

高能物理数据分析简介

苑长征 (yuancz@ihep.ac.cn)

中国科学院高能物理研究所

2026年10月6-7日

目的

- 如何将所学知识应用到具体的实验数据分析
- 进实验室前的基本训练，一般由师徒相传或师兄（师姐）带师弟（师妹）
- 将高能物理实验的范围局限到一个比较小的领域——正负电子对撞实验。
(强子对撞、固定靶实验、中微子实验、重离子碰撞等在此不讨论，但原理相同)
- 主要分为三个部分
 - 实验设备—加速器、探测器与物理过程
 - 实验基础—怎样测量物理量
 - 误差分析—怎样保证测量结果正确

一般说明

- 数据分析很重要
 - 得到正确的中心值
 - 得到合理的误差
 - ➔ 得到正确的物理结论
- 不要（过于）迷信（迷恋）数学方法
 - 物理学家要用物理学家的方法
 - 改变测量方法、改善测量精度和灵敏度
 - 采集更多数据增大统计量，提高信号显著性
 - ➔ 不要 cook data! = 不要对数据超加工!
- 无偏！NO BIAS！
 - 一定要看到信号！（一定没有信号！）— Wrong!
 - 误差一定要比别人小（安全吗？）— Wrong!

授课计划

- 加速器、探测器、物理过程、本底过程
- 物理量的实验测量
 - 亮度测量
 - 截面测量
 - 分支比测量
 - 总数测量
 - 质量、宽度、寿命
 - 量子数确定（角分布测量）
- 误差的估计或确定

穿插概率和统计知识是
如何应用到研究当中的

加速器、探测器、
物理过程、本底过程

加速器/对撞机

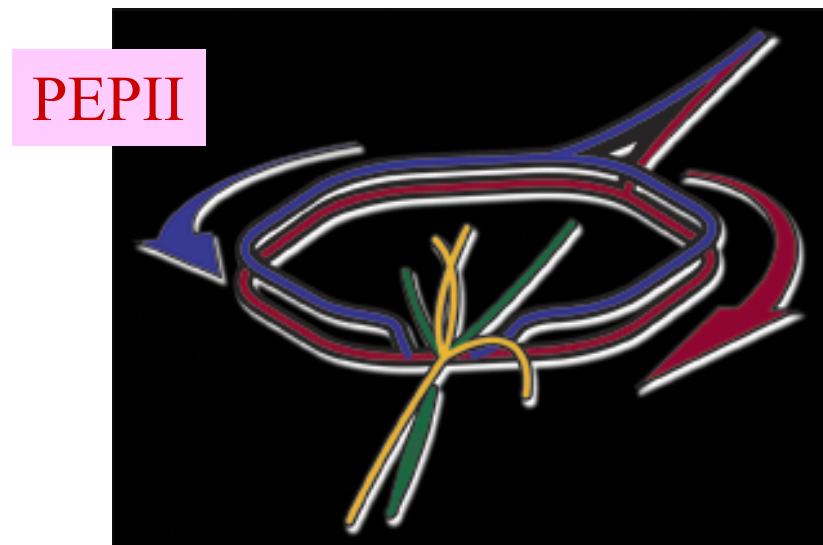
- 单环/双环
- 能量 (0.3-13000 GeV)
- 亮度 (10^{30} - $10^{35}\text{cm}^{-2}\text{s}^{-1}$)

$$1 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1} = 1 \text{ nb}^{-1}\text{s}^{-1}$$

如果物理截面为1 nb，则在亮度为
 $1 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ 的对撞机上每秒产生
一个事例。

- 能散 (1-10MeV)

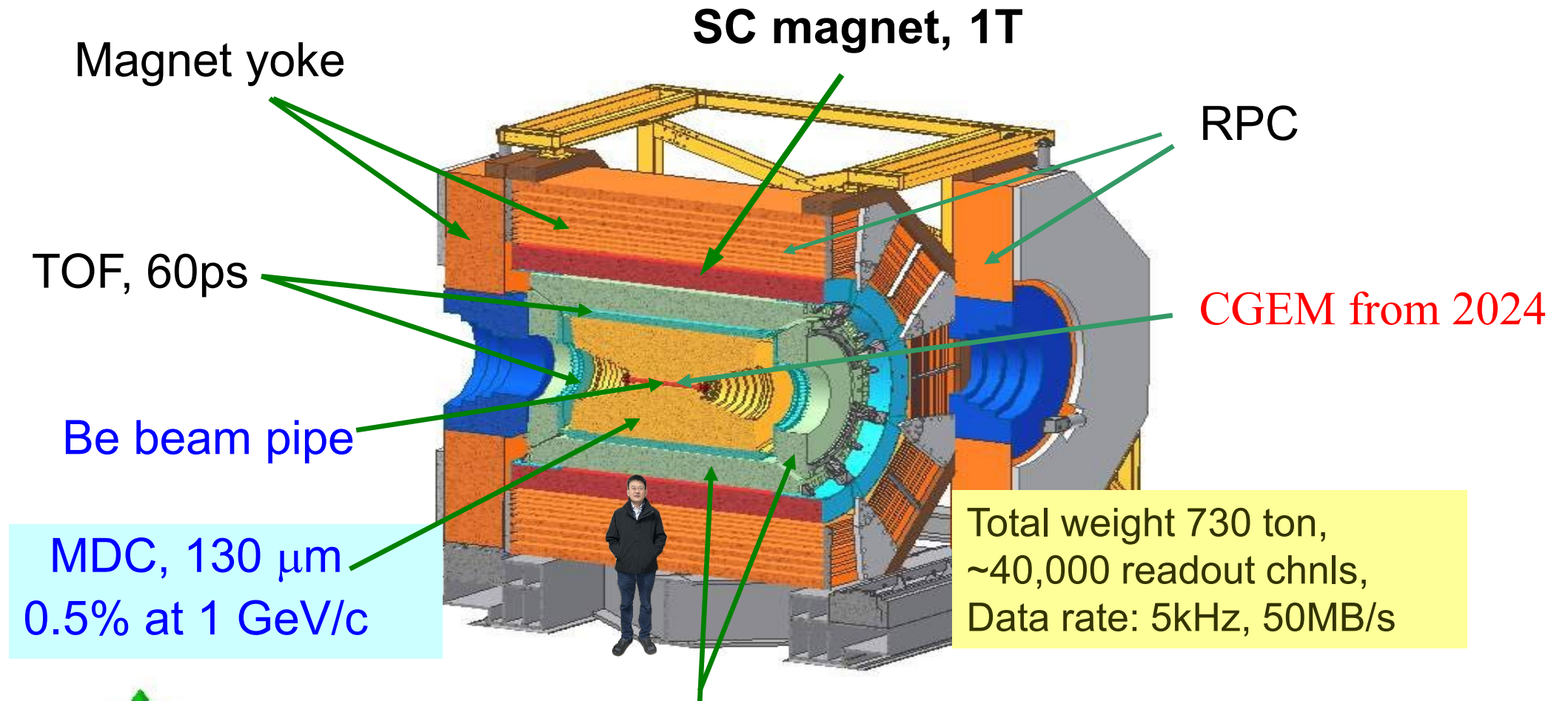
束团中大量粒子的能量分布基本为
高斯分布，标准偏差称为能散。对窄
共振粒子的产额有很大影响。



探测器（磁谱仪）

- 测量粒子四动量 (p, iE) , $E^2=p^2+m^2$
- 动量
 - 带电粒子: Silicon/Drift Chamber/TPC/...
(带电粒子在磁场中偏转。动量分辨0.5-2%)
 - 中性粒子
 - γ : 电磁量能器（取样型/全吸收型）—分辨(~20%--~2%)
 - K_L 、中子: 强子量能器（取样型）—分辨(~80%)
 - 中微子: 磁谱仪中无法探测
- 质量
 - dE/dx : Bethe-Bloch公式 (读音: 贝特-布洛赫公式)
 - TOF: 动量相同的粒子, 质量大(小)飞得慢(快)
 - Mu-Counter: μ 子穿透能力强
 - Rich/DIRC: 契伦科夫光
 - ...

BESIII Detector

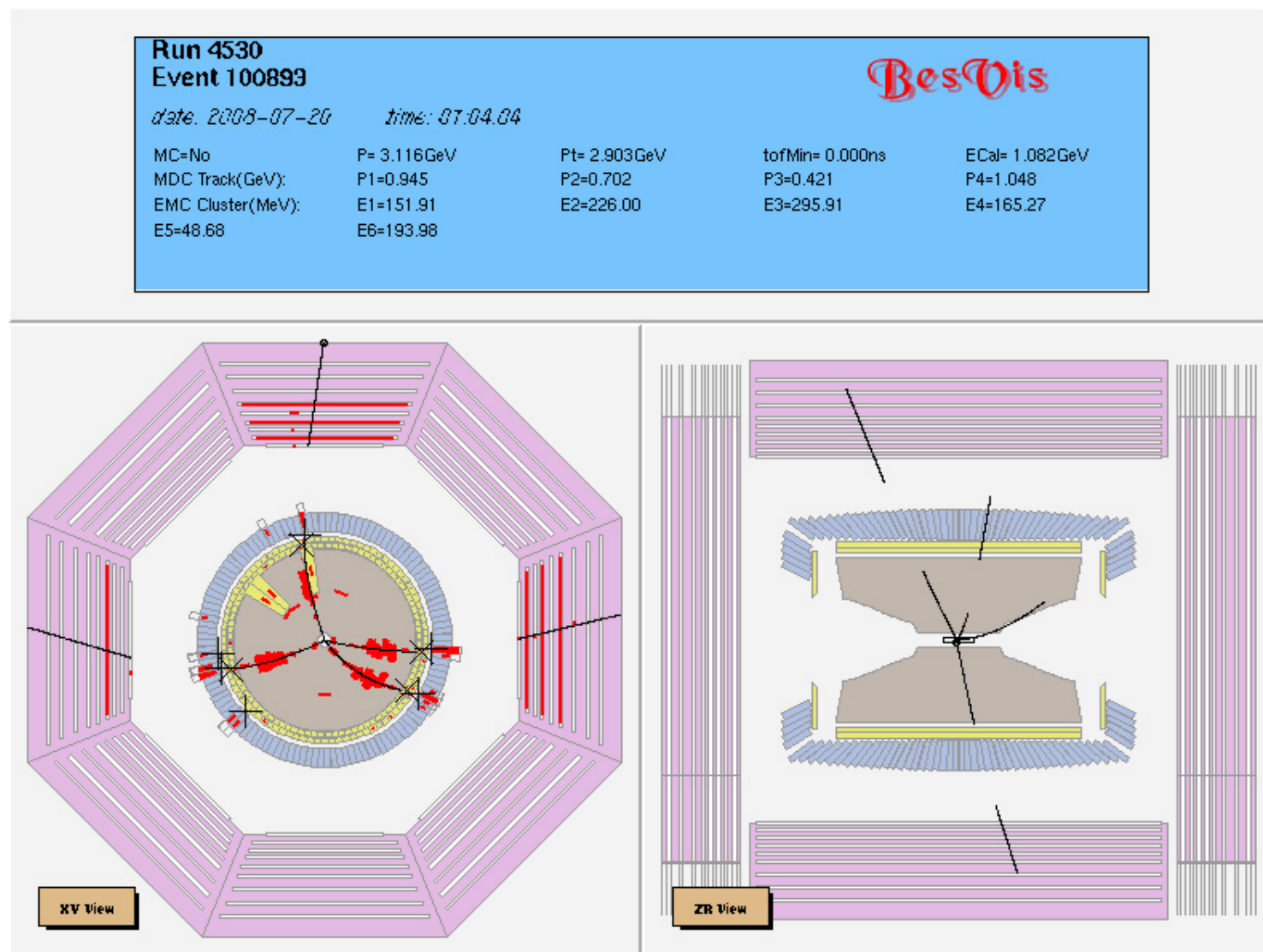


CsI(Tl) calorimeter, 2.5% @ 1 GeV

BESIII detector



典型对撞事例：2008年7月20日



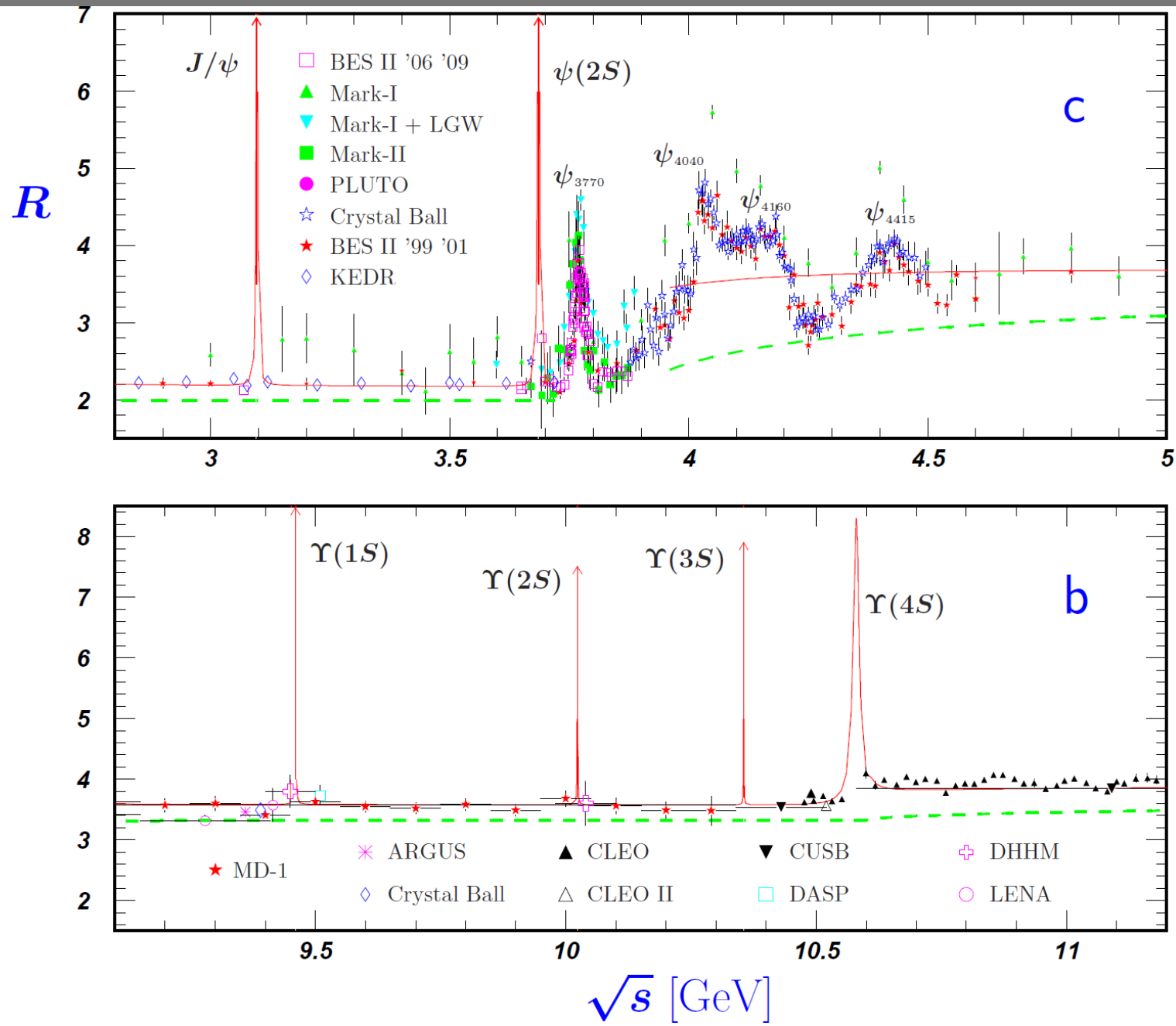
Events in detector

- e^+e^- annihilation
- Synchrotron radiation
- Beam-gas
- Beam wall (e-p, e-n, ...)
- Beam-beam interaction
- Lost beam particles
- Electronic noise (junk event)
- Nuclear interaction (final state particle + detector)
- Cosmic rays (angular distribution, rate)
- Air shower

Physics processes far below Z-pole

- QED
 - Bhabha
 - Di-mu
 - Di-tau
 - Di-gamma
- Hadrons at continuum (R-measurement)
 - Continuum production of hadrons
 - Two-photon processes
 - ISR (initial state radiation)

R values



Physics processes

- Resonance
 - BW (Breit-Wigner)
 - Radiative correction
 - Beam spread
 - Resonance scan for parameters
- Resonance decays
 - Various decay modes (below/above open charm)
- Possible interference between resonance and continuum
 - Irreducible background and impact on BF measurements
 - Phase between EM and strong amplitudes

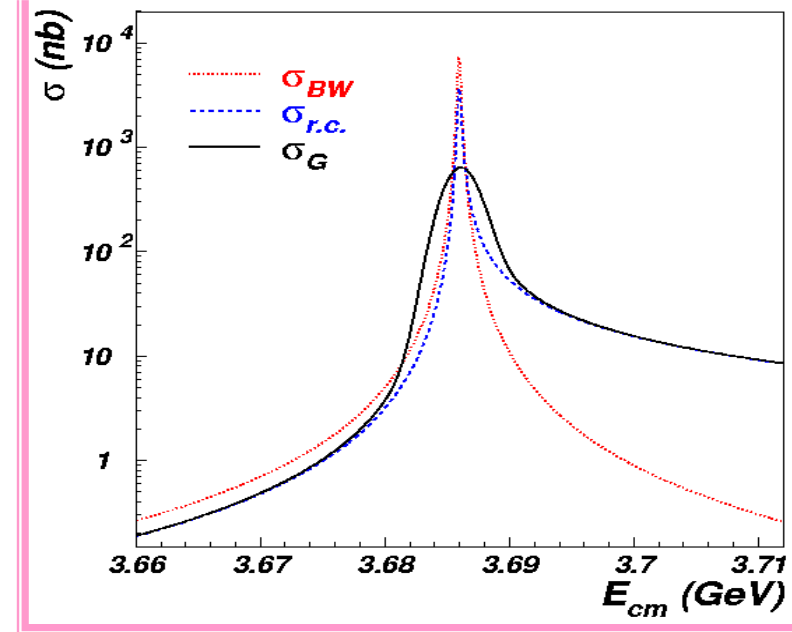
At resonance peak

$$\sigma_{BW}(W) = \frac{12\pi \cdot \Gamma_e \Gamma_f}{(W^2 - M^2)^2 + \Gamma_t^2 M^2}$$

$$\sigma_{r.c.}(W) = \int_0^{x_m} dx F(x, s) \frac{1}{|1 - \Pi(s(1-x))|^2} \sigma_{BW}(s(1-x))$$

$$\sigma_{exp}(W) = \int_0^\infty dW' \sigma_{r.c.}(W') G(W', W)$$

$$G(W, W') = \frac{1}{\sqrt{2\pi}\Delta} e^{-\frac{(W-W')^2}{2\Delta^2}}$$

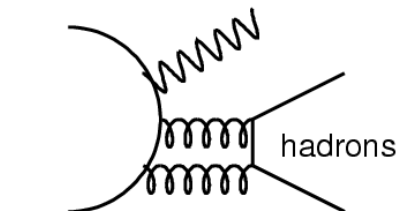
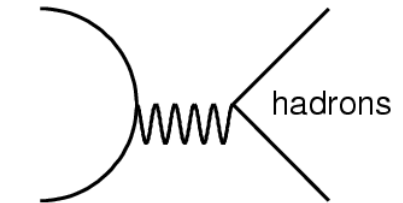
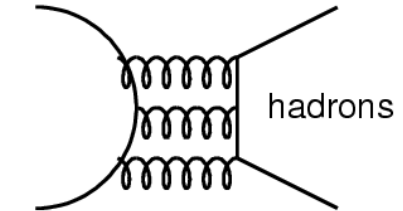
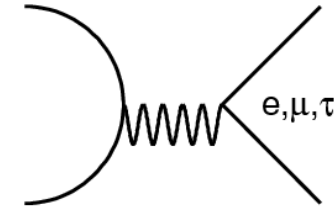
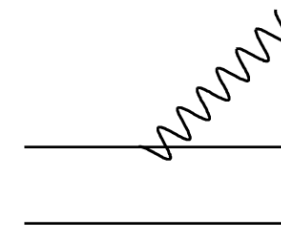
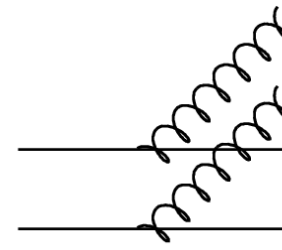


At $\psi(2S)$	Born	ISR	$\Delta=1.3\text{MeV}$
σ_{RES} (nb)	7887	4046	640
σ_{CON} (nb)	~14	~14	~14

Continuum contribution becomes larger after considering ISR and beam spread!

ψ' decays

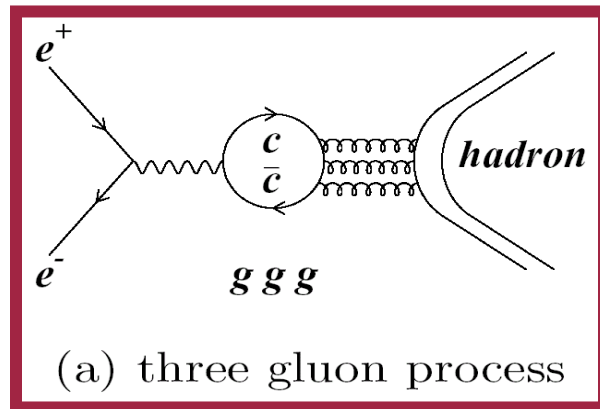
- Transitions ($\sim 82\%$)
 - Hadronic transitions ($\sim 54\%$)
 - Radiative transitions ($\sim 28\%$)
- Leptonic decays ($\sim 2\%$)
- Hadronic decays ($\sim 15\%$)
 - Strong decays ($\sim 13\%$)
 - EM decays ($\sim 2\%$)
- Radiative decays ($\sim 1\%$)
- Rare decays and beyond SM ($\ll 1\%$)



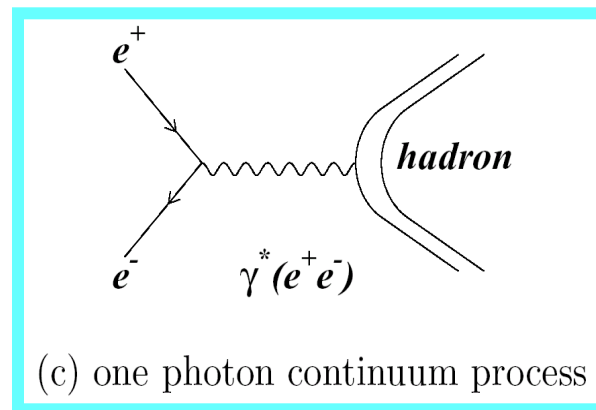
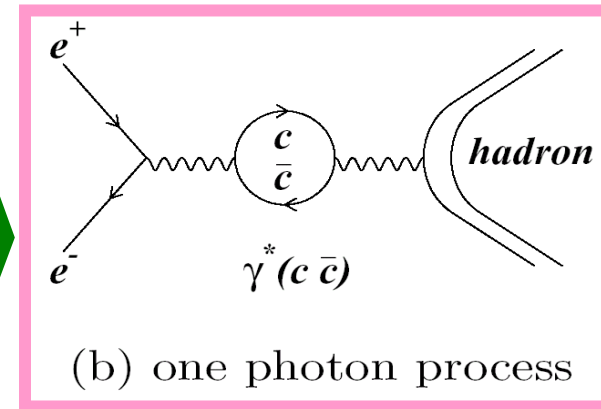
Similar for J/ψ decays, but BR different.

Continuum amplitude is important for high precision BF measurement

Branching fractions



phase



Branching fraction

$$B = \frac{n^{obs}/\varepsilon}{N_{tot}} \quad (?)$$

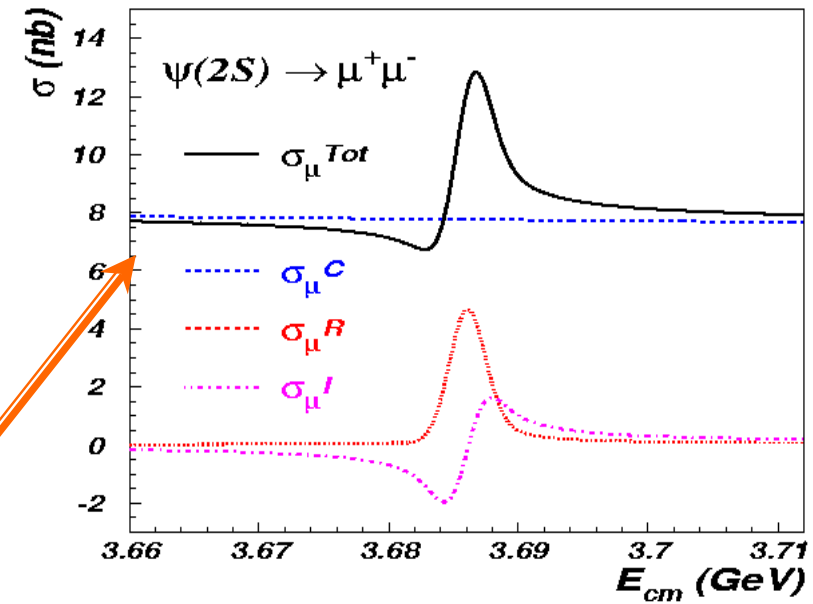
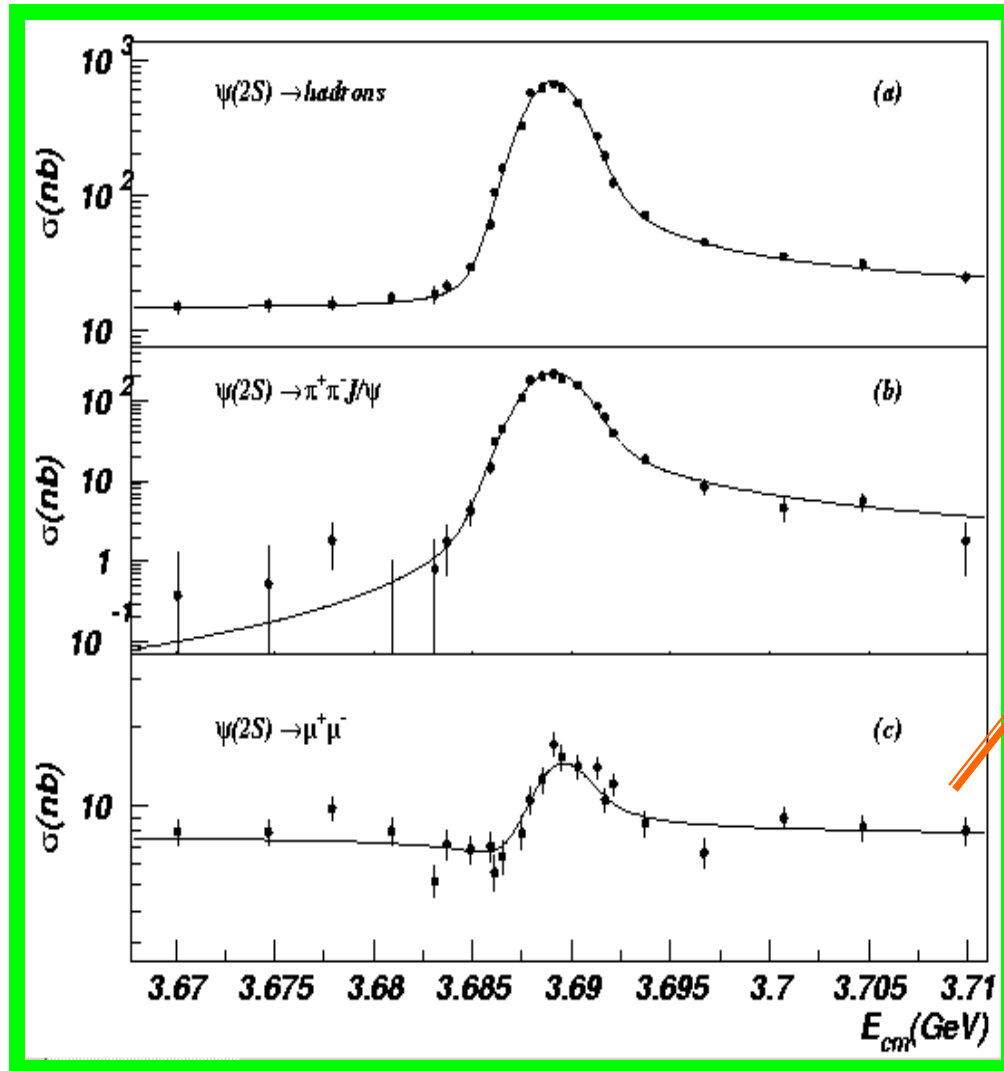
$$B = \frac{(n^{obs} - n^{cont})/\varepsilon}{N_{tot}} \quad (?)$$

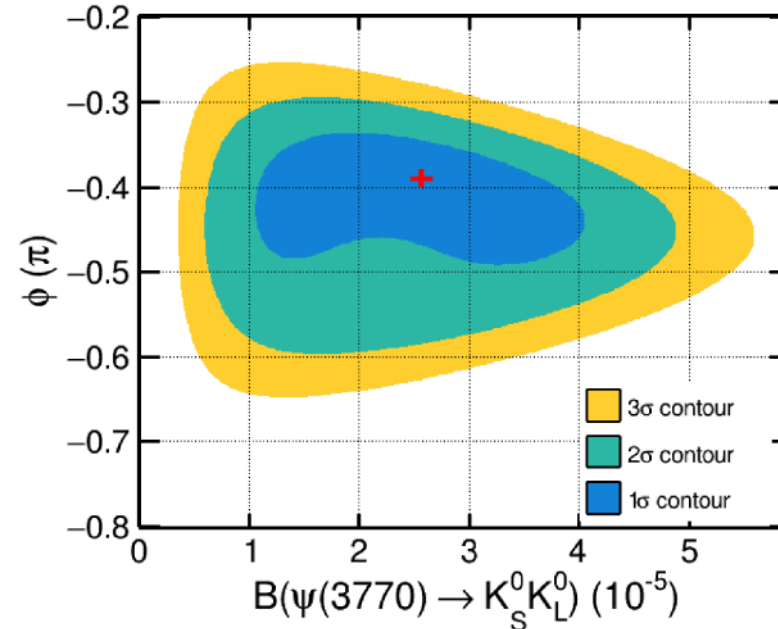
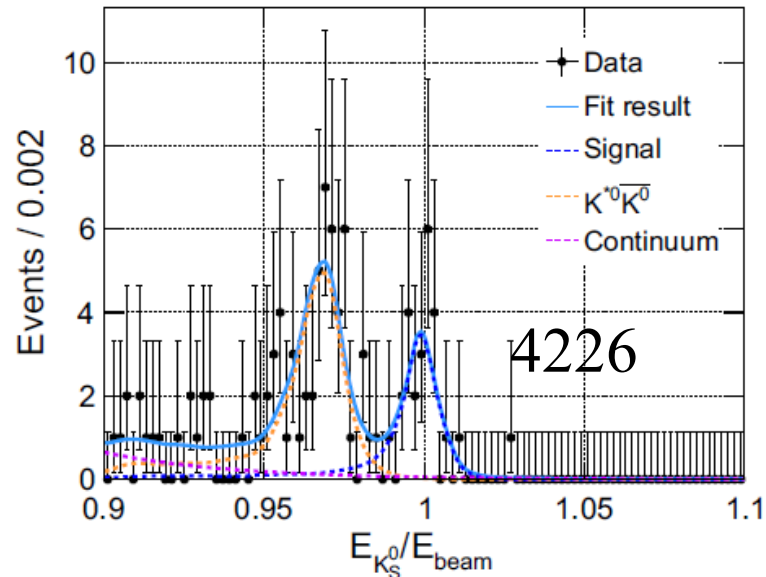
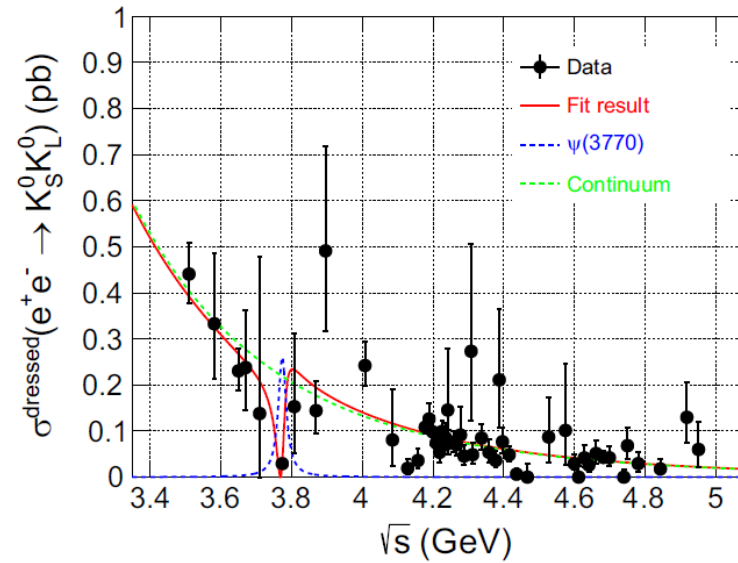
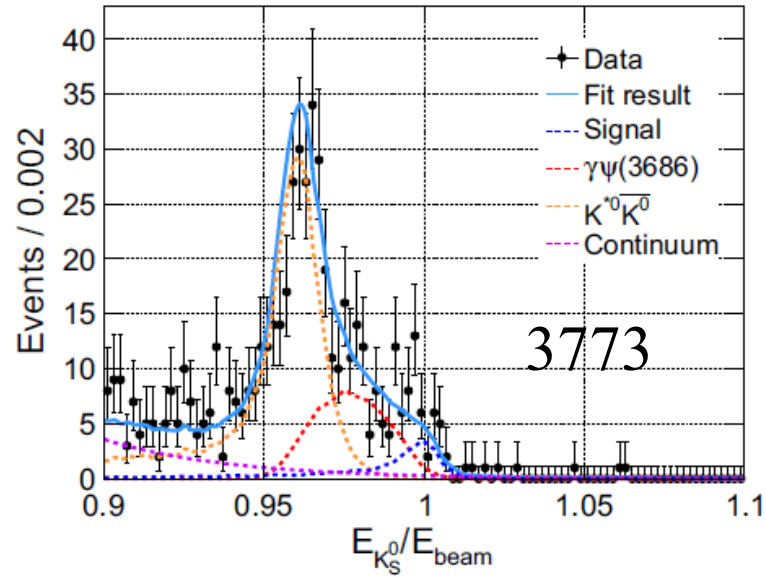
[eg. ψ rare decays, EM decays, strong decays, transitions]

Interference between continuum amplitude and resonance amplitude is very complicated, needs to be very careful.

[Y. P. Guo and C. Z. Yuan, PRD 105, 114001 (2022)]

$e^+e^- \rightarrow \psi(2S) @ \text{BESII}$



10 σ First ψ'' light
hadron decay!

数据采集

- 信号与触发
 - Cosmic rays
 - Electronic noise
 - Beam-gas interaction
 - Beam-wall interaction
 - Other non-physics events
 - Reduce QED background
- 触发系统 (trigger)
 - 尽量降低本底, 但保持高效率 (接近100%)
- 数据采集 (DAQ)
 - 记录通过触发的事例的电子学读出信号

数据刻度与重建

- **刻度**：寻找电子学信号与物理量之间的比例关系（或复杂的函数关系）——特征过程或特殊粒子的选择。
 - TOF记录的时间与粒子飞行时间之间的关系
 - EMC电荷量与光子能量的对应关系
- **重建**：将探测器记录的信息通过刻度常数，转化为粒子四动量的过程。
 - 由MDC记录的位置到带电径迹动量的大小
 - 由TOF时间或 dE/dx 大小到粒子质量
 - 由EMC电荷量到光子能量
 - 各子探测器信息的匹配等

数据里包括什么

- 物理过程/本底 (cosmic ray, beam-gas,...)
- 物理过程互为本底



海、天、岛、
树、人
互为本底！

数据分析的流程

- 物理目标（研究动机：物理+老师要求）
 - 测量什么物理量（截面，质量，宽度， I^{GJPC} ，角分布，分支比，寿命等）
 - 如何测量（设计实验或用已采集的数据）
- 信号与本底—选择条件确定
 - Monte Carlo模拟的应用（确定条件，确定本底）
- 系统误差的可能来源与测量（估计）
- 得到结果
- 由结果推出结论 → 物理目标实现了吗？

Monte Carlo 模拟

- Monte Carlo 模拟在高能物理实验中广泛应用
 - 物理事例产生子，产生末态粒子四动量
 - 粒子与探测器物质发生相互作用
 - 探测器响应的模拟
 - 触发、刻度、重建
 - 与数据进行同样分析，可得到效率、分辨率等
- 高能物理实验依赖Monte Carlo模拟
- Monte Carlo模拟的有效性须数据证明——数据分析和误差处理的重要内容

物理量的实验测量

亮度测量（最基本的参数）

$$N = \sigma \bullet L$$

瞬时亮度

积分亮度

$$L = \frac{N}{\sigma} = \frac{N_f^{obs} - N_{bkg}}{\sigma(e^+e^- \rightarrow f)\varepsilon_f}$$

N的误差

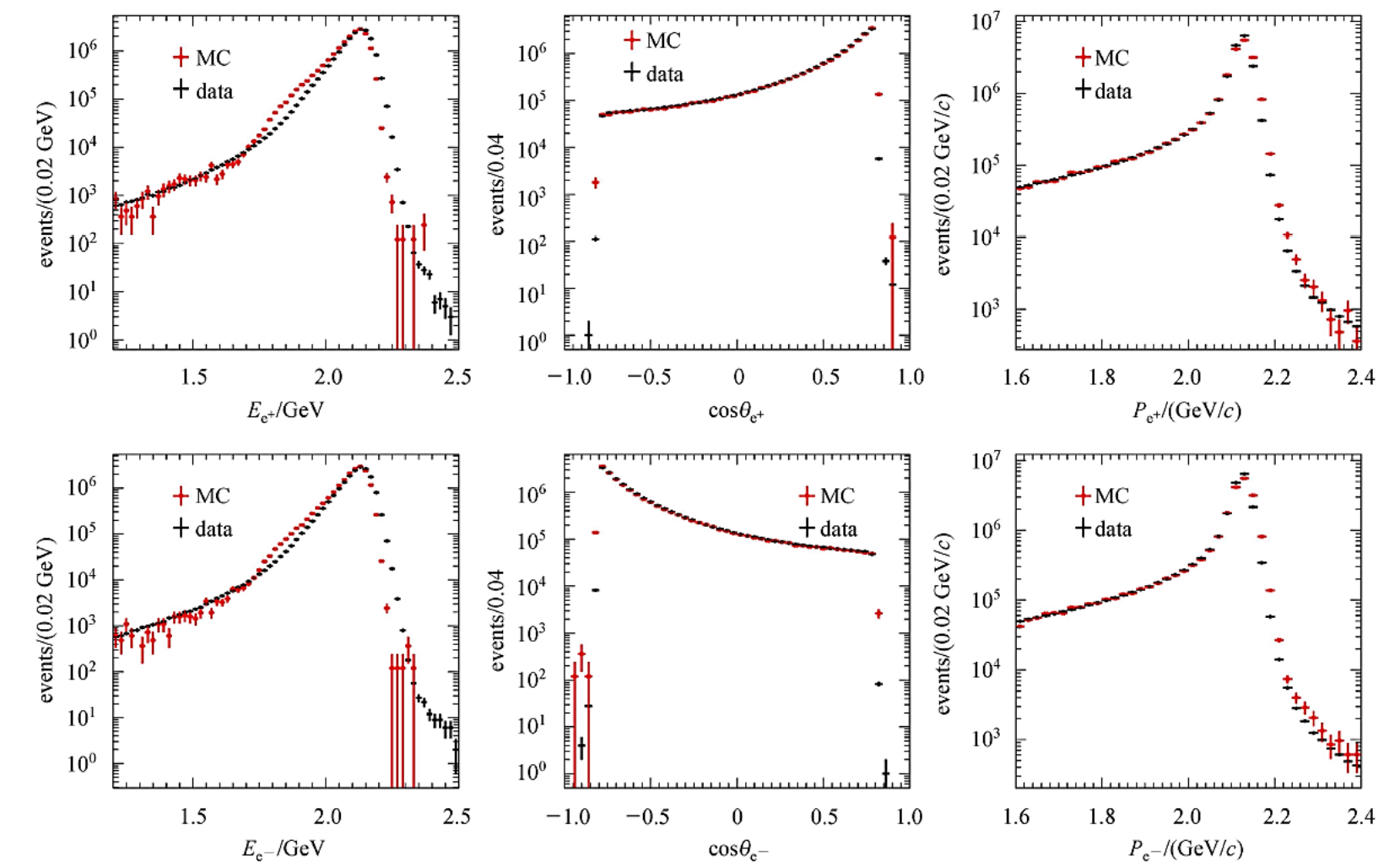
σ 的误差

ε 的误差

L的误差

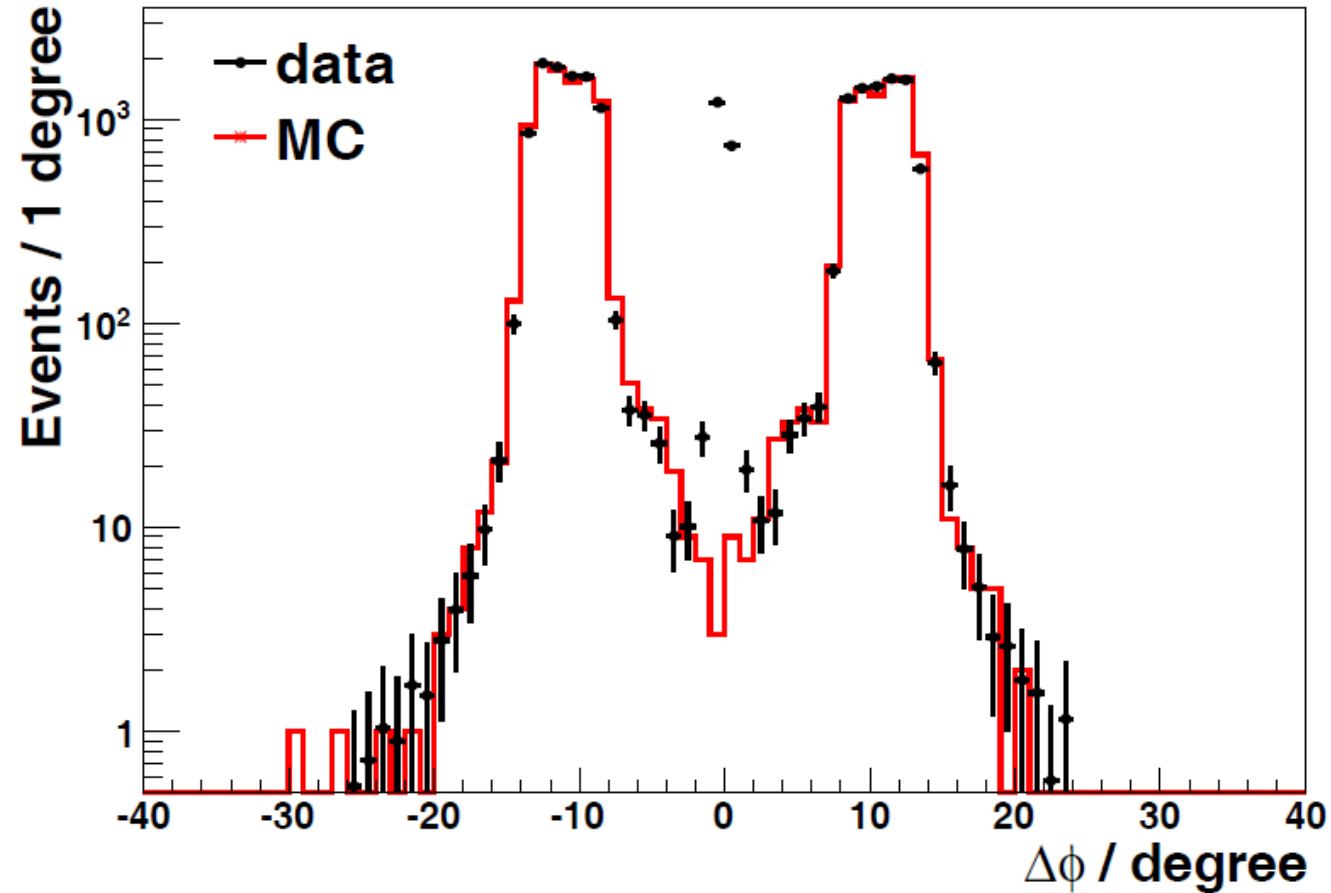
- 分布、误差传递
- 事例率估计
- 得到高精度测量结果

亮度测量 (Bhabha散射, $E_{\text{cm}}=4.26$ GeV)



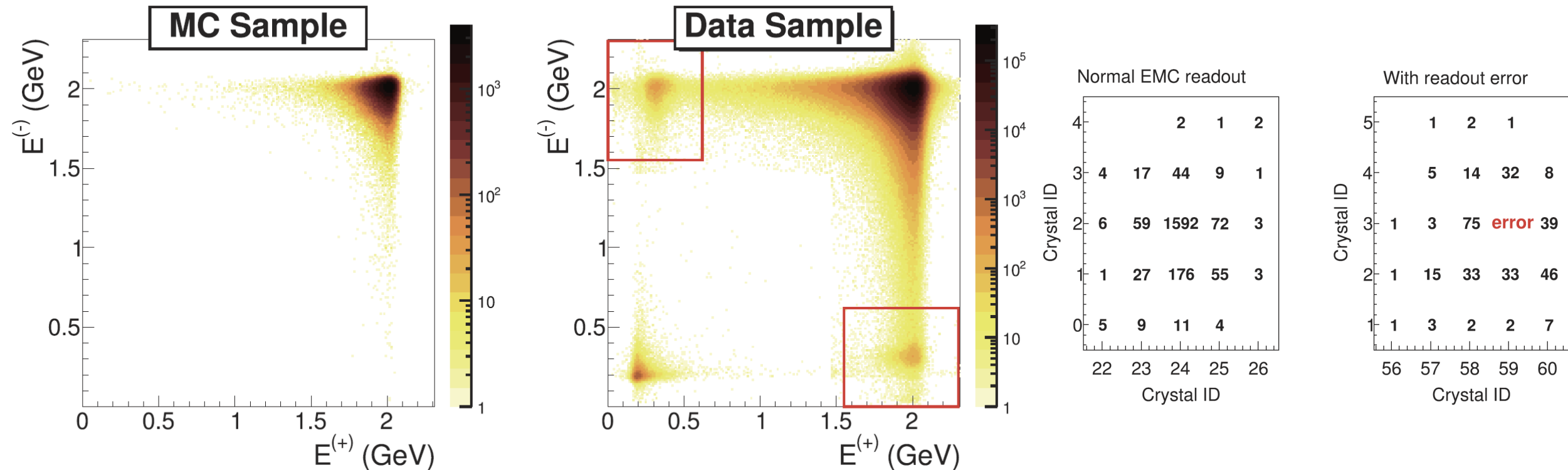
BESIII: Chin. Phys. C 39, 093001 (2015)

亮度测量 ($\text{Bhabha}+\gamma\gamma$)



$\Delta\phi=|\phi_1-\phi_2|-\pi$, ϕ measured in EMC

亮度测量 (Bhabha+ $\gamma\gamma$)



MC doesnot simulate data well, in some cases

系统误差分析并不是随便引用一个数字

BESIII: Chin. Phys. C 46, 113002 (2022)

截面测量：遍举过程

$$\sigma(e^+e^- \rightarrow f) = \frac{N_f^{obs} - N_{bkg}}{\epsilon_f \cdot \epsilon_{trg} \cdot (1 + \delta) \cdot L}$$

- 电磁形状因子的测量

$$\sigma \propto |\mathbf{F}|^2$$

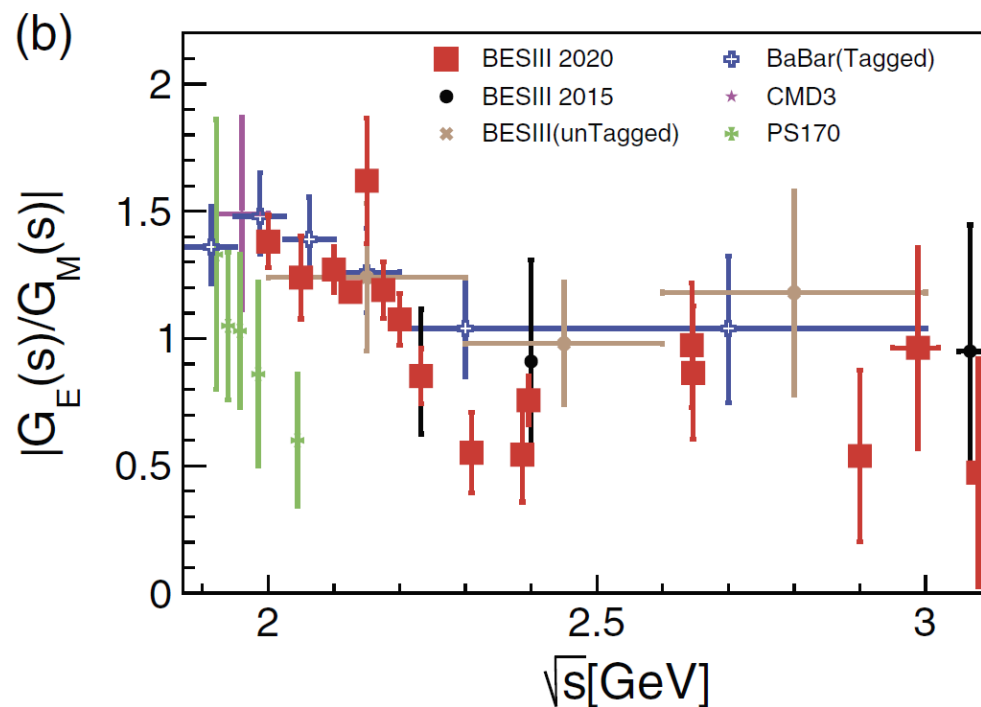
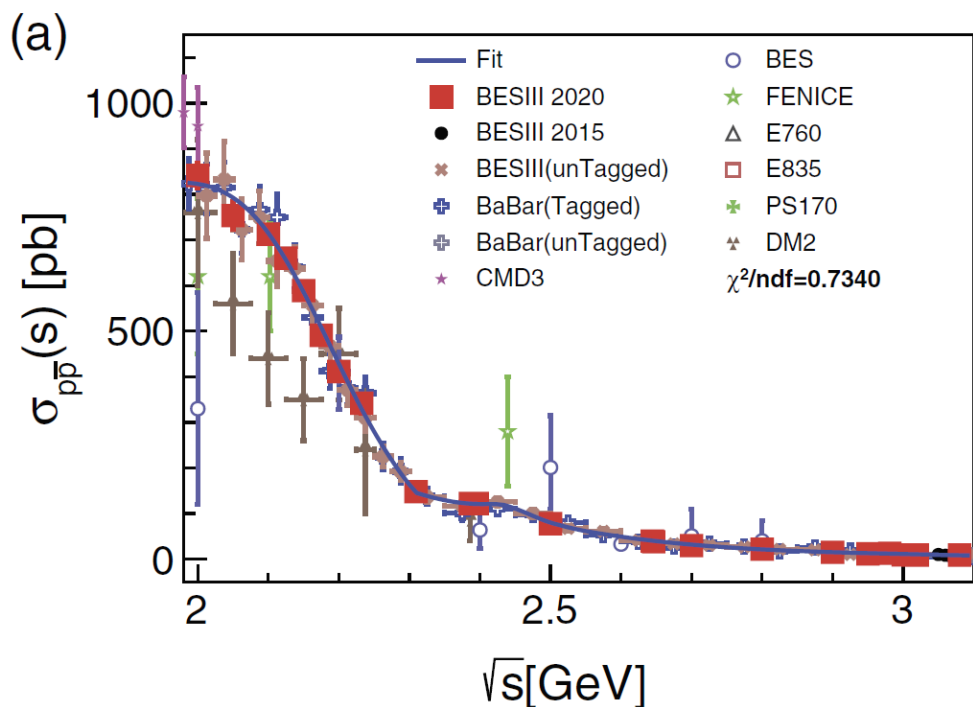
- 矢量粒子性质 (BW) , 粒子产额
- 初态辐射 (ISR) 数据的分析
 - Take data at s , but do physics at $s' < s$!
 - BaBar, Belle, KLOE, Belle II, BESIII

BESIII: PRL124, 042001 (2020)

BaBar: PRL103, 231801 (2009)

截面测量：遍举过程

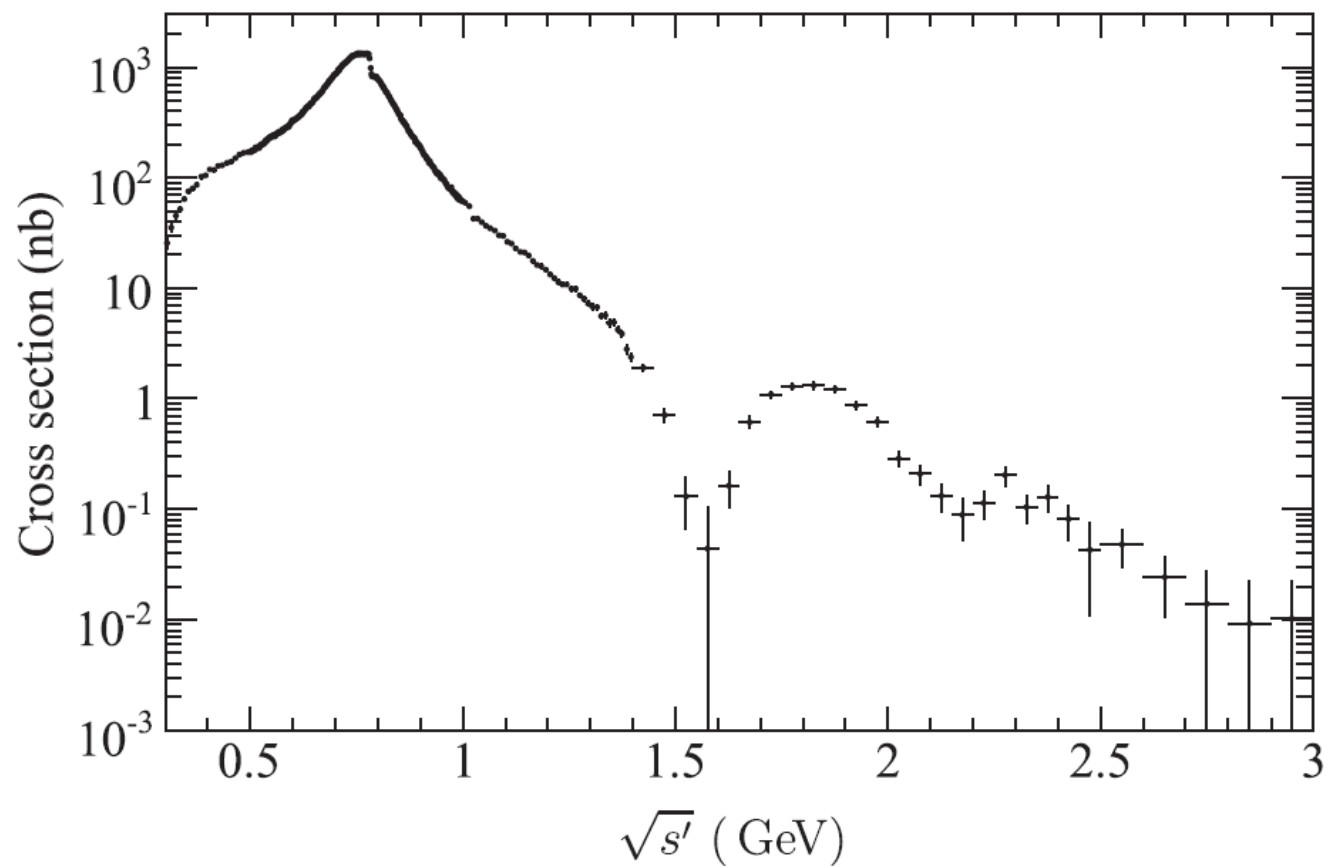
- BESIII $e^+e^- \rightarrow \bar{p}p$
- 核子结构：form factor?
- 与亮度误差部分抵消？



BESIII: PRL124, 042001 (2020)

截面测量：遍举过程

- BaBar $e^+e^- \rightarrow \pi^+\pi^-$; 与 $\mu\mu$ 过程比值?
- 真空极化 $\rightarrow a_\mu, \alpha$, 新物理



BaBar: PRL103, 231801 (2009)

截面测量：R值

- 研究动机—强子对真空极化的影响
 - Running α ; $(g-2)_\mu$; 新粒子的圈图效应
- **测量量**: $R = \sigma(e^+e^- \rightarrow \text{强子}) / \sigma^0(e^+e^- \rightarrow \mu^+\mu^-)$
- $\sigma(e^+e^- \rightarrow \text{强子}) = \sum \sigma(e^+e^- \rightarrow f)$, 为什么不测f?
- 公式: $\sigma = N/L$

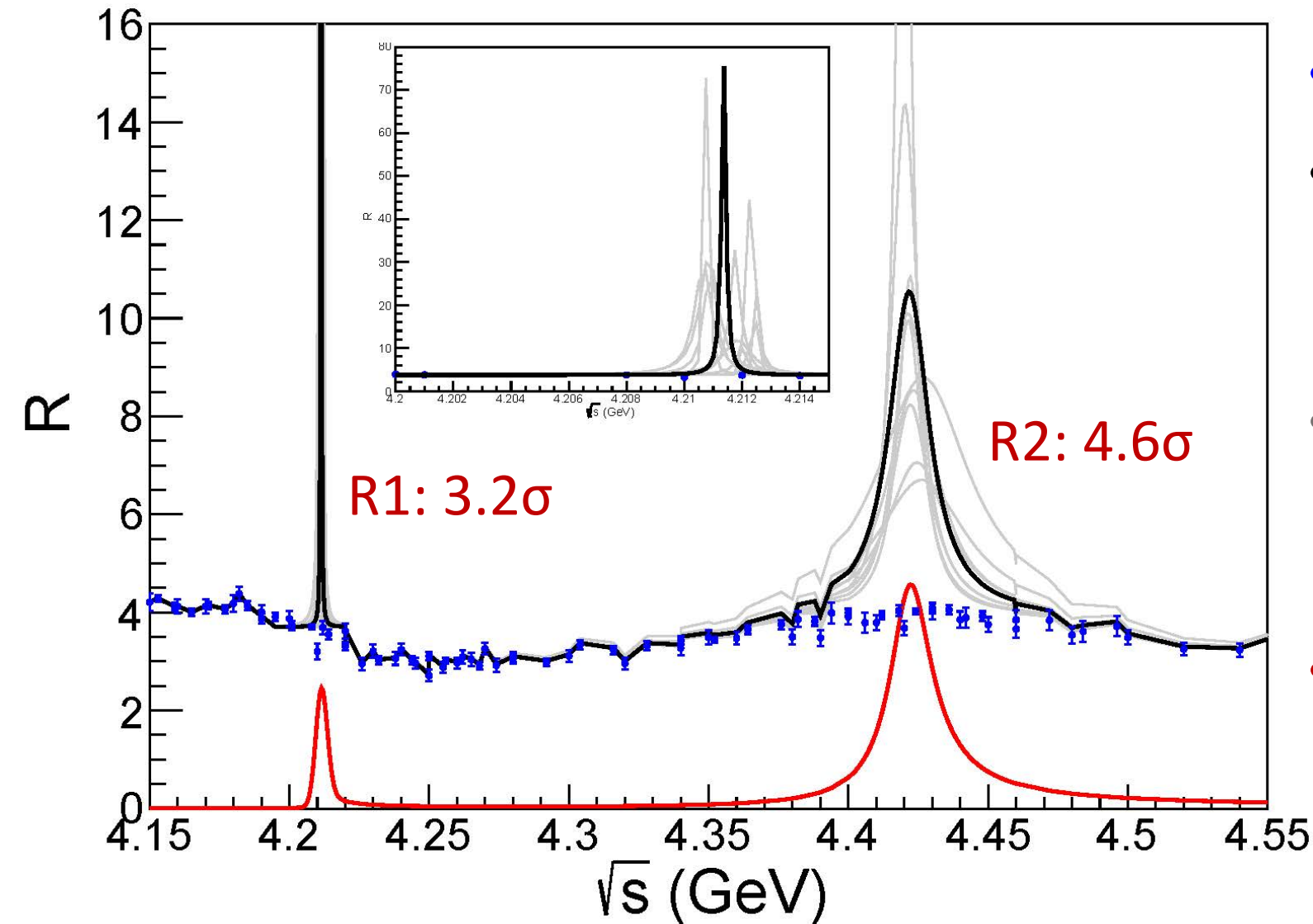
$$R = \frac{N_{had}^{obs} - N_{bkg} - \sum_l N_{ll} - N_{\gamma\gamma}}{\sigma_{\mu\mu}^0 \cdot \epsilon_{had} \cdot \epsilon_{trg} \cdot (1 + \delta) \cdot L}$$

截面测量：R值

$$R = \frac{N_{had}^{obs} - N_{bkg} - \sum_l N_{ll} - N_{\gamma\gamma}}{\sigma_{\mu\mu}^0 \cdot \epsilon_{had} \cdot \epsilon_{trg} \cdot (1 + \delta) \cdot L}$$

- N_{had} 的测量—从所有事例中扣除非强子事例
 - 强子：总能量大，能量平衡，径迹数多，来自对撞点
 - 非对撞本底：总能量小，能量不平衡
 - 轻子对：径迹数少，背对背
 - 双光子过程：总能量小
 - 选择后剩余本底由Monte Carlo模拟估计
- 效率和辐射修正因子由Monte Carlo模拟估计
- 亮度利用QED过程测量（Bhabha/dimu/two-gamma）

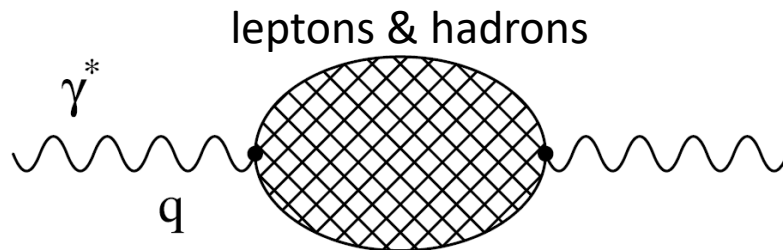
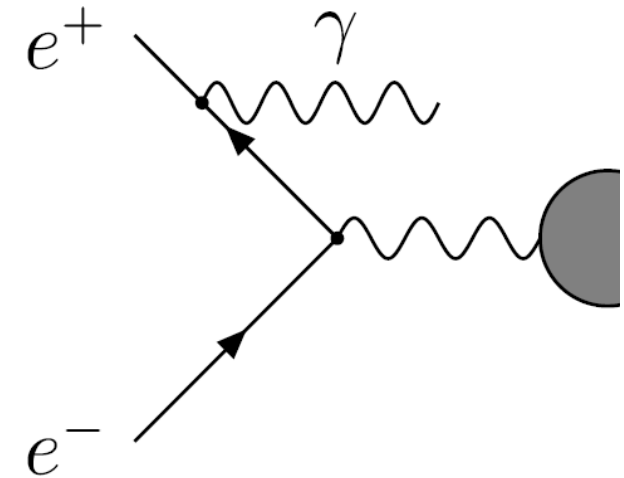
Predicted vs. **measured** $R = \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}$



- Blue points: $R_{\text{had}}^{\text{PDG}}$
- Black curve: predicted R_{had} based on measured $e^+e^- \rightarrow \mu^+\mu^-$ cross sections from BESIII
- Gray curves: 10 randomly drawn solutions from the covariance matrix.
- Red curves show R1 & R2 as they would be observed, due to beam energy spread and ISR smearing.

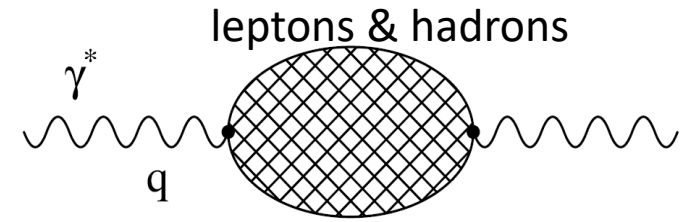
Two ways of investigating total e^+e^- cross sections

- ISR from high CM energy (BaBar, Belle, Belle II, ...)
 - Single photon trigger
 - Photon energy resolution
- $e^+e^- \rightarrow \mu^+\mu^-$ cross sections
 - Vacuum polarization includes contribution from all hadronic final states
 - High precision measurement $\rightarrow$ predicted R_{had}



We compare the predicted R_{had} calculated from $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ with that from experimental measurement to search for invisible states in e^+e^- annihilation
 $\equiv$ Compare measured $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ with that calculated with $R_{\text{had}}^{\text{PDG}}$

Calculate $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$



$$\sigma_{\mu^+\mu^-}(s) = \frac{\sigma_{\mu^+\mu^-}^B(s)}{|1 - \Pi(s)|^2}$$

$$\sigma^B(e^+e^- \rightarrow \mu^+\mu^-) = \frac{4\pi\alpha^2}{3s}$$

$$\Pi(s) \equiv \sum_{j=e,\mu,\tau} \Pi_l(s, m_j^2) + \Pi_h(s)$$

QED

Hadronic cross sections, use experimental data
+ pQCD@continuum energy region

$$\Pi_h(s) = -\frac{s}{4\pi^2\alpha} P \int_{4m_\pi^2}^{\infty} \frac{\sigma(s')}{s' - s} ds' - i \frac{s}{4\pi\alpha} \sigma(s)$$

$$\begin{aligned} \Re \Pi_h(s) = & \frac{3s}{\alpha} \sum_j \frac{\Gamma_{e^+e^-}^j}{M_j} \frac{s - M_j^2}{(s - M_j^2)^2 + M_j^2 \Gamma_j^2} \\ & + \frac{\alpha}{3\pi} R(s_1) \ln \left| \frac{s - s_1}{s_1} \right| \\ & - \frac{s}{4\pi^2\alpha} \int_{4m_\pi^2}^{s_1} \frac{\sigma_{\text{nr}}(s') - \sigma_{\text{nr}}(s)}{s' - s} ds' \\ & - \frac{s\sigma_{\text{nr}}(s)}{4\pi^2\alpha} \ln \left| \frac{s_1 - s}{4m_\pi^2 - s} \right|, \end{aligned}$$

分支比测量

- 分支比计算公式

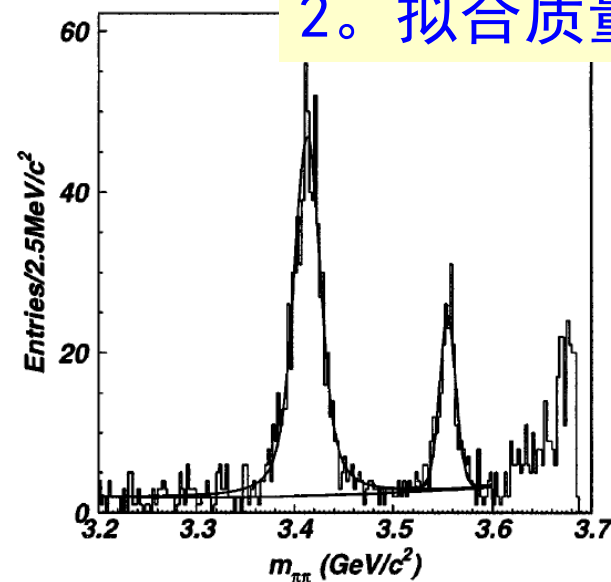
$$B(X \rightarrow f) = \frac{n^{obs} - n^{bkg}}{\epsilon N_X}$$

- 影响分支比的因素

1. 观测事例数
2. 本底事例数
3. 选择效率
4. 事例总数

本底的估计:

1. 数数
2. 拟合质量谱



关于效率

- 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、探测末态的分支比等。
- 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。一般情况下这个假设不成立。
- 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。

关于效率

- 一般地

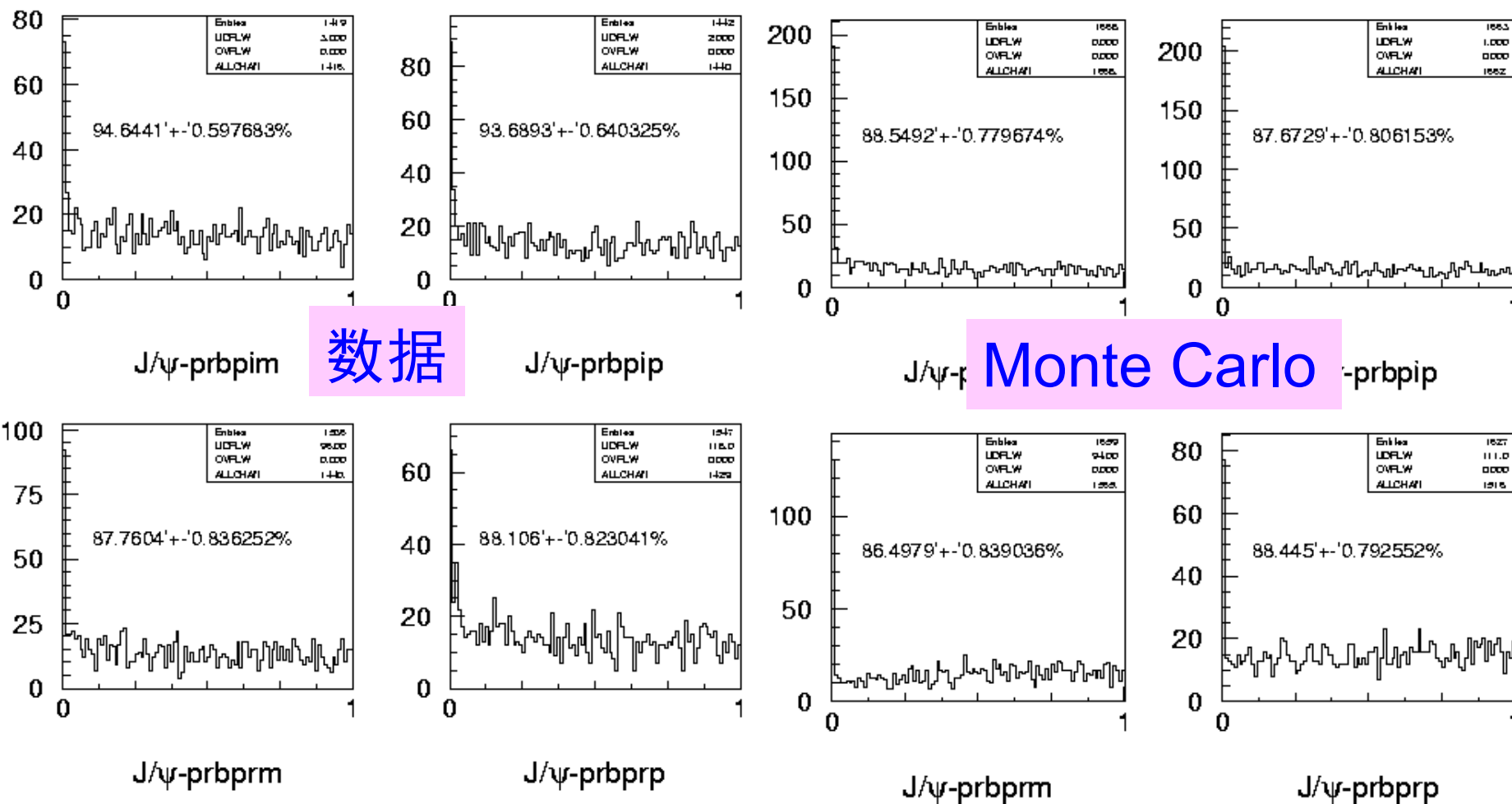
$$\mathcal{E}_{DT} = \mathcal{E}_{DT}^{Geom} \cdot \mathcal{E}_{DT}^{Track} \cdot \mathcal{E}_{DT}^{Ptid} \cdot \mathcal{E}_{DT}^{\gamma ID} \dots$$

$$\mathcal{E}_{MC} = \mathcal{E}_{MC}^{Geom} \cdot \mathcal{E}_{MC}^{Track} \cdot \mathcal{E}_{MC}^{Ptid} \cdot \mathcal{E}_{MC}^{\gamma ID} \dots$$

$$\mathcal{E}_{DT}^i = \mathcal{E}_{MC}^i \cdot f^i$$

测量 “ f^i ”，即可修正Monte Carlo效率得到真实数据对应的效率。这是物理分析的重要内容之一。

粒子鉴别效率的测量



数据和Monte Carlo对 π^+ 、 π^- 和质子、反质子粒子鉴别的效率。由 $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ 测得。

粒子鉴别效率的测量

Table 2: Particle ID efficiencies ($Prob_{ID} > 0.01$) for π^+ , π^- , p and $\bar{p}$ samples selected from $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ (in %).

Particle	data	Monte Carlo
π^-	94.64 ± 0.60	88.55 ± 0.78
π^+	93.69 ± 0.64	87.67 ± 0.81
$\bar{p}$	87.76 ± 0.84	86.50 ± 0.84
p	88.11 ± 0.82	88.44 ± 0.79

数值修约 (rounding)

MC efficiency corr. factor: 1.154 ± 0.028

关于效率估计的说明

- 数据与Monte Carlo的不一致永远存在，因此利用Monte Carlo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，数据与模拟的差别对结果影响很小，但对本底估计时需特别关注。
- 对于大统计量情形，期望高精度结果，数据与模拟的差别不容忽视。
- 对于稀有事例的寻找，本底情况的模拟可能会与真实情况有很大差别。
 - $J/\psi \rightarrow e\mu$ 的寻找，Bhabha和 $dimu$ 本底（ $e \rightarrow \mu$ 、 $\mu \rightarrow e$ 误判）

D衰变分支比测量

$$\psi'' \rightarrow D\bar{D}$$

Single Tag :

$$n_s^i = N_{D\bar{D}} \cdot B(D \rightarrow i) \cdot \varepsilon^i$$

Double Tag :

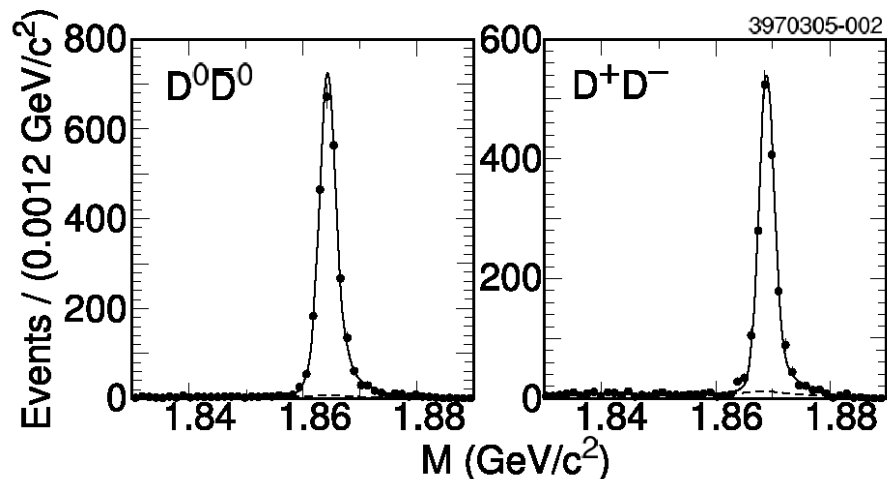
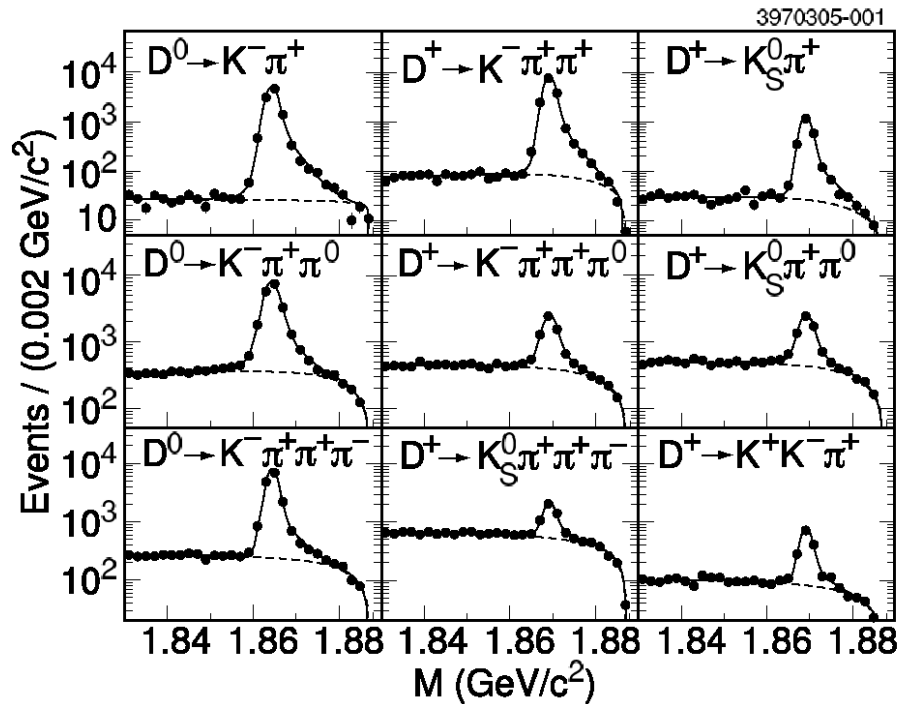
$$n_d^{ij} = N_{D\bar{D}} \cdot B(D \rightarrow i) \cdot B(\bar{D} \rightarrow j) \cdot \varepsilon^{ij}$$

Branching fraction :

$$B(\bar{D} \rightarrow j) = \frac{n_d^{ij} / \varepsilon^{ij}}{n_s^i / \varepsilon^i}$$

测比值时部分
系统误差抵消，
有利于提高精度。

D衰变分支比测量



Two important quantities used in D analysis:

1. Beam constrained mass:

$$M^2 = E_b^2 - p_D^2$$

2. $\Delta E = E_D - E_b$

(E_b is used to improve the resolutions)

ARGUS function for the background shape

[PLB241, 278 (1990)]

CLEO, PRL95, 121801 (2005)

总数测量

- 对于 e^+e^- 对撞产生的窄共振，可以利用R值测量的方法得到事例总数
- 也可以根据分支比计算公式，利用分支比已被高精度测量的道得到事例总数

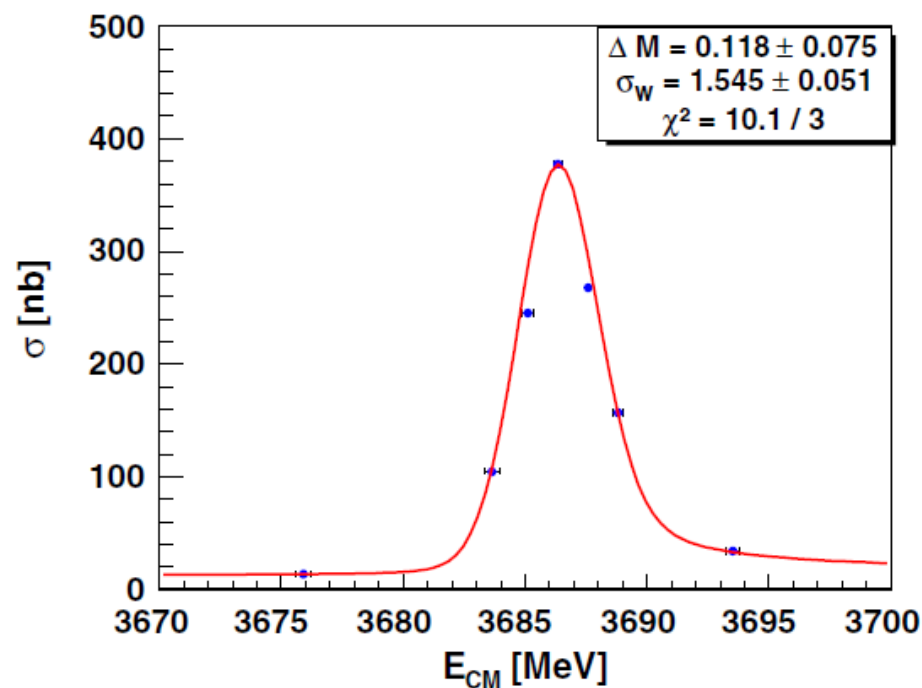
$$B(X \rightarrow f) = \frac{n^{obs} - n^{bkg}}{\epsilon N_X}$$

BESIII: Chin. Phys. C 37, 063001 (2013)

总数测量

- 对于 e^+e^- 对撞产生的窄共振，可以利用R值测量的方法得到事例总数
- Technique: $N_{\text{signal}} = N_{\text{peak}} - f \cdot N_{\text{continuum}}$

Interference?



BESIII: Chin. Phys. C 37, 063001 (2013)

总数测量

- 带电径迹多重数模拟不对
- Technique: solve overdetermined equations

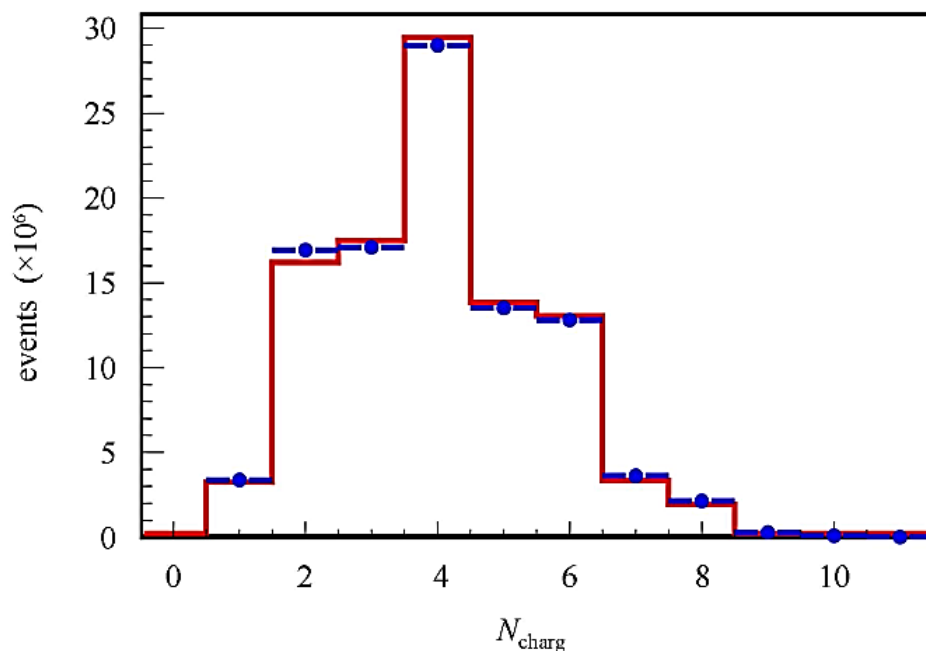


Fig. 9. The charged-track multiplicity distribution. Dots with error bars are data; the histogram is MC simulation.

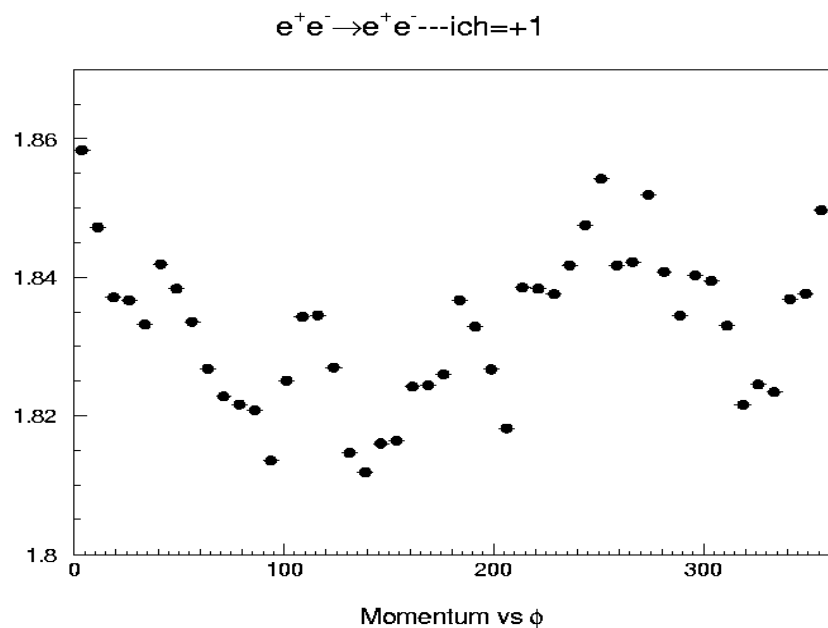
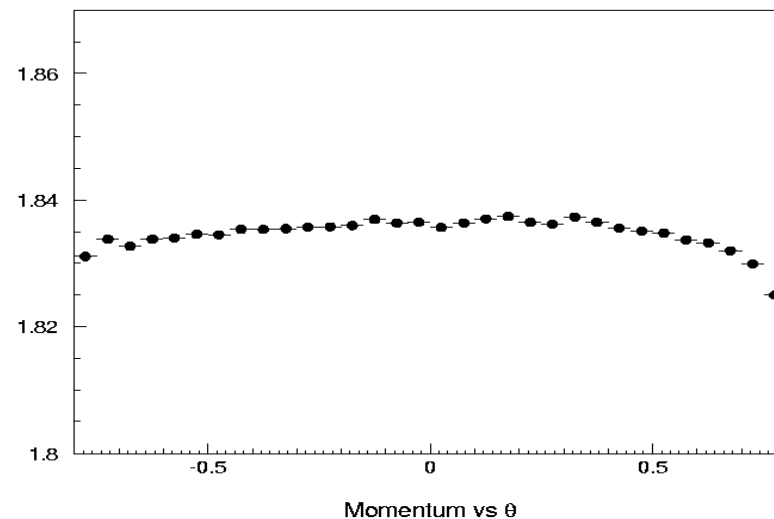
$$\chi^2 = \sum_{i=1}^{10} \frac{\left(N_i^{\text{obs.}} - \sum_{j=0}^{10} \epsilon_{ij} \cdot N_j \right)^2}{N_i^{\text{obs.}}},$$

质量测量：衰变产物

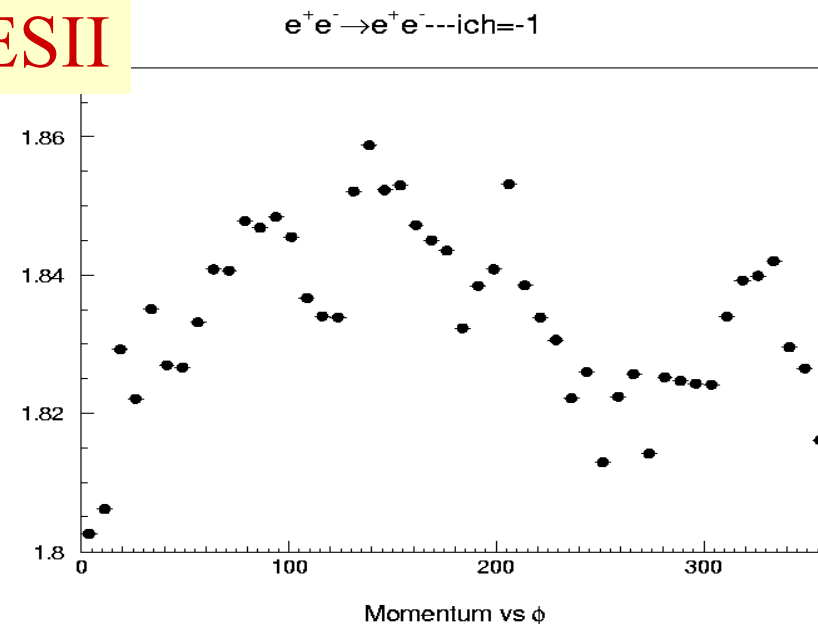
- $M^2 = (E_1 + E_2 + \dots)^2 - (\mathbf{p}_1 + \mathbf{p}_2 + \dots)^2$
 $E_i^2 = \mathbf{p}_i^2 + m_i^2$
- 影响质量测量的关键因素：
 - 动量测量（带电径迹）
 - 能量测量（中性径迹）
 - 本底形状（质量谱拟合）
- 例如
 - X_{c0} 、 η_c 、Higgs粒子质量测量
 - 一般轻强子或重子激发态质量测量

动量测量

- 需要检查和重新刻度动量。
- 这不是刻度的任务。



BESII



- 能量和方向。
- 与能量、极角、方位角的依赖。

能量测量

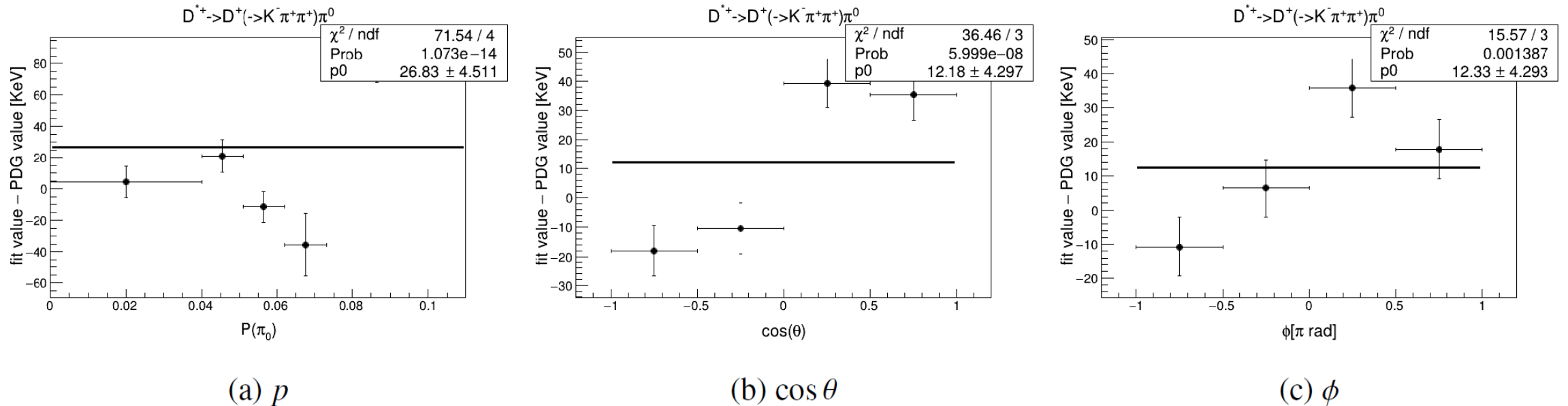
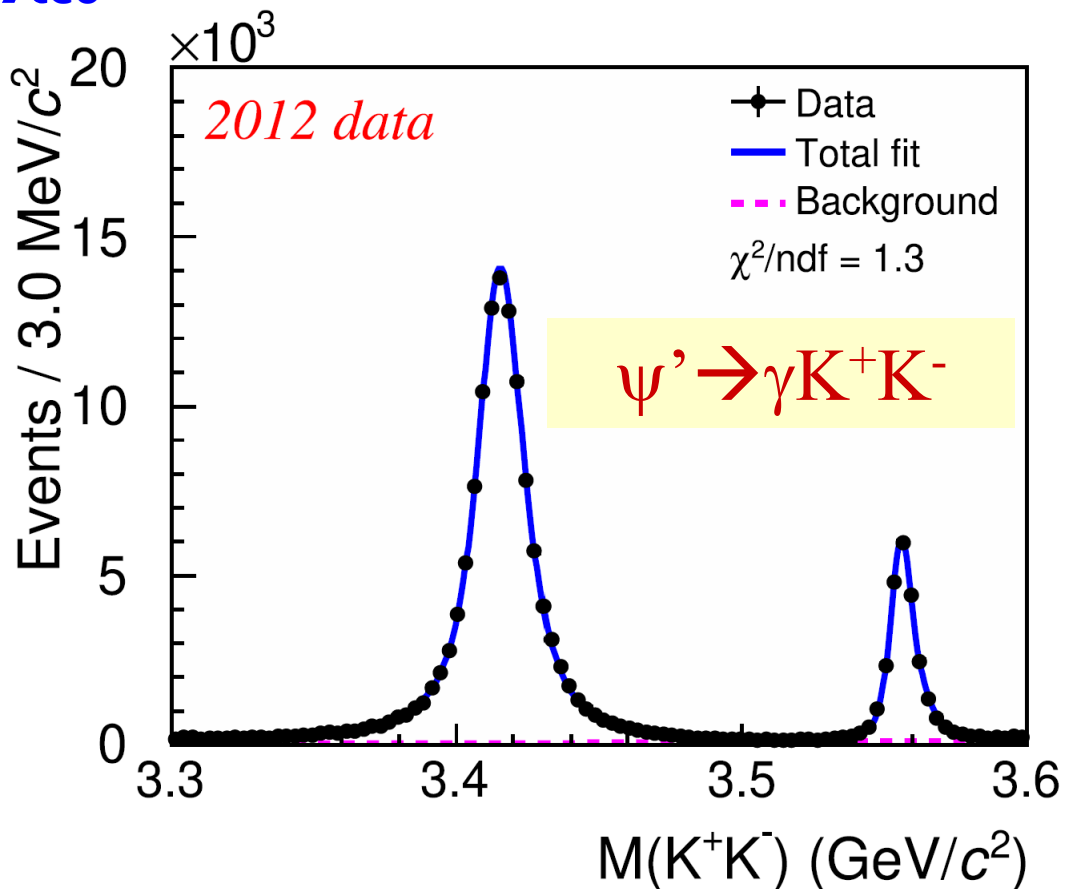


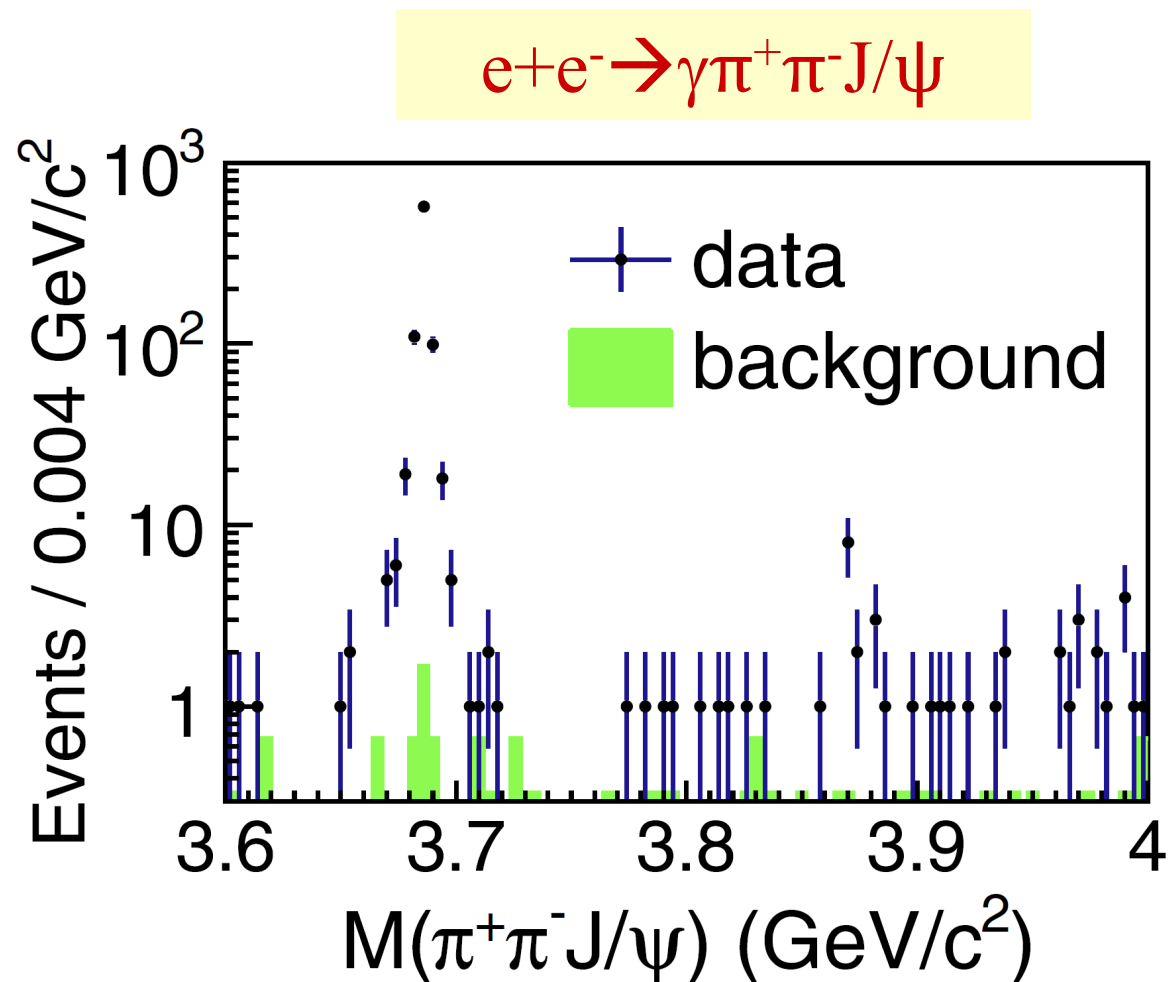
Fig. 19: Δm_s measurements from different ranges of the momentum magnitude p , polar angle $\cos \theta$ and the azimuth angle of π^0 ϕ of π^0 .

变通的办法

- 用精确已知的粒子刻度待测粒子。
- 例 χ_{c0} 、 $X(3872)$ 质量测量。



BESIII, CPC49, 091001 (2025)



BESIII, PRL 112, 092001 (2014) 53

质量测量：近阈产生

- 通过截面拟合的方法得到质量

- 影响结果的主要因素：

加速器能量标定

事例选择的绝对效率

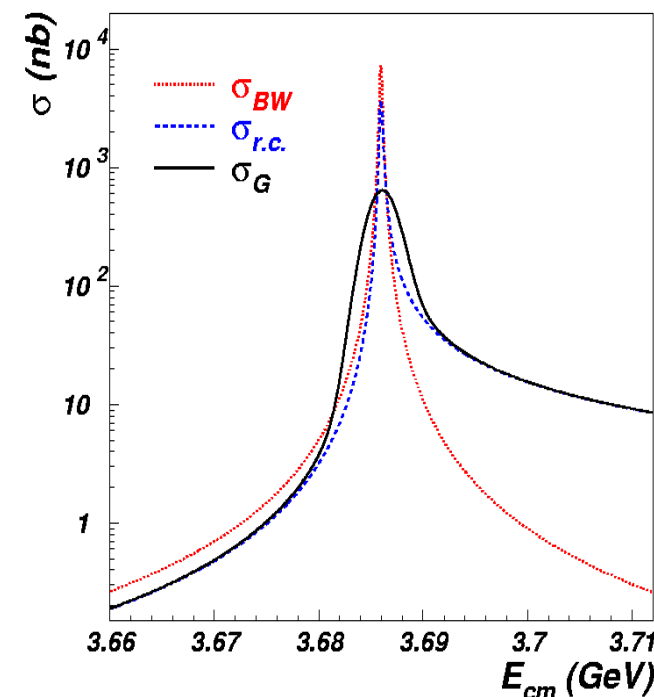
理论截面的精度

- 例

1. τ 质量测量 (PRD90,012001(2014))

2. J/ψ 、 $\psi(2S)$ 共振参数测量

(PLB355,374(1995), PLB550,24(2002))

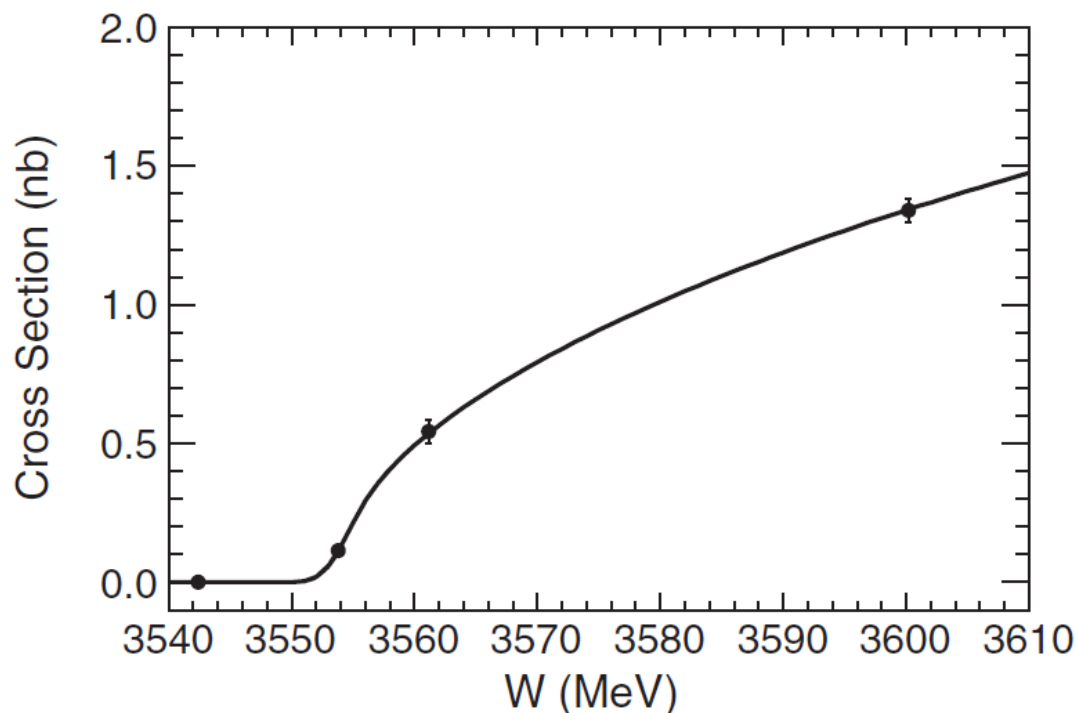


τ 质量测量

$$\sigma(W, m_\tau) = \frac{1}{\sqrt{2\pi}\Delta} \int_{2m_\tau}^{\infty} dW' e^{\frac{-(W-W')^2}{2\Delta^2}} \int_0^{1-\frac{4m_\tau^2}{W'^2}} dx F(x, W') \sigma_1(W' \sqrt{1-x}, m_\tau)$$

where σ_1 is given by

$$\sigma_1(W, m_\tau) = \frac{4\pi\alpha^2}{3W^2} \frac{\beta(3-\beta^2)}{2} \frac{F_c(\beta)F_r(\beta)}{[1-\Pi(W)]^2};$$



取多少个数据点？

在哪儿取数？

怎样分配取数时间？

怎样得到tau质量？

如何控制误差？

BES的发明，经典案例。BESIII: PRD90,012001(2014)

τ 质量测量

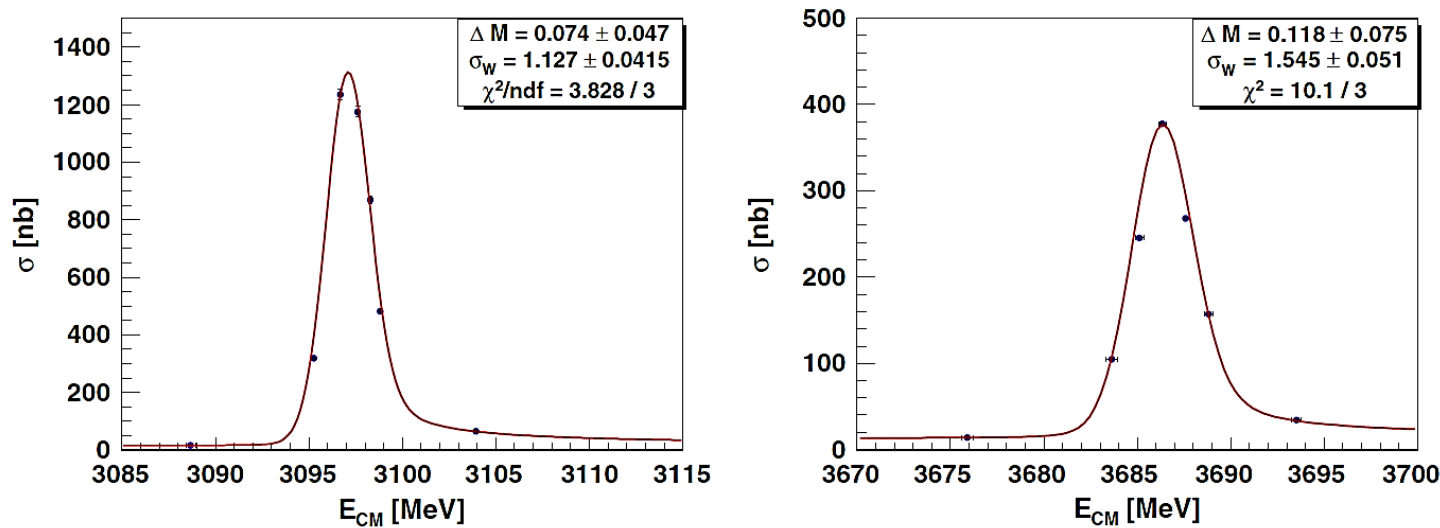


FIG. 4 (color online). Fits of the J/ψ (left) and ψ' (right) hadronic cross sections.

能量刻度

能散测量

$$L(m_\tau, \mathcal{R}_{\text{data/MC}}, \sigma_B) = \prod_{i=1}^4 \frac{\mu_i^{N_i} e^{-\mu_i}}{N_i!}, \quad (13)$$

极大似然法

where N_i is the number of observed τ pair events at scan point i ; μ_i is the expected number of events and calculated by

$$\mu_i = [\mathcal{R}_{\text{data/MC}} \times \epsilon_i \times \sigma(E_{\text{c.m.}}^i, m_\tau) + \sigma_B] \times \mathcal{L}_i. \quad (14)$$

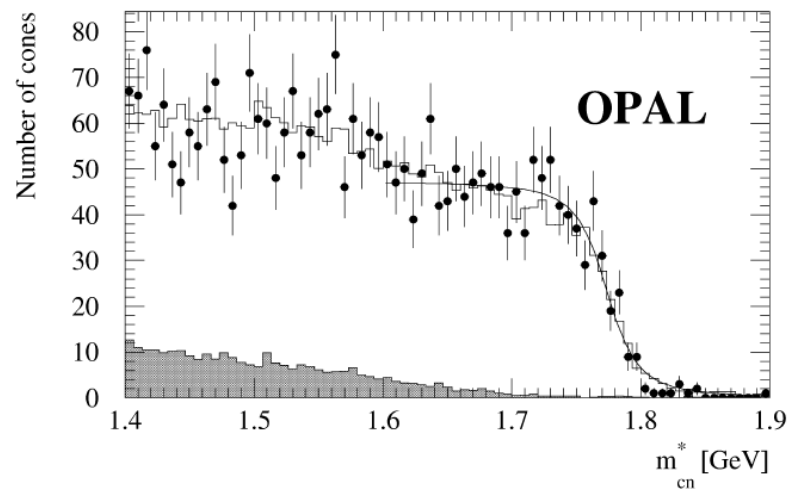
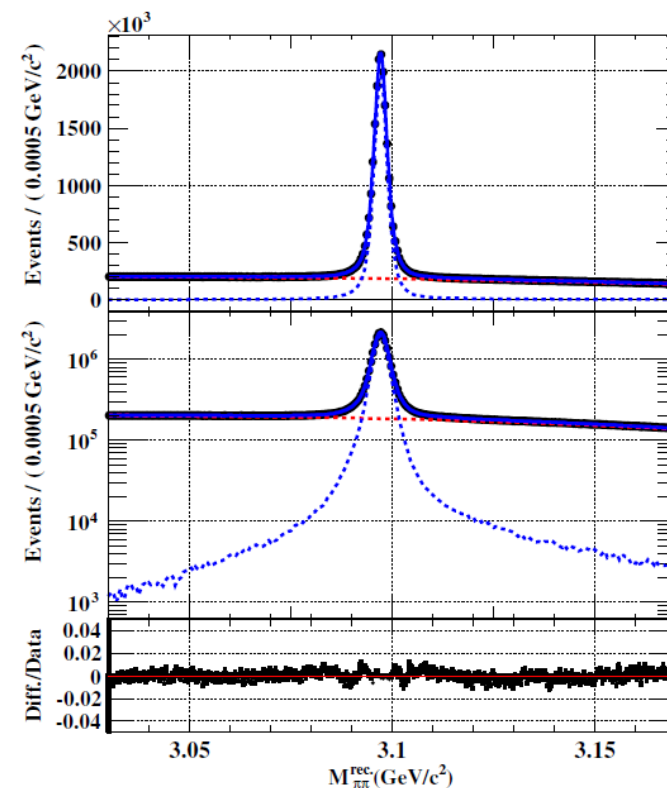
质量测量

- 反冲质量

- $e^+e^- \rightarrow \Sigma h_i + X$
- 初态四动量已知,
- h_i 四动量已知
- $m_X = ?$

- 质质量

- $e^+e^- \rightarrow \tau^+\tau^-$, $\tau^- \rightarrow \pi^+\pi^-\pi^-\nu$
- 初态四动量已知
- π^+ 、 π^- 、 π^- 四动量已知
- $m_\tau = ?$
- Most recent: Belle II



宽度测量

- 衰变产物

拟合不变质量谱

- 对撞产生

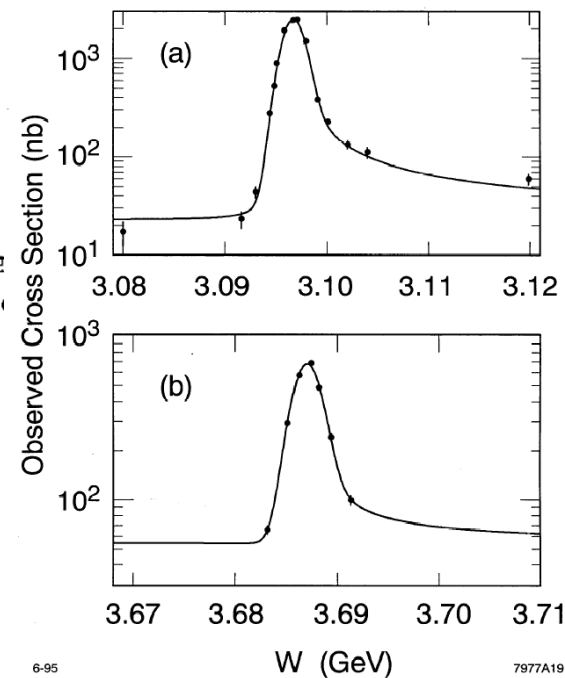
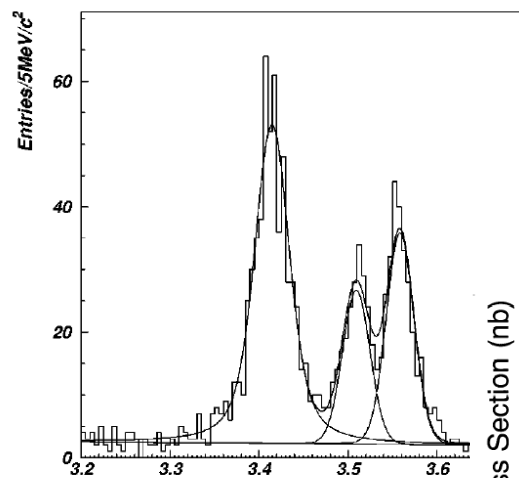
拟合产生截面

- 影响宽度测量的主要因素

1. 质量分辨或能散

2. 本底形状

- 往往质量和宽度同时测量，但前者对质量分辨依赖很小。

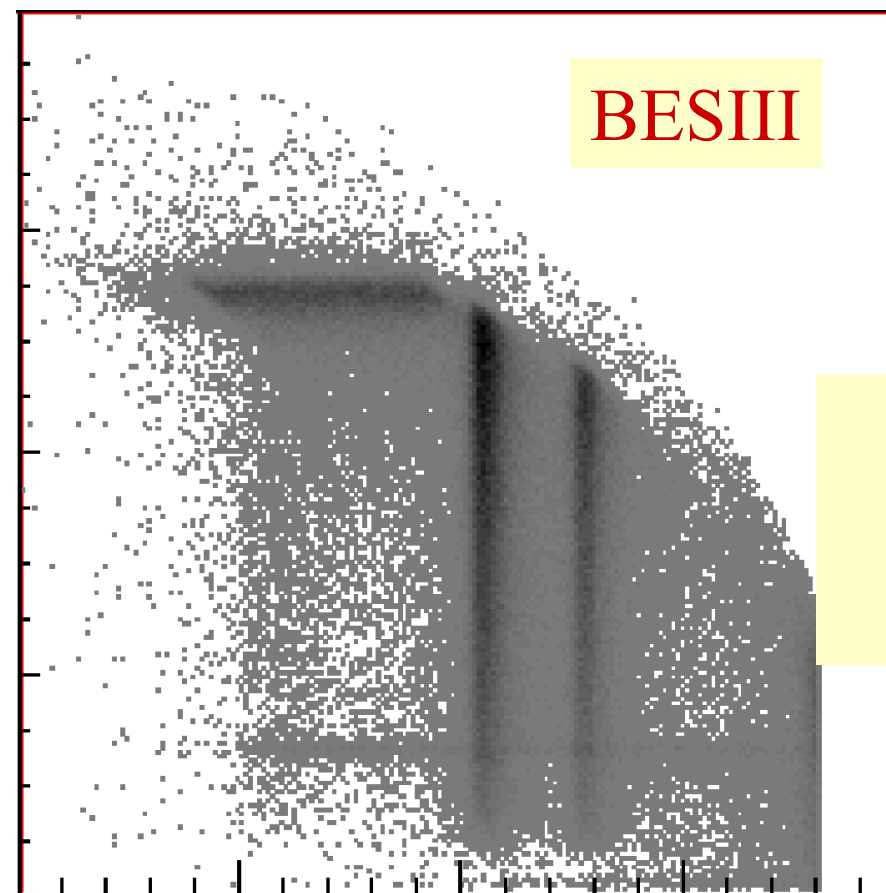
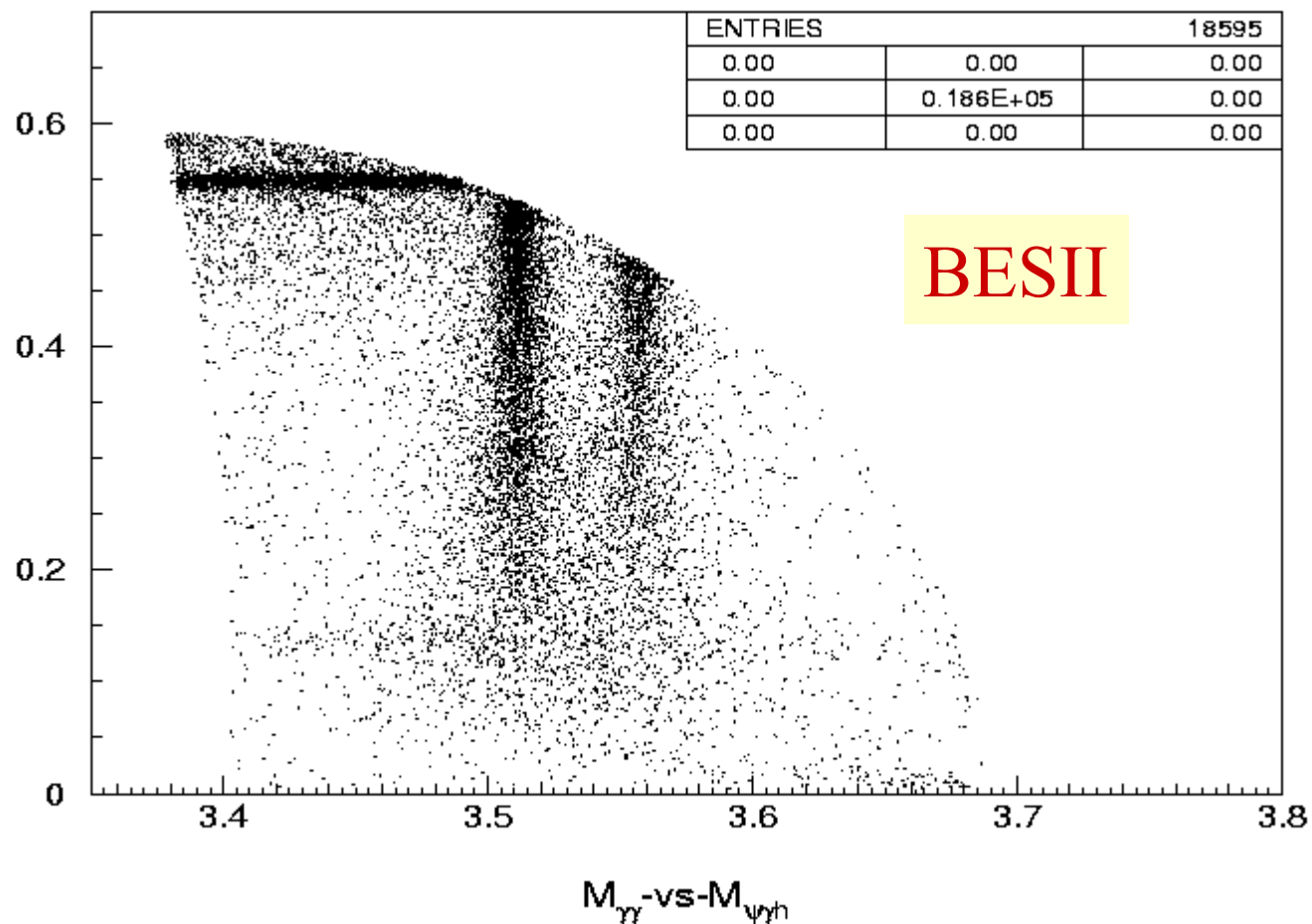


6-95

7977A19

非高斯形式的质量分辨

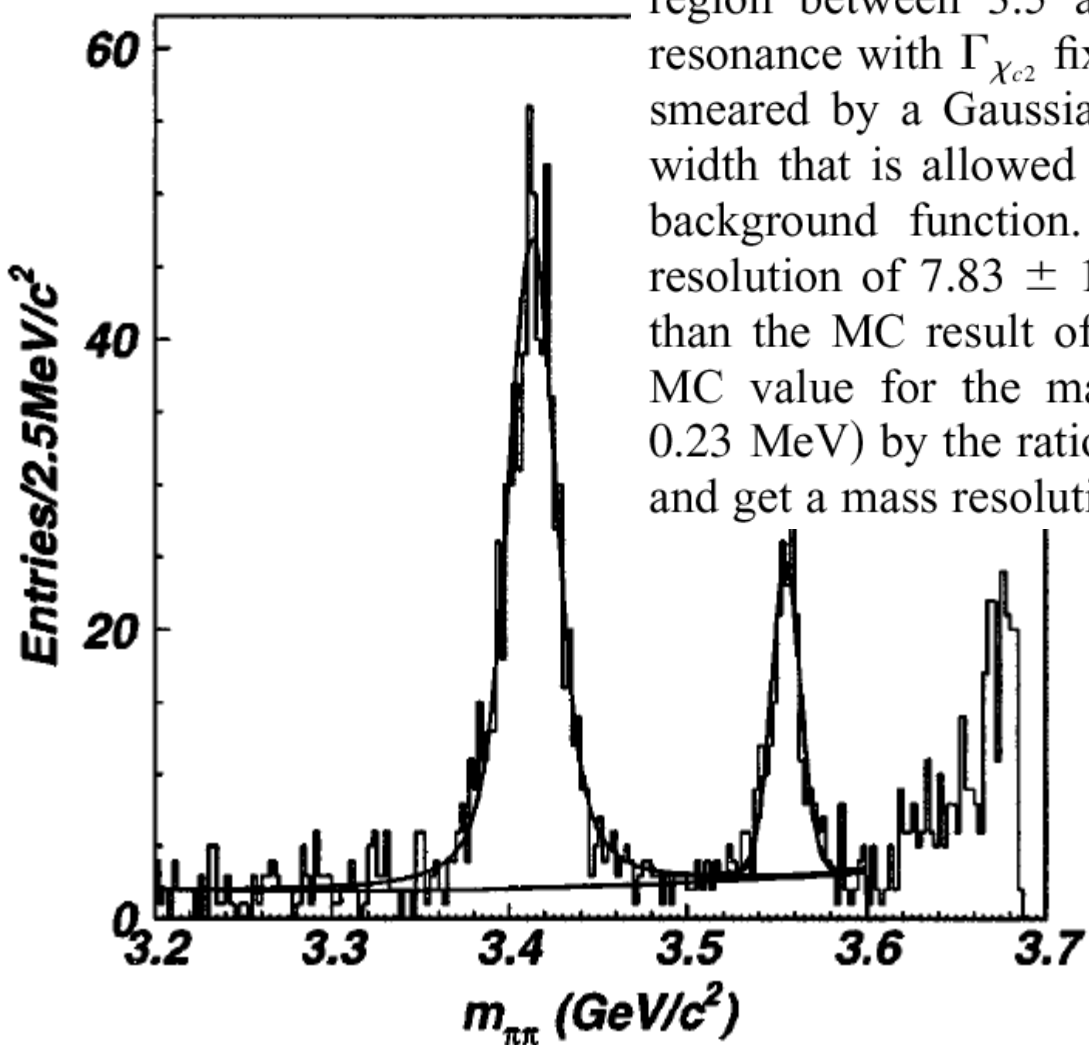
$$\psi(2S) \rightarrow \gamma\gamma J / \psi \rightarrow \gamma\gamma l^+ l^-$$



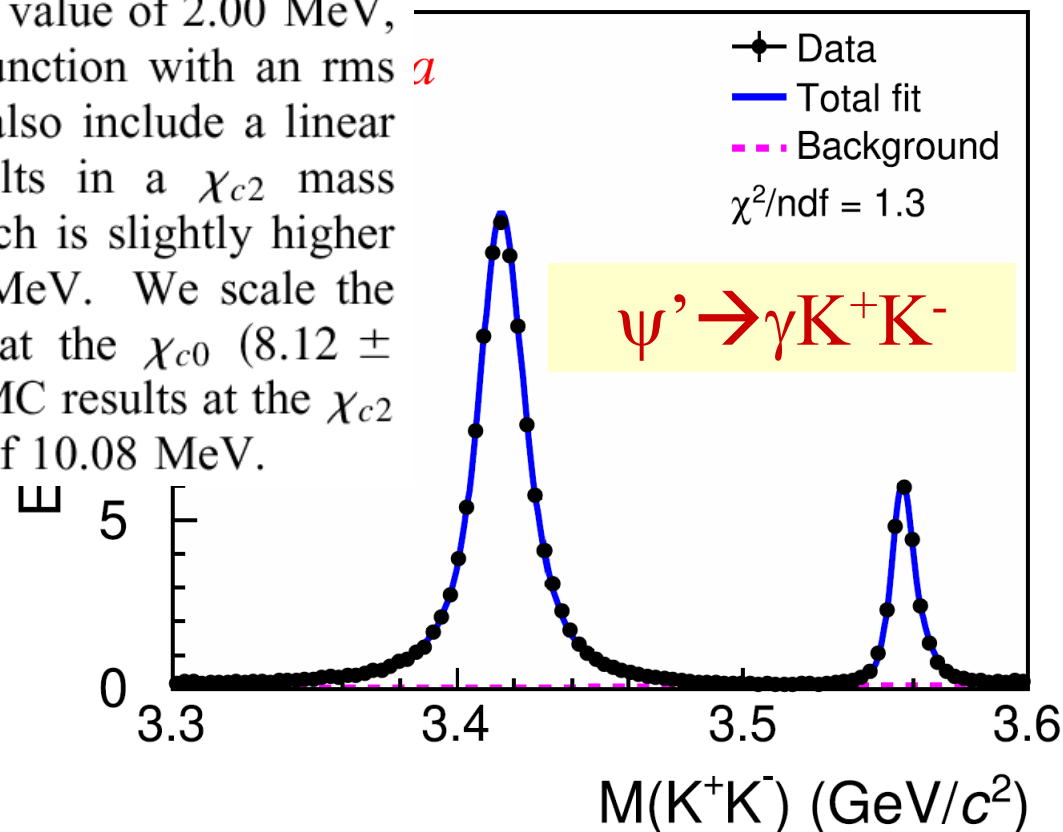
χ_{c0} 宽度测量!

由数据刻度质量分辨

For the $\pi^+\pi^-$ channel, we first fit the invariant mass region between 3.5 and 3.6 GeV with a Breit-Wigner resonance with $\Gamma_{\chi_{c2}}$ fixed at the PDG value of 2.00 MeV, smeared by a Gaussian resolution function with an rms width that is allowed to float. We also include a linear background function. The fit results in a χ_{c2} mass resolution of 7.83 ± 1.04 MeV, which is slightly higher than the MC result of 6.31 ± 0.11 MeV. We scale the MC value for the mass resolution at the χ_{c0} (8.12 ± 0.23 MeV) by the ratio of the fitted MC results at the χ_{c2} and get a mass resolution at the χ_{c0} of 10.08 MeV.



BES, PRL81, 3091(1998)

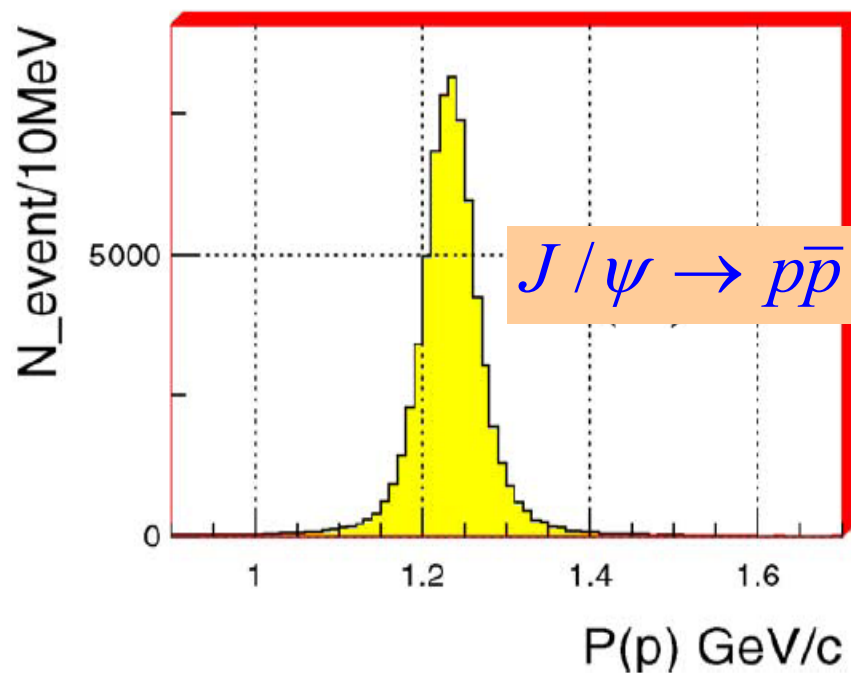


BESIII, CPC49, 091001 (2025)

角分布测量

- 分波分析中除了用到不变质量信息，还用到末态粒子的角分布信息，因此可以更好地确定共振态的参数（尤其自旋-宇称量子数）。
- 角分布的测量实际上要用到与角度有关的效率。（一般的效率实际是对角度积分的效果，如果模拟有问题，同样会对结果造成影响。）
- 可以通过对已知角分布过程的测量标定与角分布有关的效率。

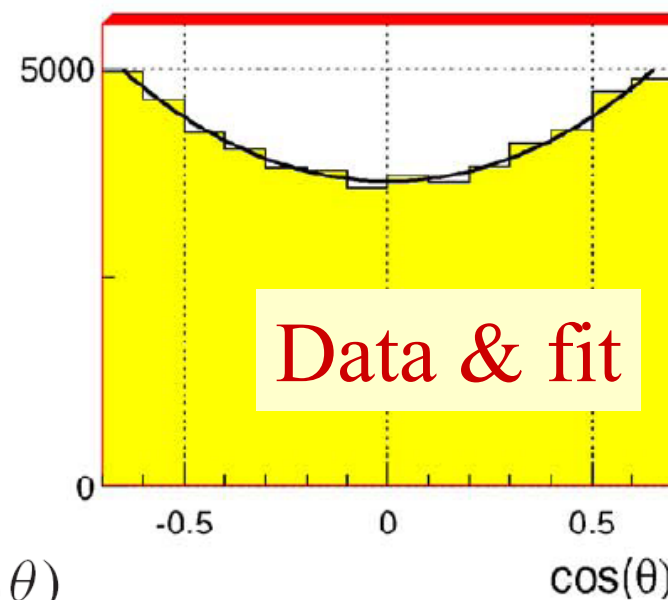
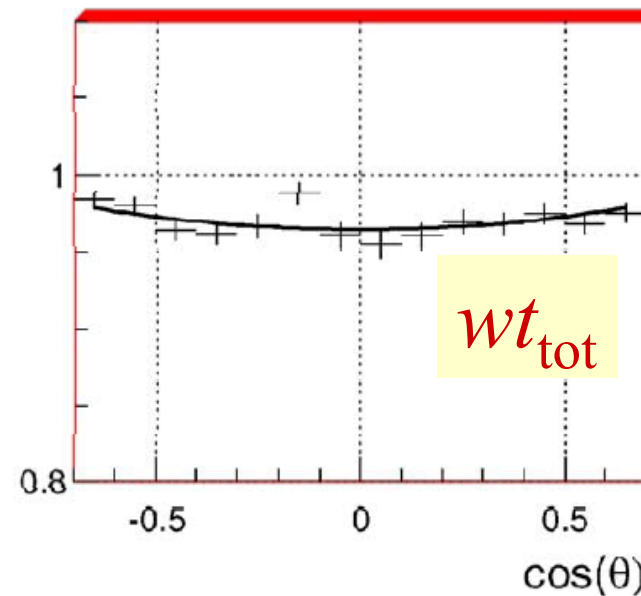
角分布测量



$$\frac{dN}{d\cos\theta_B} \propto 1 + \alpha \cos^2 \theta_B$$

$$wt_{\text{tot}}(\cos\theta) = wt_p(\cos\theta)wt_{\bar{p}}(\cos\theta)wt_{pid}(\cos\theta) \\ \times wt_{\bar{p}id}(\cos\theta),$$

$$f(\cos\theta) = A(1 + \alpha \cos^2 \theta) \varepsilon_{\text{mc}}(\cos\theta) wt_{\text{tot}}(\cos\theta)$$

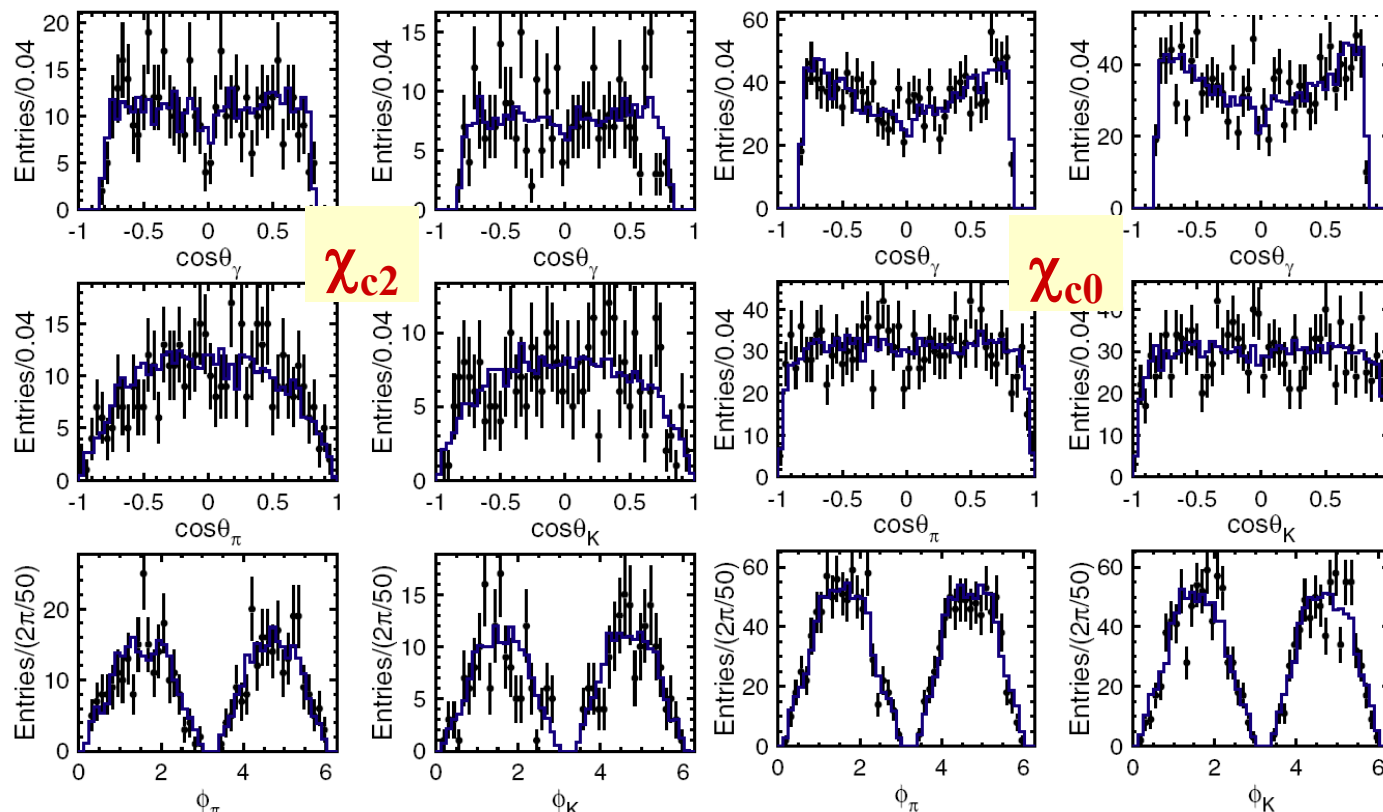
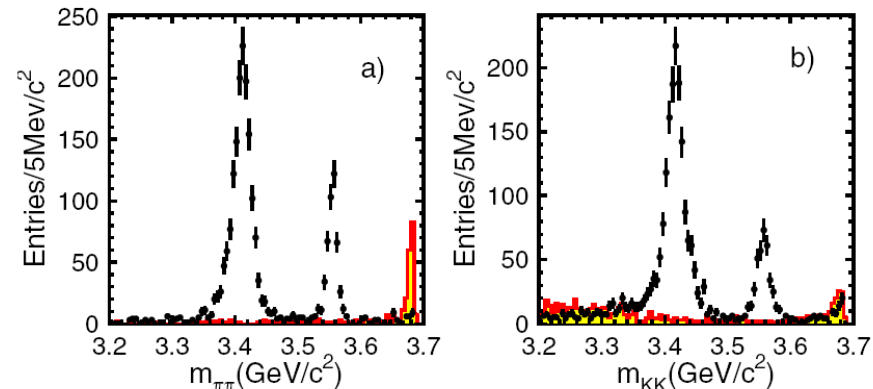


角分布测量

Measurement of the χ_{c2} polarization in $\psi(2S) \rightarrow \gamma\chi_{c2}$

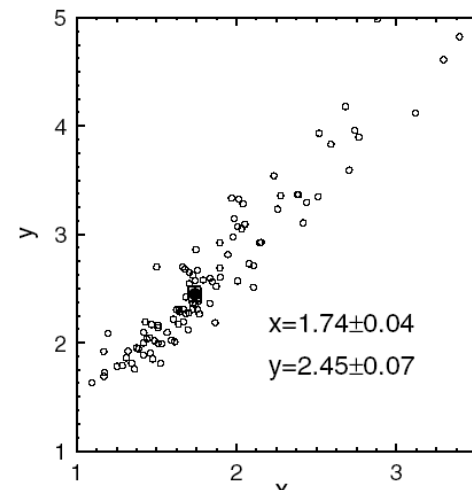
最简单的PWA例子，两个自由参数由极大似然法给出。

(BES, PRD70, 092004 (2004); BESIII, PRD84, 092006 (2011))



$$f(x_i; \alpha) = dI/d(LIPS)$$

$$\mathcal{L} = \prod_{i=1}^{N_{events}} f(x_i; \alpha)$$



$$W_2(\theta_\gamma, \theta_M, \phi_M) = 3x^2 \sin^2 \theta_\gamma \sin^2 2\theta_M + (1 + \cos^2 \theta_\gamma) \left[(3 \cos^2 \theta_M - 1)^2 + \frac{3}{2} y^2 \sin^4 \theta_M \right] + \sqrt{3} x \sin 2\theta_\gamma \\ \times \sin 2\theta_M \left[3 \cos^2 \theta_M - 1 - \frac{1}{2} \sqrt{6} y \sin^2 \theta_M \right] \cos \phi_M + \sqrt{6} y \sin^2 \theta_\gamma \sin^2 \theta_M (3 \cos^2 \theta_M - 1) \cos 2\phi_M$$

误差分析

- 误差分析的重要性
- 统计误差的确定
- 系统误差的确定

珠穆朗玛峰高程测量

<http://www.cctv.com/news/china/20051009/100402.shtml>

[国新办发布会]介绍2005珠穆朗玛峰高程数据



[国家测绘局局长陈XX]:

根据《中华人民共和国测绘法》珠峰高程新数据经国务院批准并授权，由国家测绘局公布。现在我正式宣布2005珠峰高程测量获得的新数据为：珠穆朗玛峰峰顶岩石面海拔高程8844.43米。

2005:

$$H_{\text{rock}} = (8844.43 \pm 0.21) \text{m}$$

$$D_{\text{snow}} = 3.50 \text{m}$$

- 参数：珠穆朗玛峰峰顶岩石面高程测量精度 ± 0.21 米；峰顶冰雪深度3.50米。原1975年公布的珠峰高程数据停止使用。

珠穆朗玛峰高程测量

<http://www.cctv.com/news/china/20051009/100402.shtml>

[国新办发布会]介绍2005珠穆朗玛峰高程数据



- [陈XX院士,1975年珠峰高程测定的设计者和组织者]
- 1975年公布的岩面高是8848.13米，高程的精度，是正负0.35米。在1975年的时候，雪深的测量是0.92米。

1975:

$$H_{\text{rock}} = (8848.13 \pm 0.35)\text{m}$$

$$D_{\text{snow}} = 0.92\text{m}$$

珠穆朗玛峰高程测量

<http://www.cctv.com/news/china/20051009/100402.shtml>

[国新办发布会]介绍2005珠穆朗玛峰高程数据



- [陈XX院士，解读差别]
- 因为在1975年的时候，雪深的测量是0.92米，这次是3.50米，差了差不多2.7 (2.6) 米。重力差的计算，导致大地水准基准面升高了0.7米，这样还剩下0.3-0.4 (0.4) 米，也可能是误差的问题，也可能是其他的变化。

2005-1975:

$$\Delta H_{\text{rock}} = (-3.70 \pm 0.41) \text{m}$$

9 σ difference!

$$\Delta D_{\text{snow}} = 2.58 \text{m}$$

结论：1975年测量误差估计偏小，测量结果存在两个严重偏差 (bias) 。

推论：误差估计非常重要！

误差的分类

- **统计误差**：即随机误差，由于样本大小造成的不确定性。靠增大统计量减小。
- **系统误差**：非随机误差，绝对值及符号恒定，或条件改变而按一定规律变化。靠增大信号本身统计量无法减小。判断误差来源，改善方法或进行测量并修正。
- **粗大误差**：异常测量值。

此处仅限于统计误差与系统误差。

Chebyshev大数定律

$X_1, X_2, \dots, X_n$ 为相互独立的随机变量且

$$E(X_i) = \mu_i, V(X_i) = \sigma_i^2,$$

则对于任意 $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n \mu_i \right| < \varepsilon \right\} = 1.$$

特例:

如果 $X_i \stackrel{i.i.d}{\sim} (\mu, \sigma^2)$, 则

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon \right\} = 1.$$

**利用统计的方法
进行测量的基础。**

Chebyshev不等式

对任意随机变量 X （离散或连续），且

$$E(X) = \mu, V(X) = \sigma^2,$$

则对于任意 $\varepsilon > 0$,

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

或写为（ k 为正数）:

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2},$$

即出现在 k 倍标准偏差以外的几率不大于 $\frac{1}{k^2}$.

k	任意分布 ($P \leq$)	高斯分布 ($P =$)
1	1	0.3173
2	0.25	0.0455
3	0.11	0.0027
5	0.04	5.7×10^{-7}
10	0.01	---

Bessel 公式

如果 $X_i \stackrel{i.i.d}{\sim} (\mu, \sigma^2)$, $i = 1, 2, \dots, n$, 其中 μ, σ^2 未知,

$$\text{取 } v_i = x_i - \bar{x}, \quad \bar{x} = \sum_i x_i / n,$$

则 $\hat{\mu} = \bar{x};$

$$\hat{\sigma}^2 = \sum_i v_i^2 / (n-1);$$

区分单次测量的方差
与平均值的方差!

$$\hat{\sigma}^2(\bar{x}) = \hat{\sigma}^2 / n.$$

x 的标准偏差为 $s = \sqrt{\sum_i v_i^2 / (n-1)}$

s 的标准偏差的估计值 $\hat{\sigma}_s = \frac{s}{\sqrt{2(n-1)}}$ (误差的误差)

统计误差确定

- **物理量**

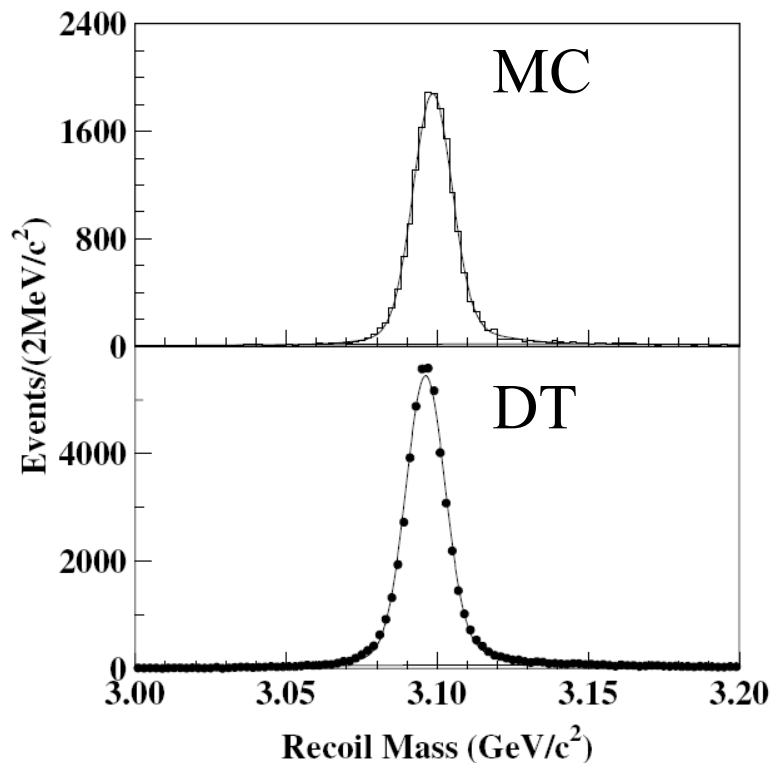
- 截面或分支比
- 质量（粒子或共振态） / 宽度或寿命
- $I^G J^{PC}$ 或角分布
- 极化或不对称性
- ...

- **基本分为两类**

- 通过粒子四动量计算（误差由样本大小及分辨，四动量不确定性决定）
- 计数（物理量满足的分布，样本大小，效率不确定性）

统计误差确定

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



$$X_i \stackrel{i.i.d}{\sim} (m_{J/\psi}, \sigma^2), i = 1, 2, \dots, n,$$

其中 $m_{J/\psi}, \sigma^2$ 未知,

$$\text{取 } v_i = x_i - \bar{x}, \quad \bar{x} = \sum_i x_i / n,$$

则

$$m_{J/\psi} = \bar{x};$$

$$\hat{\sigma}^2 = \sum_i v_i^2 / (n - 1);$$

$$\hat{\sigma}^2(\bar{x}) = \hat{\sigma}^2 / n.$$

$$\sigma m_{J/\psi} = \sqrt{\hat{\sigma}^2(\bar{x})}$$

即Bessel公式。

实际上可由程序拟合直接得到 (root, minuit等)

$$\phi(u) = E \left[e^{iux} \right] = \int_{-\infty}^{\infty} e^{iux} f(x) dx$$

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Uniform	$f(x; a, b) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{ibu} - e^{iau}}{(b-a)iu}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Binomial	$f(r; N, p) = \frac{N!}{r!(N-r)!} p^r q^{N-r}$ $r = 0, 1, 2, \dots, N ; \quad 0 \leq p \leq 1 ; \quad q = 1 - p$	$(q + pe^{iu})^N$	Np	Npq
Poisson	$f(n; \nu) = \frac{\nu^n e^{-\nu}}{n!} ; \quad n = 0, 1, 2, \dots ; \quad \nu > 0$	$\exp[\nu(e^{iu} - 1)]$	ν	ν
Normal (Gaussian)	$f(x; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp(-(x - \mu)^2 / 2\sigma^2)$ $-\infty < x < \infty ; \quad -\infty < \mu < \infty ; \quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\mathbf{x}; \boldsymbol{\mu}, V) = \frac{1}{(2\pi)^{n/2} \sqrt{ V }}$ $\times \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T V^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right]$ $-\infty < x_j < \infty ; \quad -\infty < \mu_j < \infty ; \quad \det V > 0$	$\exp \left[i\boldsymbol{\mu} \cdot \mathbf{u} - \frac{1}{2}\mathbf{u}^T V \mathbf{u} \right]$	$\boldsymbol{\mu}$	V_{jk}
χ^2	$f(z; n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2} \Gamma(n/2)} ; \quad z \geq 0$	$(1 - 2iu)^{-n/2}$	n	$2n$

"标准正态"分布?

二项分布 (binomial)

- n次试验x次成功的几率

$$p(x) = C_n^x p^x (1-p)^{n-x}$$

$$\bar{x} = n \cdot p$$

$$V(x) = n \cdot p \cdot (1-p)$$

- 应用

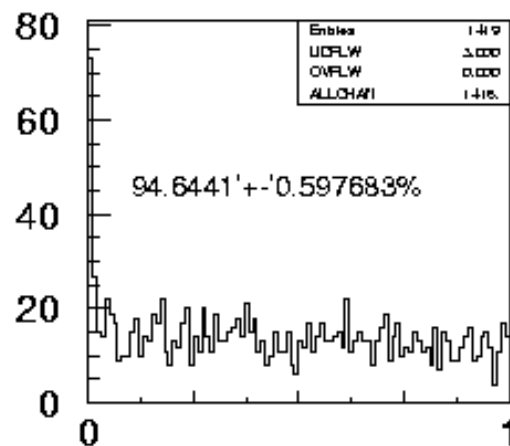
- 总体固定时样本的涨落
- 用MC样本确定效率及统计误差
- 粒子鉴别效率测量等

$$N^{prod} \xrightarrow{\text{selection}} n^{obs}$$

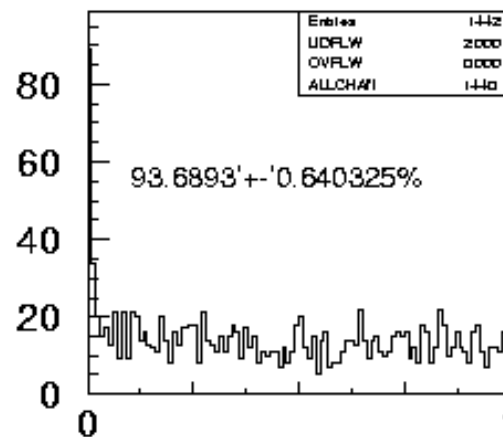
$$\varepsilon = \frac{n^{obs}}{N^{prod}}, \quad \sigma\varepsilon = \frac{\sigma n^{obs}}{N^{prod}}$$

$$\sigma n^{obs} = \sqrt{N^{prod} \cdot \varepsilon \cdot (1-\varepsilon)}$$

粒子鉴别效率的测量

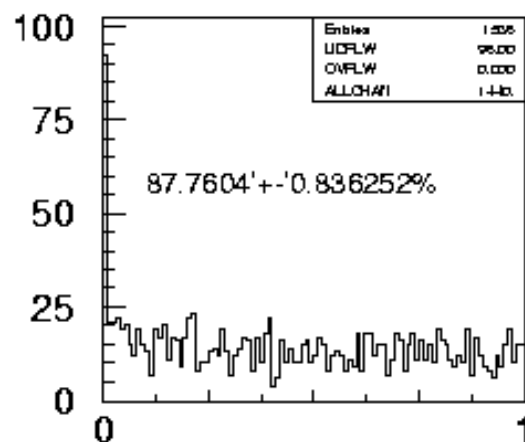


J/ψ-prbpim

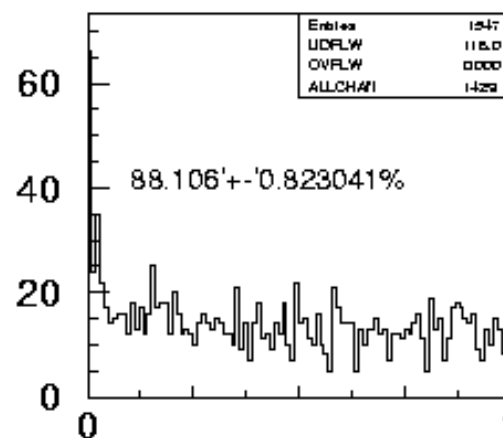


J/ψ-prbpip

误差由二项
分布确定!



J/ψ-prbprm



J/ψ-prbprp

注意有效位数!

ϵ 与 $(1-\epsilon)$ 误差一样

泊松分布 (Poisson)

- 二项分布中 $n \rightarrow \infty, p \rightarrow 0$ 时 ($\lambda = np$)

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$$

$$\bar{x} = \lambda$$

$$V(x) = \lambda$$

- 应用
 - 直方图(bin足够多, n足够大, 分布不很集中)
 - 小样本实验
 - BES τ 质量测量: PRL69,3021 (1992)
 - BES f_{DS} 测量: PRL74,4599 (1995)
 - 新物理寻找

正态分布 (Normal/Gaussian)

- Poisson分布中x较大时 ($x > 20, 10, 7?$)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, N(\mu, \sigma^2)$$

$$\bar{x} = \mu$$

$$V(x) = \sigma^2$$

$N(0,1)$: 标准正态分布

- 应用(广泛)
 - 误差分析中的 $1\sigma=68.3\%$ 几率
 - 信号显著性 ($n\sigma$, n 倍标准偏差)
 - χ^2 定义的基础
 - 中心极限定理

系统误差

- 非统计原因造成的测量结果与真值之间的差别。
- 系统误差的来源：
 1. 测量设备（测量不准）
 2. 实验、分析方法（顶点拟合、运动学拟合）
 3. 理论模型的应用（lund用在低能）
 4. Monte Carlo的模拟（核作用的模拟）
 5. 实验者的偏见（期望有新发现）
 6. 理论预期的误导（标准模型总是对的）
 7. 其它（与别的实验不一致心里不塌实）

系统误差确定

- **基本考虑**

- **待测物理量与哪些测量量有关，如何将测量量的误差传递给待测物理量**

$$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos \theta_{\gamma\gamma})$$

- **原则上可以由每个测量量误差求出。**
- **但有时几乎不可能（关系过于复杂），如**

对于 $J/\psi \rightarrow \pi^+\pi^-\gamma\gamma$ 进行了运动学拟合以后

$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos \theta_{\gamma\gamma})$ 的误差与带电径迹也有关。

- **寻找变通的办法！对数据分析很重要！**

系统误差的分析

10. Determination of systematic uncertainties

10.1. Methodology

ALEPH, Phys. Rep. 421, 191 (2005)

Wherever possible the efficiencies relevant to the analysis have been determined using ALEPH data, either directly on the $\tau\tau$ sample itself or on specifically selected control samples, as for example in the case of particle identification. The resulting efficiencies are thus measured with known statistical errors.

In some cases the procedure is less straightforward and involves a model for the systematic effect to be evaluated. An important example is given by the systematics in the simulation of fake photons in ECAL. In such cases the evaluation of the systematic error not only takes into account the statistical aspect, but also some estimate of the systematics involved in the assumed model. The latter is obtained from studies where the relevant parameters are varied in a range consistent with the comparison between data and Monte Carlo distributions.

Quite often a specific cut on a given variable can be directly studied. The comparison between the corresponding distributions, respectively in data and Monte Carlo, provides an estimate of a possible discrepancy whose effect would be to change the assumed efficiency of the cut. If a significant deviation is observed, a correction is applied to the simulation to obtain the nominal branching ratio results, while the error on the deviation provides the input to the evaluation of the systematic uncertainty. The analysis is therefore repeated with a full re-classification of all the measured τ decay candidates, changing the incriminated cut by one standard deviation. Since the new samples slightly differ from the nominal ones because of feedthrough between the different channels, the modified results are affected both by the systematic change in the variable value and the statistical fluctuation from the event sample which is uncommon to both samples. In this case the final systematic uncertainty is obtained by adding linearly the modulus of the systematic deviation observed and the statistical error from the monitored independent sample. This procedure is followed in a systematic way for the studies listed below and its description will of course not be repeated each time.

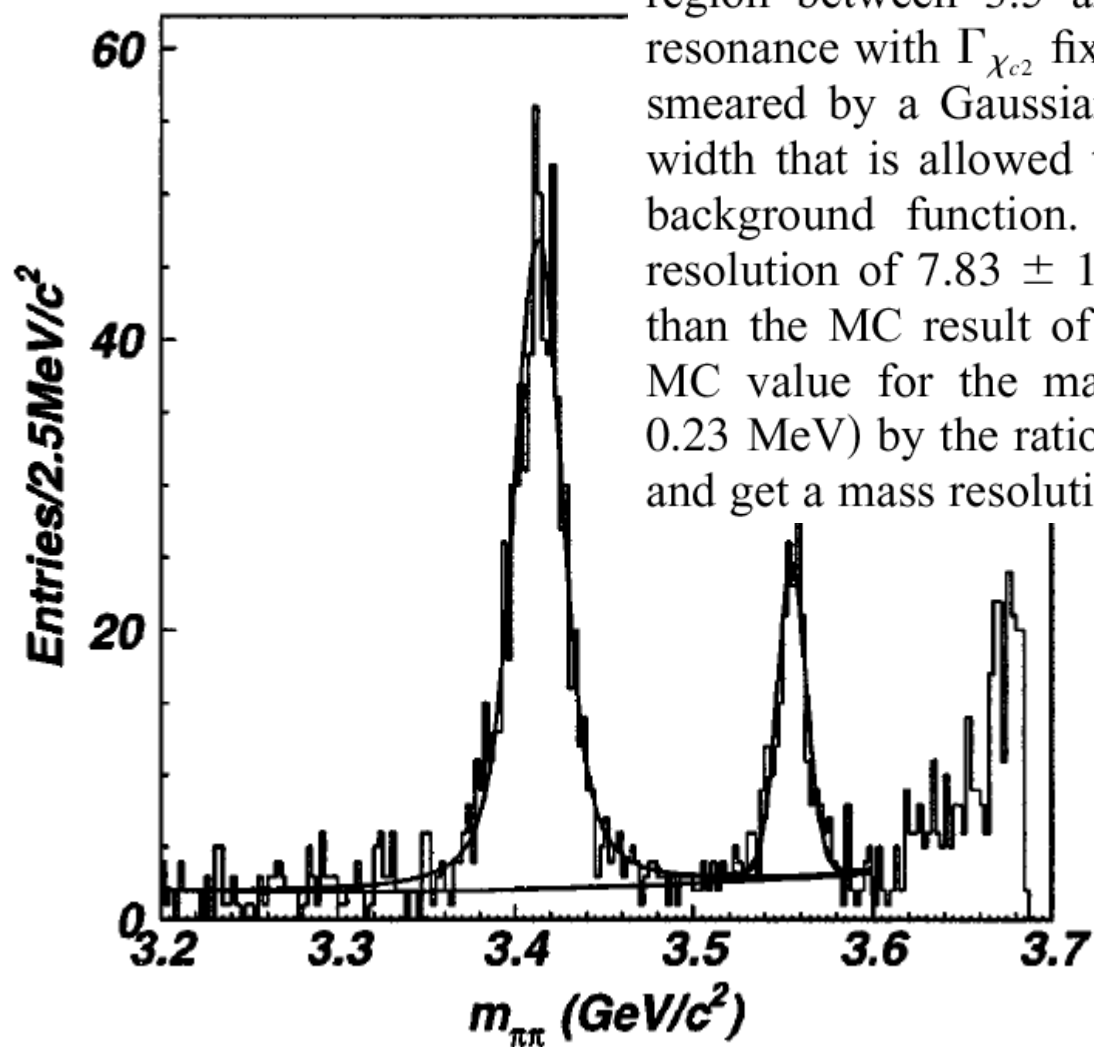
Finally, the systematic deviations for each study are obtained with their sign in each measured decay channel, thus providing the full information on the correlations between the results and allowing the corresponding covariance matrix to be constructed.

系统误差的分析

- 模型误差、理论误差等较容易考虑（至少不必由实验者负责）。
- 事例选择误差确定是实验者的基本任务。
- 一言以蔽之，事例选择系统误差表征
“测量数据与Monte Carlo之间的吻合程度”
- 在“效率修正”部分已经包括：
系统误差即“f因子”的误差。
- 质量分辨等造成的系统误差同样考虑。

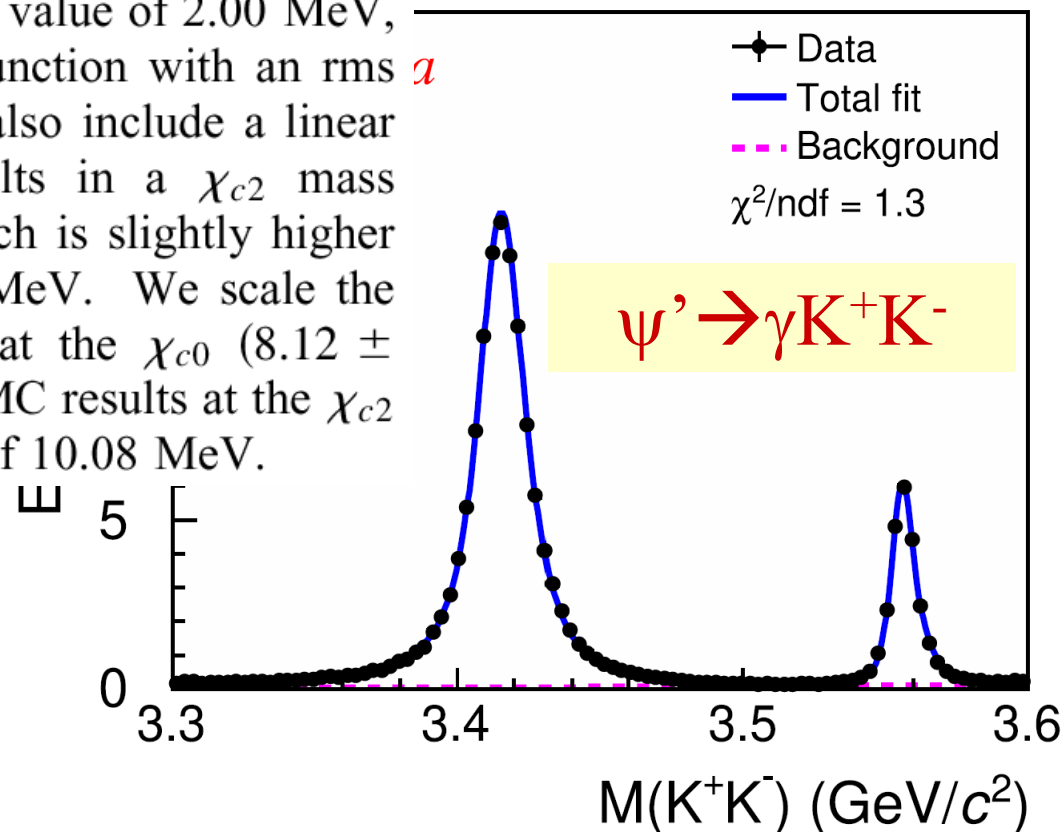
χ_{c0} 宽度测量!

质量分辨对宽度造成的系统误差



BES, PRL81, 3091(1998)

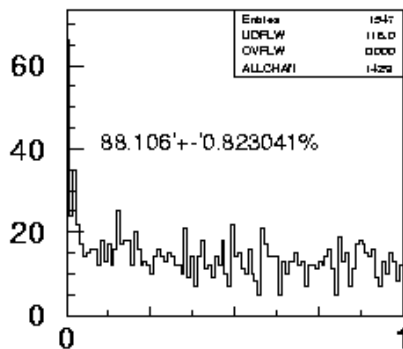
For the $\pi^+\pi^-$ channel, we first fit the invariant mass region between 3.5 and 3.6 GeV with a Breit-Wigner resonance with $\Gamma_{\chi_{c2}}$ fixed at the PDG value of 2.00 MeV, smeared by a Gaussian resolution function with an rms width that is allowed to float. We also include a linear background function. The fit results in a χ_{c2} mass resolution of 7.83 ± 1.04 MeV, which is slightly higher than the MC result of 6.31 ± 0.11 MeV. We scale the MC value for the mass resolution at the χ_{c0} (8.12 ± 0.23 MeV) by the ratio of the fitted MC results at the χ_{c2} and get a mass resolution at the χ_{c0} of 10.08 MeV.



BESIII, CPC49, 091001 (2025)

粒子鉴别效率的系统误差

Table 2: Particle ID efficiencies ($Prob_{ID} > 0.01$) for π^+ , π^- , p and $\bar{p}$ samples selected from $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ (in %).



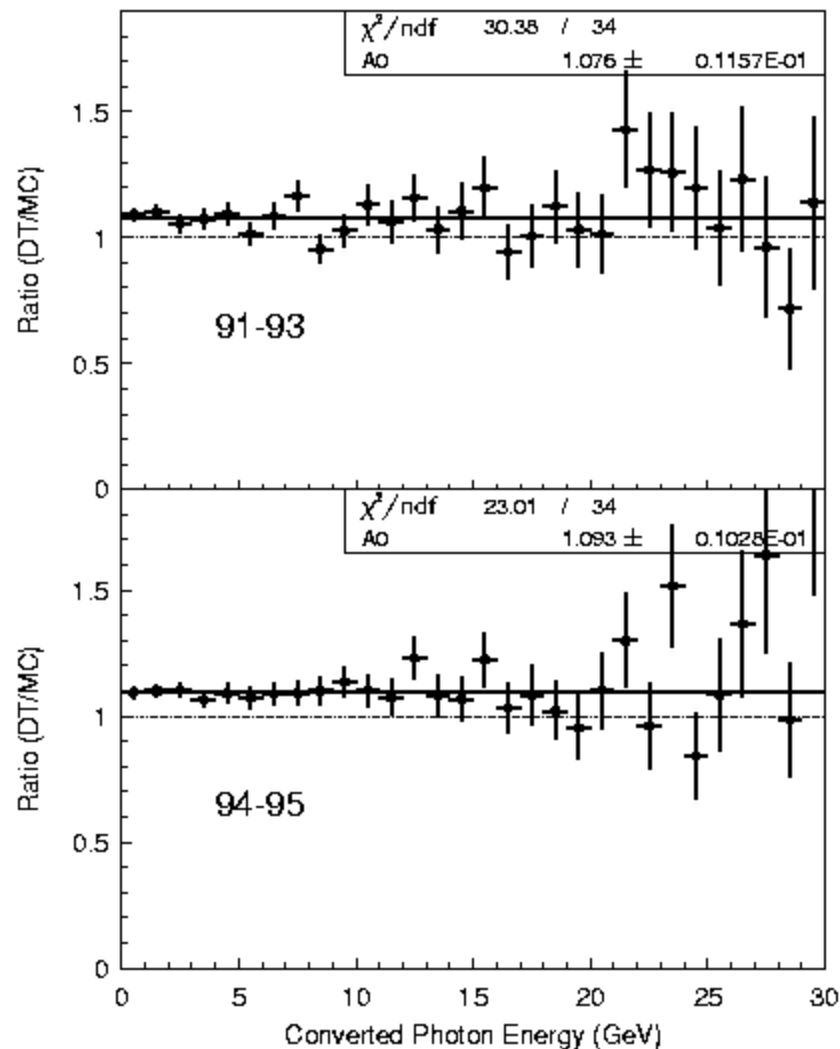
Particle	data	Monte Carlo
π^-	94.64 ± 0.60	88.55 ± 0.78
π^+	93.69 ± 0.64	87.67 ± 0.81
$\bar{p}$	87.76 ± 0.84	86.50 ± 0.84
p	88.11 ± 0.82	88.44 ± 0.79

数据和Monte Carlo模拟对 $\pi^+\pi^-$ 和质子、反质子粒子鉴别的效率。由 $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ 测得。

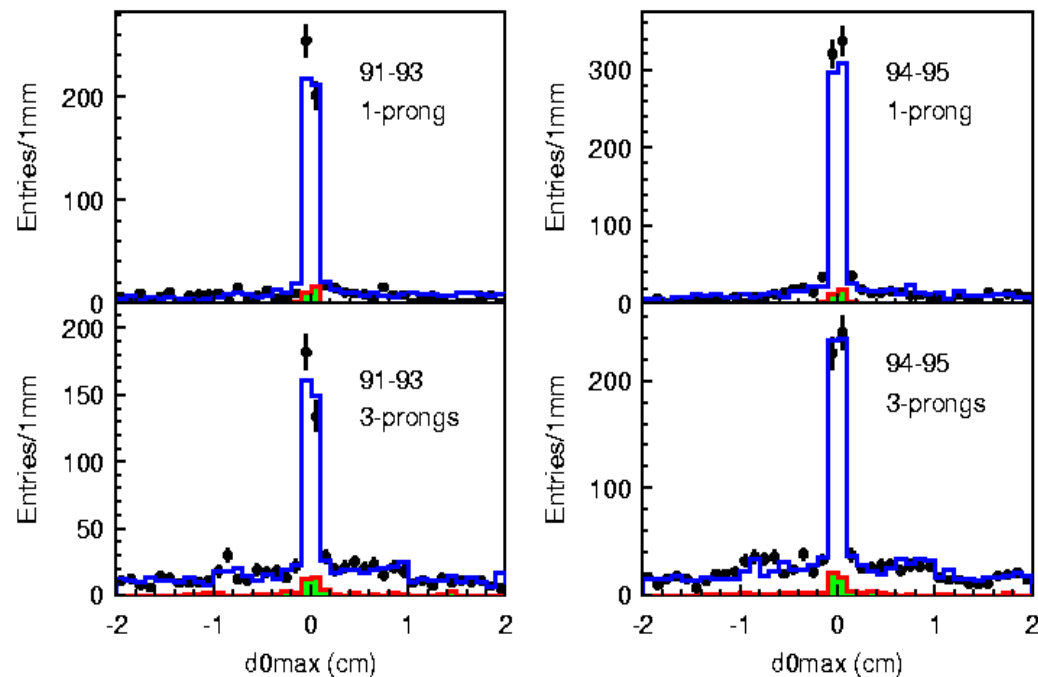
MC efficiency corr. factor: 1.154 ± 0.028

ALEPH对探测器的模拟

Gamma-conversion, 数据比Monte Carlo大近10%。



末态粒子与探测器物质的核相互作用，数据比Monte Carlo小约10%。



相关误差

- 来源不完全独立的误差。

- 最简单的例子

寻迹效率、几何接受度

计算事例总数时要计算其误差，在计算分支比时仍要计算。事实上，二者会有一部分抵消。

- 一个实验组中总数总是100%相关的，在本组内对比时其误差不应考虑在内。相同的软件，如运动学拟合也带来相关的误差。两个独立做同一道分析的人结果仍然不是相互独立的，因为他们的数据样本有很大程度的重叠。
- 高精度测量时非常重要。

有效数字修约

• 测量误差一般取1-2位有效数字

对于位数很多的近似数，当有效位数确定后，其后面多余的数字应予舍去，而保留的有效数字最末一位数字应按下面的舍入规则进行凑整：

- 1) 若舍去部分的数值，大于保留部分的末位的半个单位，则末位加 1。
- 2) 若舍去部分的数值，小于保留部分的末位的半个单位，则末位不变。
- 3) 若舍去部分的数值，等于保留部分的末位的半个单位，则末位凑成偶数，即当末位为偶数时则末位不变，当末位为奇数时则末位加 1。

例如，按上述舍入规则，将下面各个数据保留四位有效数字进行凑整：

原有数据	舍入后数据
3.14159	3.142
2.71729	2.717
4.51050	4.510
3.21550	3.216
6.378501	6.379
7.691499	7.691
5.43460	5.435

误差合并与传播

- 直接测量误差与间接测量误差

$$Y=f(X_1, X_2, \dots, X_N)$$

$$\sigma^2(Y) = \left[\frac{\partial f}{\partial X_1} \right]^2 \sigma^2(X_1) + \left[\frac{\partial f}{\partial X_2} \right]^2 \sigma^2(X_2) + \dots + \left[\frac{\partial f}{\partial X_N} \right]^2 \sigma^2(X_N) \\ + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial X_i} \frac{\partial f}{\partial X_j} \frac{1}{n} \sum_{k=1}^n \delta X_{ik} \delta X_{jk}$$

不同的误差合并与传播方式有差别，高能物理实验中常假设误差间无关联，所有误差都较小且近似按正态分布。（add in quadrature $\neq$ quadratic sum）

关联项

**各项独立时
为零。**

常用误差传播公式

(1) $U = aX \pm bY$

$$\sigma^2(U) = a^2\sigma_X^2 + b^2\sigma_Y^2 \pm 2ab\text{cov}(X, Y)$$

(2) $U = aXY$

$$\frac{\sigma^2(U)}{U^2} = \frac{\sigma_X^2}{X^2} + \frac{\sigma_Y^2}{Y^2} + \frac{2\text{cov}(X, Y)}{X \cdot Y}$$

(3) $U = aX^{\pm b}$

$$\frac{\sigma(U)}{U} = \frac{b \cdot \sigma_X}{X}$$

(4) $U = ae^{\pm bX}$

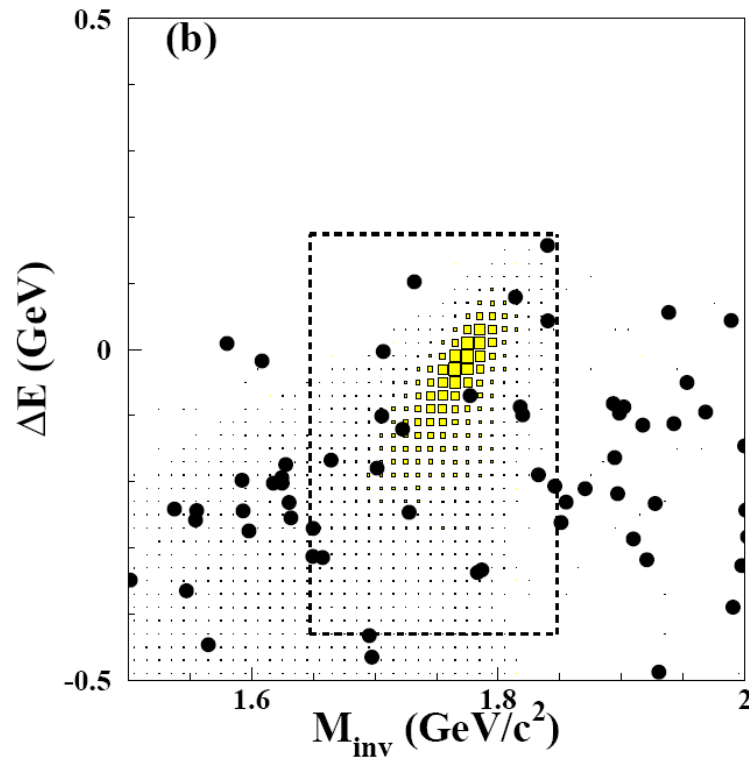
$$\frac{\sigma(U)}{U} = b \cdot \sigma_X$$

(5) $U = a \ln(\pm bX)$

$$\sigma(U) = \frac{a \cdot \sigma_X}{X}$$

**非常有用，
学会推导！**

避免Bias—Blind Analysis



Belle II:
search for $\tau \rightarrow \mu \gamma$ and $\tau \rightarrow e \gamma$

BESIII:
Precision R measurement

$$R = \frac{N_{had}^{obs} - N_{bkg} - \sum_l N_{ll} - N_{\gamma\gamma}}{\sigma_{\mu\mu}^0 \cdot \epsilon_{had} \cdot \epsilon_{trg} \cdot (1 + \delta) \cdot L}$$

总 结

- 简单介绍了正负电子对撞物理的数据分析的基本概念，介绍了误差分析的基本方法。
- 了解探测器，了解其工作原理和性能，了解Monte Carlo模拟。永远记住，完美的模拟是不可能存在的。
- 严肃对待分析的每一个环节。不要用似是而非的语言描述实验。
- 对结果的评价要与数据给出的信息匹配。决不说超出实验数据给出的信息的话。

谢谢！

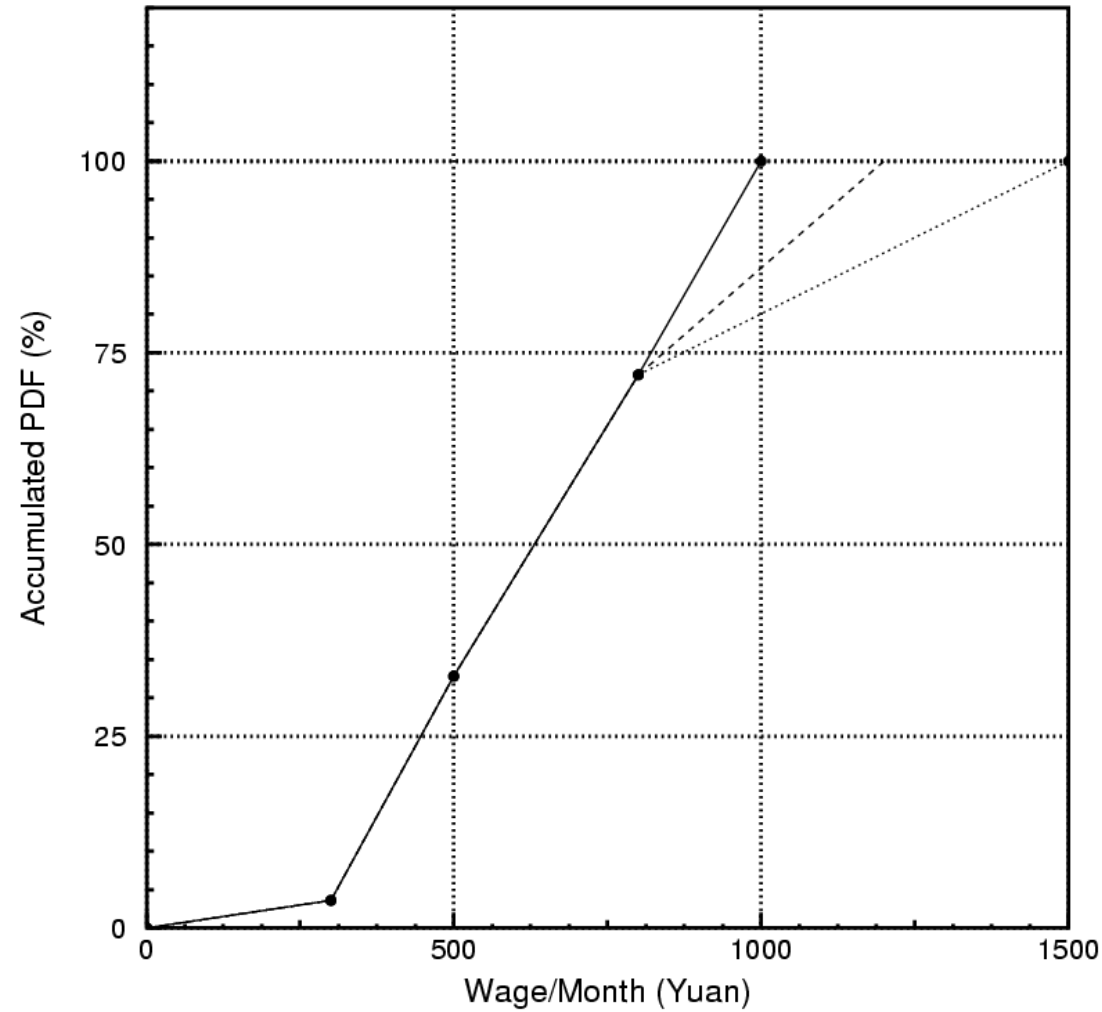
希望对大家有所帮助。

我国农民工收入主要集中在500至800元之间

<http://www.cctv.com/news/china/20060416/100220.shtml>

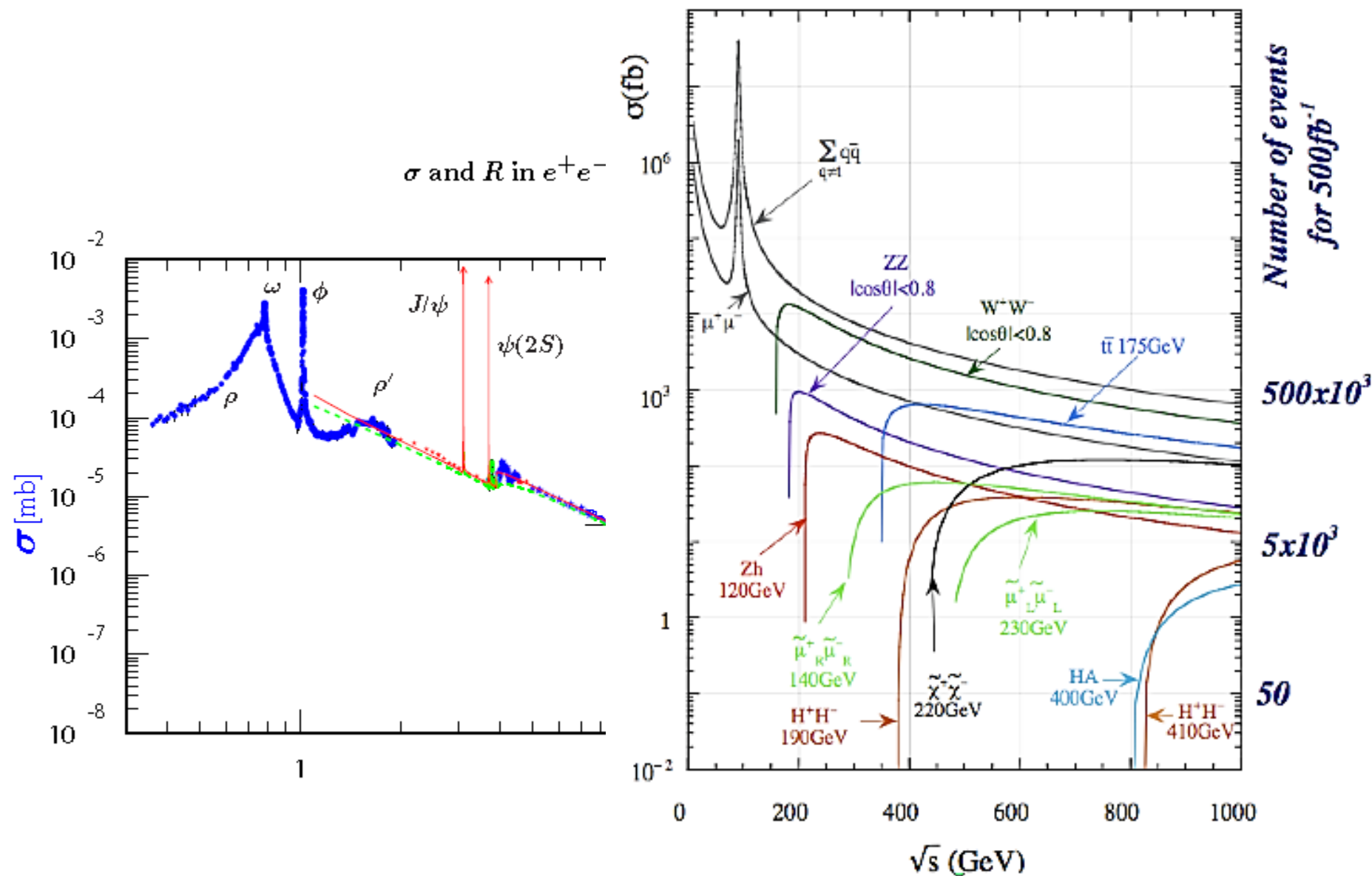
- 新华社北京2006年4月16日电（记者 徐京跃、刘奕湛）国务院研究室近日发布了《中国农民工调研报告》，对正在城镇务工和返乡农民工进行了问卷调查。
- 在收入方面，被调查农民工的月工资均比较低，主要集中在500至800元之间。
- 其中，每月收入在
 - 300元以下的占 3.58%，
 - 300至500元的占 29.26%，
 - 500至800元的占 39.26%，
 - 800元以上的占 27.90%。

我国农民工收入主要集中在500至800元之间



Wrong conclusion!

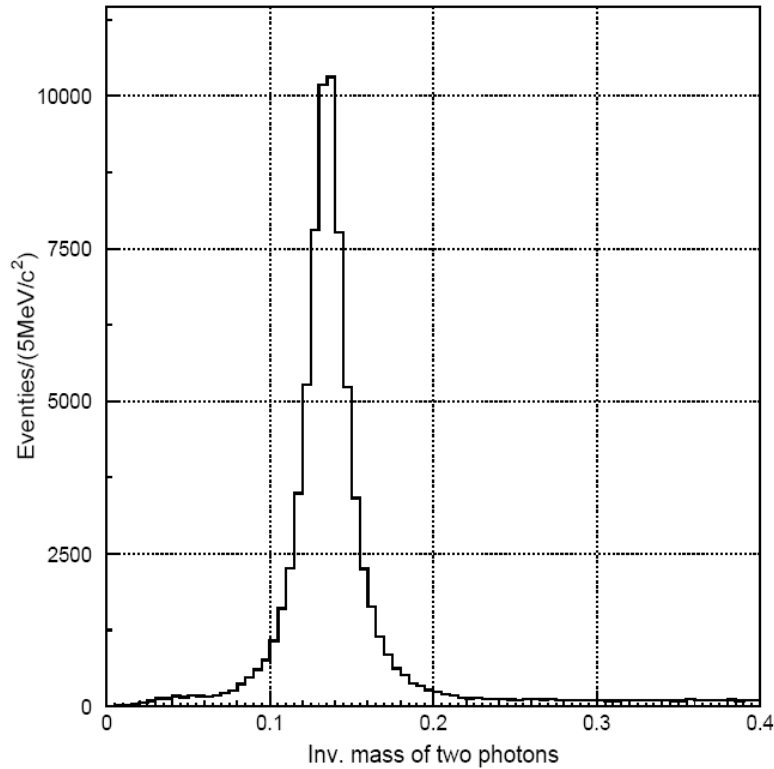
正负电子对撞机上的截面——物理事例 (从低能到高能)



什么是误差

- **误差：给出值与真值之间的差异。** (error/uncertainty)
 - 给出值：测量值，实验值，标称值，计算近似值等
 - 真值：客观值
- **一般说来：**
 - 真值是未知的——测量的意义
 - 因而误差也未知——误差分析的意义
- **有些情况真值是可以知道的；有些情况真值在相对意义上是知道的——因而误差分析是可能的。**
- **真值已知的情况**
 - 理论真值：三角形内角和； $e^+e^- \rightarrow l^+l^-$ 角分布
 - 定义值：米的定义；绝对零度
 - 相对真值：研究强相互作用时，弱电理论可以认为是完美的；
BESIII分析 ($J/\psi, \psi'$ 衰变) 中可以忽略弱作用

为什么需要好的探测器指标



$$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos \theta_{\gamma\gamma})$$

$$\Rightarrow \sigma m_{\gamma\gamma} \propto \sigma E_{\gamma}$$

$$\Rightarrow \sigma \bar{m}_{\gamma\gamma} \propto \sigma E_{\gamma} / \sqrt{n}$$

可见能量测量的不确定性要通过增大事例数来补偿！
带电径迹亦然。

Performance	BESII	BESIII	CLEOc
$\sigma p/p$	$1.7\%/\sqrt{1+p^2}$	$0.5\% @ p=1\text{GeV}$	$0.6\% @ p=1\text{GeV}$
$\sigma E/E$	$22\% / \sqrt{E}$	$2.5\% @ E=1\text{GeV}$	$2.2\% @ E=1\text{GeV}$
PartID	dE/dx+TOF	dE/dx+TOF	dE/dx+RICH
Coverage	80%	93%	93%

多项分布 (multinomial)

- **n次试验, A_i 出现 r_i 次成功的几率**

$$M(\vec{r}; n, \vec{p}) = \frac{n!}{r_1! r_2! \cdots r_l!} p_1^{r_1} p_2^{r_2} \cdots p_l^{r_l}$$

$$E(r_j) = n \cdot p_j$$

$$V(r_j) = n \cdot p_j \cdot (1 - p_j)$$

$$\text{cov}(r_i, r_j) = -n \cdot p_i \cdot p_j$$

- **应用**
 - 直方图
 - J/ ψ 衰变分支比 (p_i 很小, 可忽略其它贡献, 按Poisson分布处理)
 - τ 衰变分支比

多项分布 (multinomial)

216

ALEPH Collaboration / Physics Reports 421 (2005) 191–284

Table 4

Number of reconstructed events in 1991–1993 and 1994–1995 data sets in the different considered topologies

Reconstr. class	n_i^{obs} (91–93)	n_i^{obs} (94–95)
e	22,405	33,100
μ	22,235	32,145
h	15,126	22,429
$h\pi^0$	32,282	49,008
$h2\pi^0$	12,907	18,317
$h3\pi^0$	2681	3411
$h4\pi^0$	458	499
$3h$	11,610	17,315
$3h\pi^0$	6467	9734
$3h2\pi^0$	1091	1460
$3h3\pi^0$	124	150
$5h$	60	105
$5h\pi^0$	36	59
Class 14	4834	7100
Sum	132,316	194,832

小的统计误差得益于服从多项分布。