	0.10	0.23
Fit Method	0.37	0.37	0.37
Trigger	0.10	0.30	0.30
Number of $\psi(3686)$	0.81	—	—
Sum in quadrature	1.28	0.62	0.63

小结（回顾）

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$

推荐阅读：李刚老师关于 $\pi^+\pi^-J/\psi$ 分析的 memo 与 draft
在哪？上 HyperNews 或 Documentation 去找
它们又在哪？bes3.ihep.ac.cn 主页右下角有链接

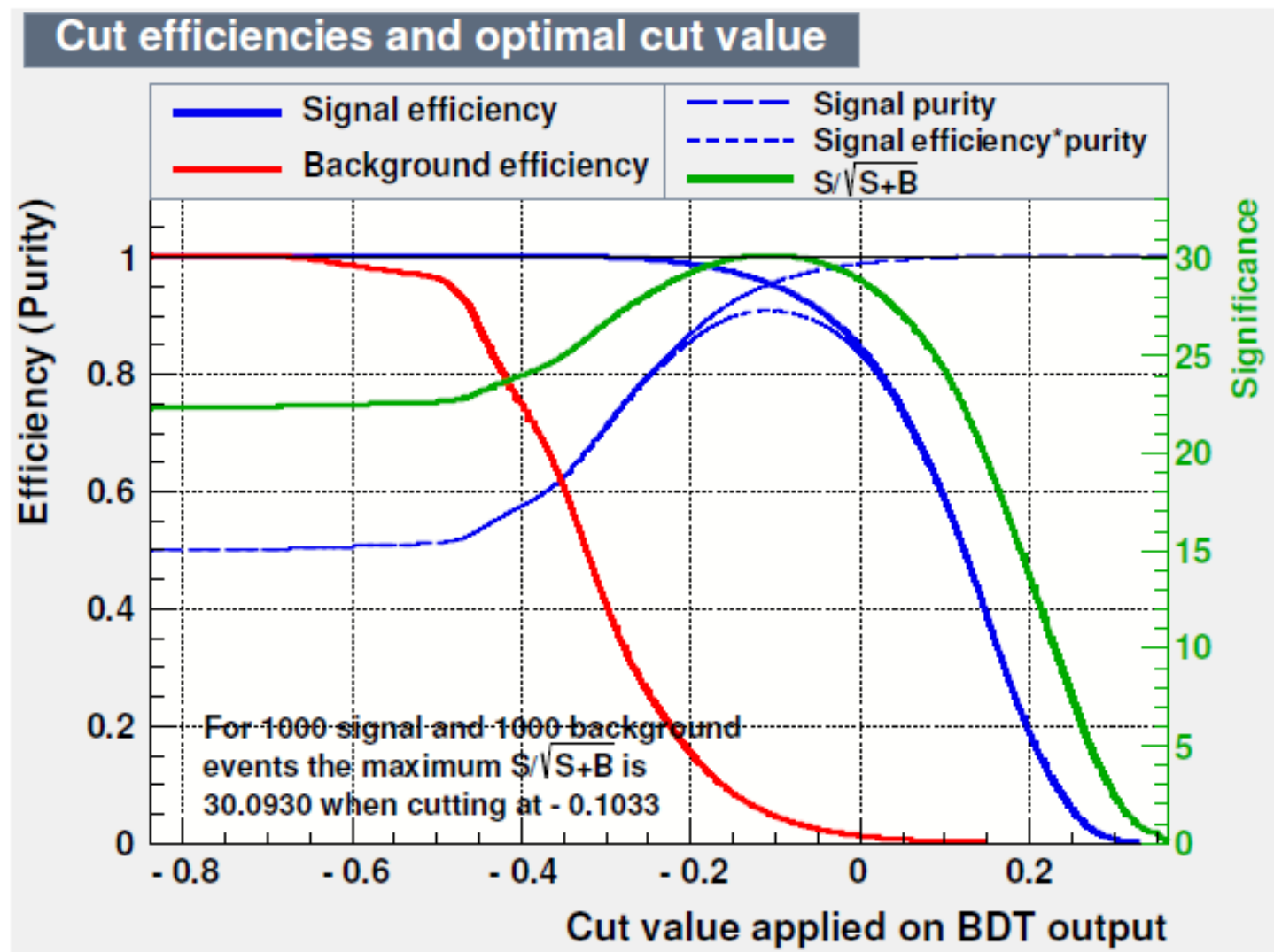
BESIII 物理分析一般流程

- 确定选题（此测量以前没人做过不是一个好动机！）
- 构造事例选择程序，利用信号MC检查各种分布，确定程序正确
- run inclusive MC 了解本底，优化事例选择
- run数据，进一步优化事例选择
- 做各种检查，排除可能的错误
- 分析数据，得到初步结果
- 估计系统误差，得到最终结果
- 撰写memo，会议报告
- 漫长的：物理组内审核，物理软件会，合作组内评阅，全合作组
阅读文章, 英语编辑, 发言人最终审核，杂志评阅，发表

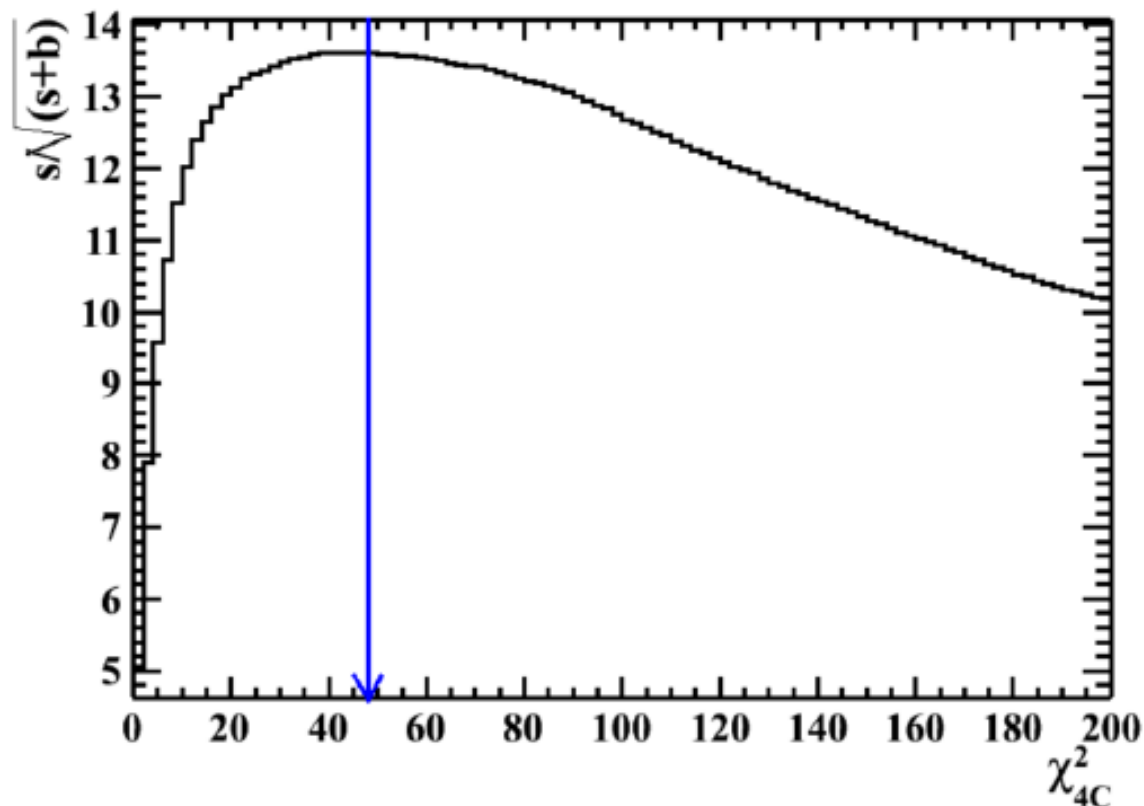
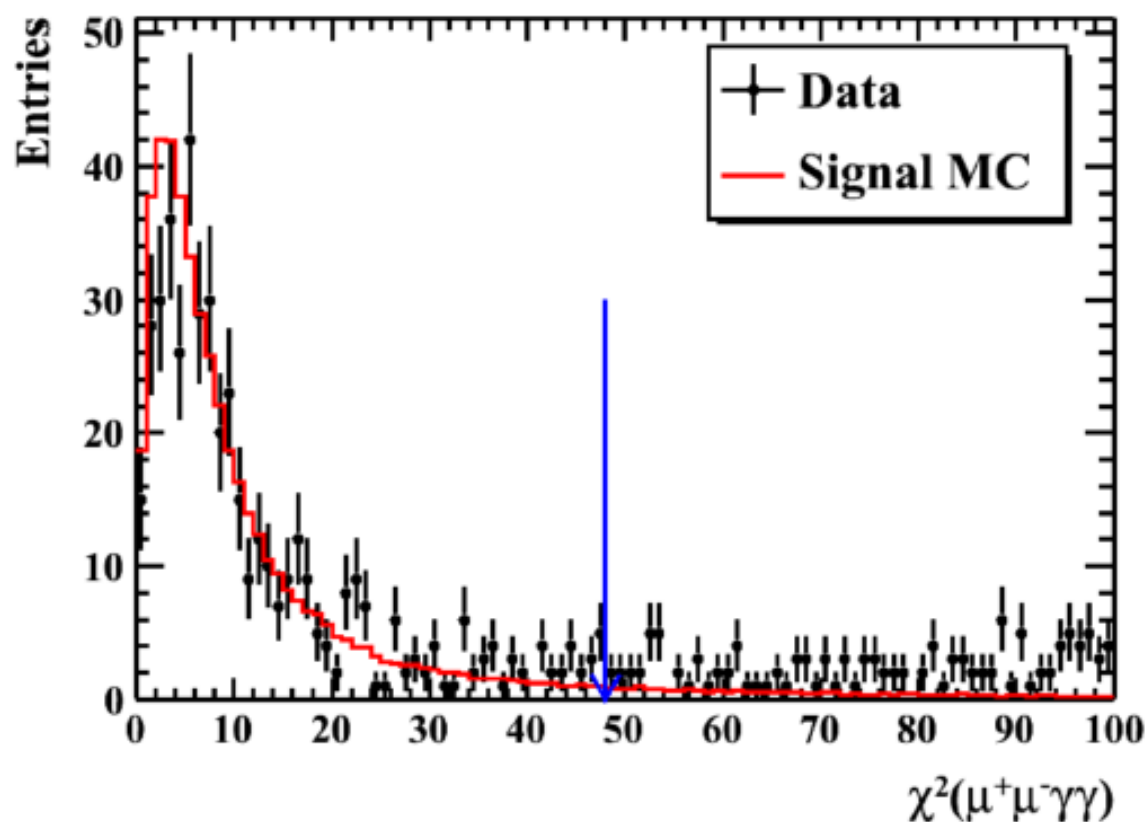
以上过程可（/常需要）迭代

事例选择的优化 (专题)

- 尽可能在保证效率的同时压低本底
 - Why?
 - 提高信号的显著性【significance】
 - 即分析的灵敏度【sensitivity】
 - How?
 - 基于 FOM (figure-of-merit) 的优化
- $$FOM = \frac{S}{\sqrt{S+B}}$$



➤ $\chi^2_{4C} < 48$ ($J/\psi\eta \rightarrow \mu^+\mu^-\gamma\gamma$)

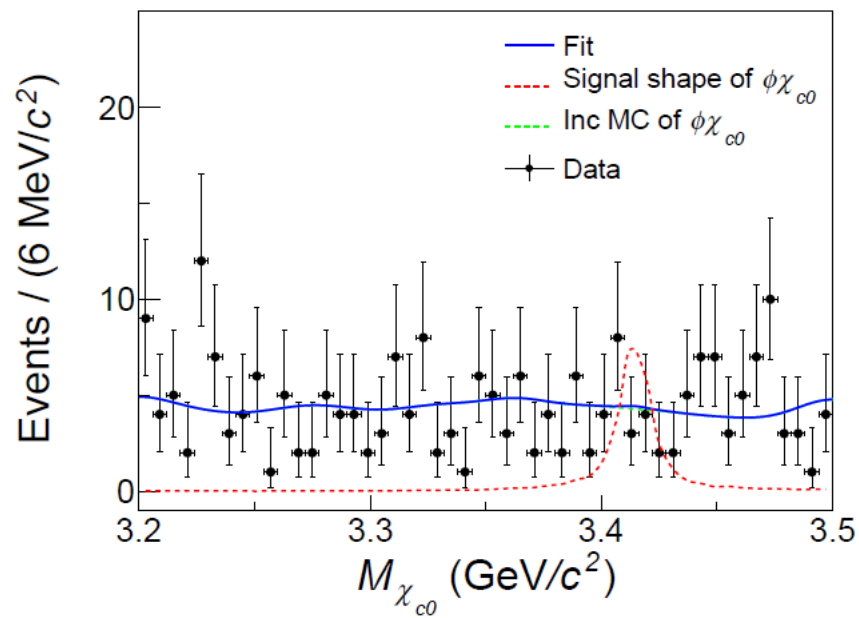


• 注意事项:

1. 计算 FOM 时只针对信号区间的事例
2. 不要使用数据进行优化，特别是小统计量的情况下，以免造成偏差

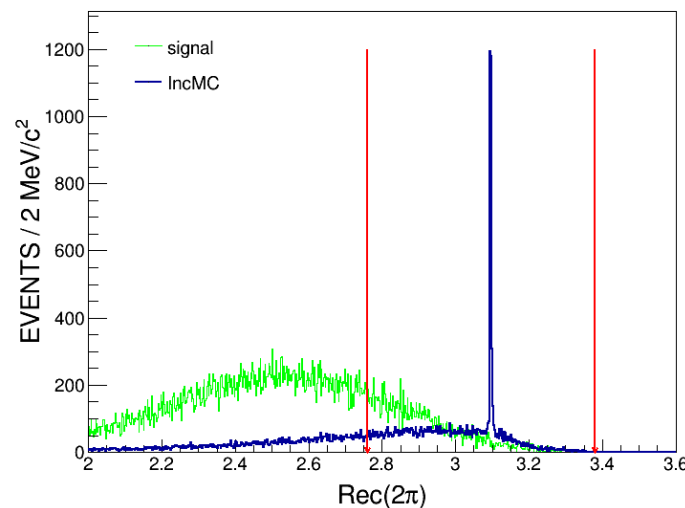
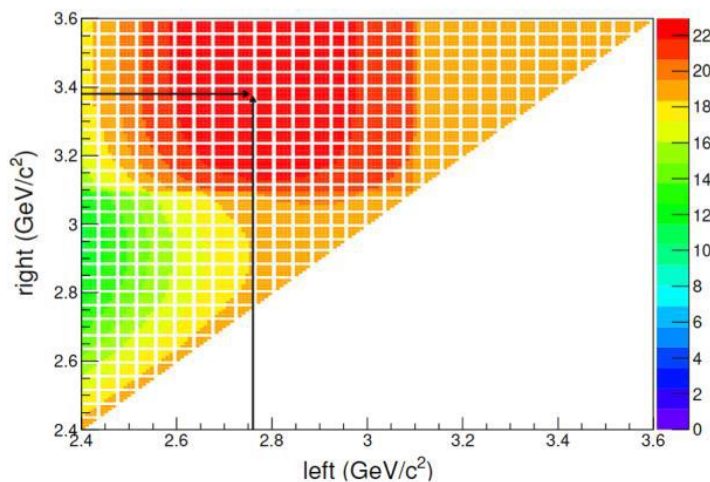
FOM 的变形

- 针对稀有过程（小信号甚至无信号）
 - 不能简单把 $\frac{S}{\sqrt{S+B}}$ 分母中的 S 置为 0
- Punzi method
 - $\frac{S}{\frac{\alpha}{2} + \sqrt{B}}$ (α 对应显著性，一般取 3 或 5)



对FOM的“背叛”

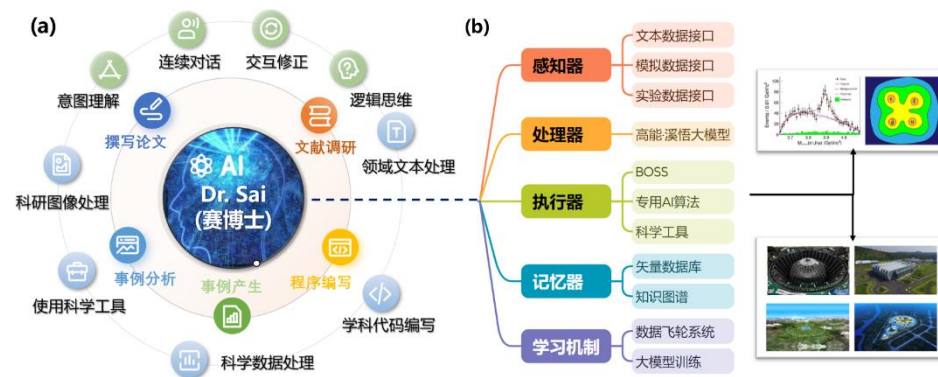
- FOM: 统计上对效率与本底的**平衡**
- 有时候, 当某 cut 损失过多效率引起大的系统误差时, 需要考虑统计显著性与系统误差的**平衡**



AI时代的物理分析

明天下午李老師

- 更简单也更困难危险
 - 多变量分析、深度学习 “黑匣子”
 - 查找资料与自动代码生成 “需要判断力”
- 理解原理！



在人工智能时代，我们正在重新定义物理分析，重新定义研究生教学，还可能要重新定义自我

古典名著对分析工作的启发

1. 独立思考
 2. 不耻“下”问
 3. 学会向上管理，利用好导师
 4. 探究问题根源
-



祝大家物理分析快乐！