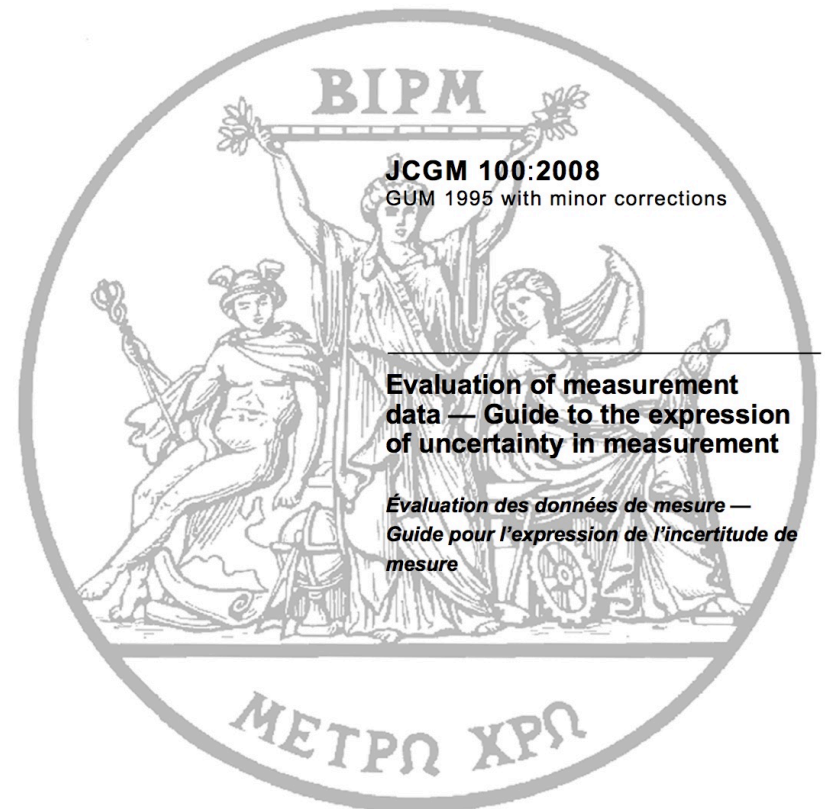
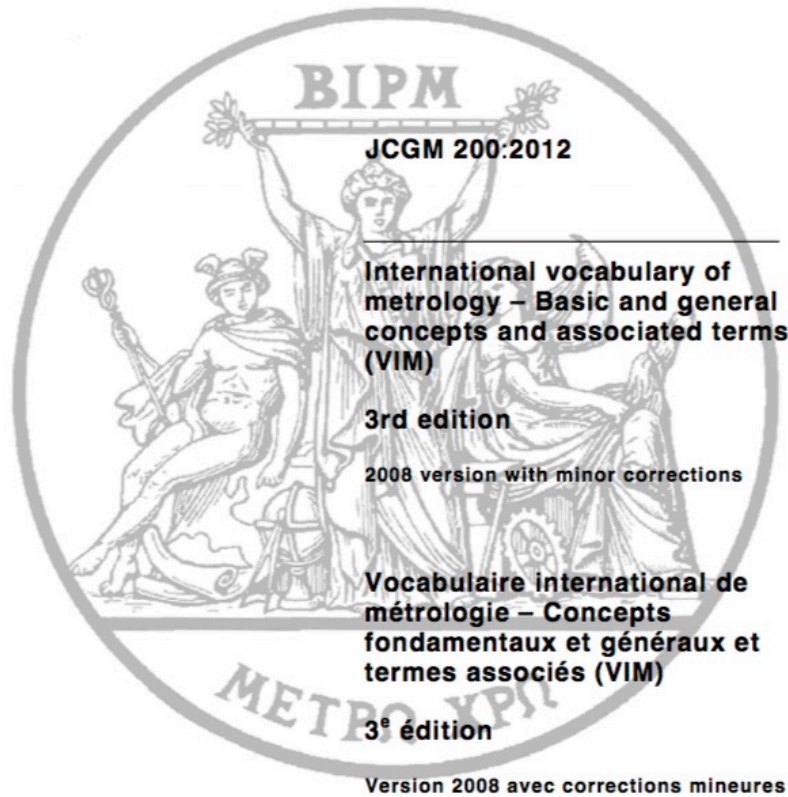


高能物理数据分析中的 误差

李刚

Li.Gang@mail.ihep.ac.cn

Two references



相关的概念和方法 有历史的发展和一定的标准

The following seven organizations* supported the development of this *Guide*, which is published in their name:

- BIPM: Bureau International des Poids et Mesures
- IEC: International Electrotechnical Commission
- IFCC: International Federation of Clinical Chemistry**
- ISO: International Organization for Standardization
- IUPAC: International Union of Pure and Applied Chemistry
- IUPAP: International Union of Pure and Applied Physics
- OIML: International Organization of Legal Metrology

Outline

- ◆ 测量和粒子物理中的测量
- ◆ 误差相关的概念
- ◆ 误差估计 – – 举例说明
- ◆ 小结

测量

输出新的知识/信息
必须输入知识/信息

Measurement

1  [uncountable] the act or the process of finding the size, quantity or degree of something

- *the metric system of measurement*
- *Accurate measurement is very important in science.*

 Collocations

 Culture

2  [countable, usually plural] the size, length or amount of something

- *to take somebody's **chest/waist measurement***
- *Do you know your measurements (= the size of parts of your body)?*
- *The exact measurements of the room are 3 metres 20 by 2 metres 84.*

- **定义:** Set of operations having the object of determining a value of a quantity (以确定一个量的值为目的的一系列操作)
- **描述:** The objective of a **measurement** is to determine the **value** of the **measurand**, that is, the value of the **particular quantity** to be measured. A measurement therefore begins with an appropriate specification of the measurand, **the method of measurement**, and the **measurement procedure**.

- 测量是个不断理解（系统误差）的过程
- 测量是个不断理解测量对象的过程
- 测量的精华在于修正（定义，测量方法，误差）

几个🍎



- 测体重：人(苍蝇)
- 测小行星的质量
 - $F=ma$
 - $F=GM_1M_2/r^2$

Error and uncertainty

It is now widely recognized that, when all of the known or suspected components of **error** have been **evaluated** and the **appropriate corrections** have been applied, there still remains an **uncertainty** about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured.

测量问题的核心

- 精度，精度，精度！
- 精度决定你的工作是否有意义
- 精度决定你的工作量
 - 随着要求精度的提高，测量难度和工作量会几何量级的增加
- 最终，在足够好的精度下，你开始怀疑测量（定义，测量方法，测量原理，测量过程）
—— 恭喜你！

Bearing in mind: 一些测量是可以自动化的

但在高能物理中很难做到:

测量设备复杂, 原理复杂, 环境条件复杂、不理想, 测量量的定义不完备...

高能物理测量的复杂性

- 用探测器采集的，并经过刻度、重建的数据里得到测量结果。实际上数据的采集到处理的整个过程都需要监控和了解。
- 检查、比较数据和模拟以及更多因素，做必要的修正
- 分析中往往还需要做二次修正，修正后会给出更新的误差，一般修正后的误差会更小，也使结果更加合理和可靠

知止而后有定，定而后能静，静而后能安，安而后能虑，虑而后能得

其实后面可以看到，系统误差的一大类也是由统计方法来估计的，因此也受统计量的限制。

→

粗略来说，大多数情况下把系统误差做到统计误差相当，差不多就到极限了。

大多数情况下，可以预期达到的精度。



数据：基础和出发点

Monte Carlo: reference/guiding star

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

测量是一个比较过程，本质上是做除法

做数据分析必要的知识准备

- 需要了解和掌握的知识和工具
 - 物理知识：基本概念，比如粒子物理基本知识，运动学(相空间);有量子场论的知识，不一定需要能做复杂的推导，但一定要有比较清楚的物理概念，读懂理论文献。
 - 仪器工具：加速器，探测器，计算机编程，分析工具（c++, fortran, root, RooFit, Minuit）
 - C++, g++, Linux及其相关软件，比如程序的编译调试（(g)Make），脚本语言，shell, python, ...
- 开始最消耗时间的两件事情：画图和发作业。早点把时间解放出来，做更有意义的事情——注意自动工具的学习

测量的出发点和参考

已经知道什么？

将要测量什么？

需要、能达到到什么精度

Phase space, 相空间

一般形式

$$d\Gamma = \frac{(2\pi)^4}{2M} |\mathcal{M}|^2 d\Phi_n (P; p_1, \dots, p_n),$$

两体衰变

$$d\Gamma = \frac{1}{32\pi^2} |\mathcal{M}|^2 \frac{|\mathbf{p}_1|}{M^2} d\Omega ,$$

三体衰变

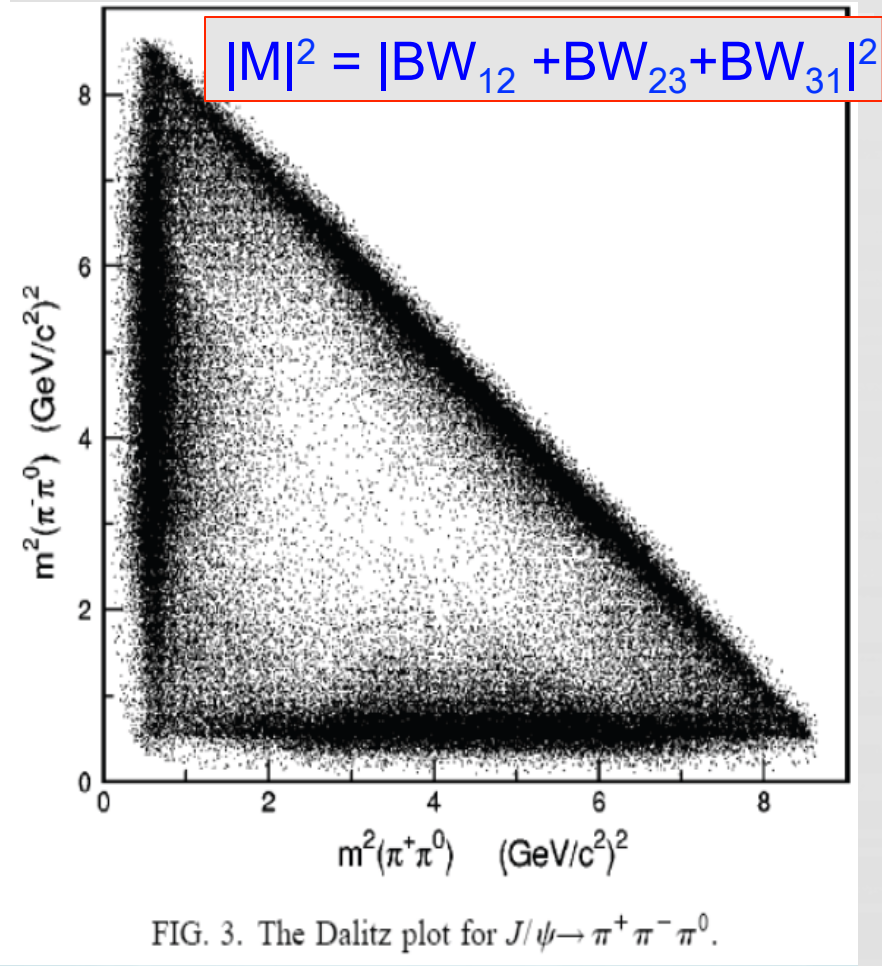
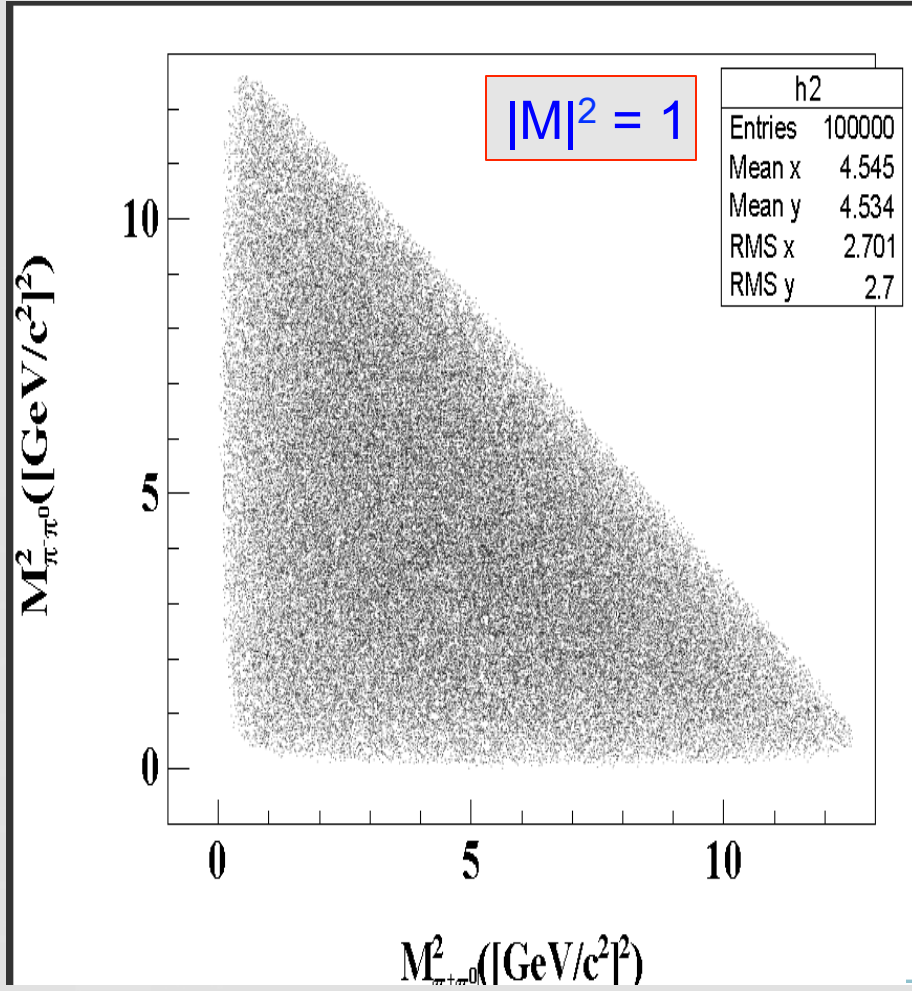
$$d\Gamma = \frac{1}{(2\pi)^5} \frac{1}{16M} |\mathcal{M}|^2 dE_1 dE_2 d\alpha d(\cos\beta) d\gamma .$$

$$d\Gamma = \frac{1}{(2\pi)^5} \frac{1}{16M^2} |\mathcal{M}|^2 |\mathbf{p}_1^*| |\mathbf{p}_3| dm_{12} d\Omega_1^* d\Omega_3 ,$$

不要忽视 Phase Space

- 它是平庸的，PHSP产生子是最简单的考虑
- 它是观测的参考点---任何偏离相空间分布的现象都意味这某种动力学效应；不熟悉相空间及其分布就意味着丧失了对动力学现象的敏感
- 相空间边缘的物理现象要多考察→排除相空间效应。
- 在分波分析当中用来代表直接N体衰变的贡献，这点一直在讨论当中，而且具体形式也不完全是真正的相空间，但是还叫做相空间。

$J/\psi \rightarrow \pi^+\pi^-\pi^0$ 的naive相空间和数据分布



截面测量

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

- 公式里各项都是误差的可能来源
- 主要由分析人员来测量的误差包含在 ϵ 中
 - 主要是由于模拟的不完美导致的：包含理论和实验两方面

质量测量：衰变产物

- $M^2 = (E_{cm} - E_1 - E_2 - \dots)^2 - (p_1 + p_2 + \dots)^2$

$$E_i^2 = p_i^2 + m_i^2$$

- 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）

干涉效应（动力学）

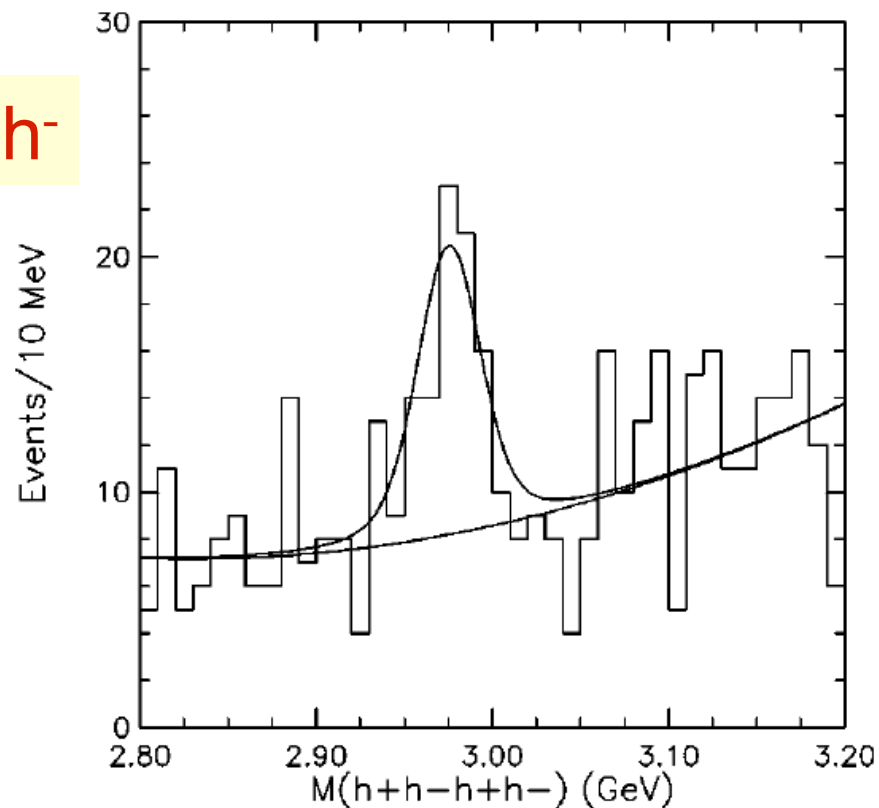
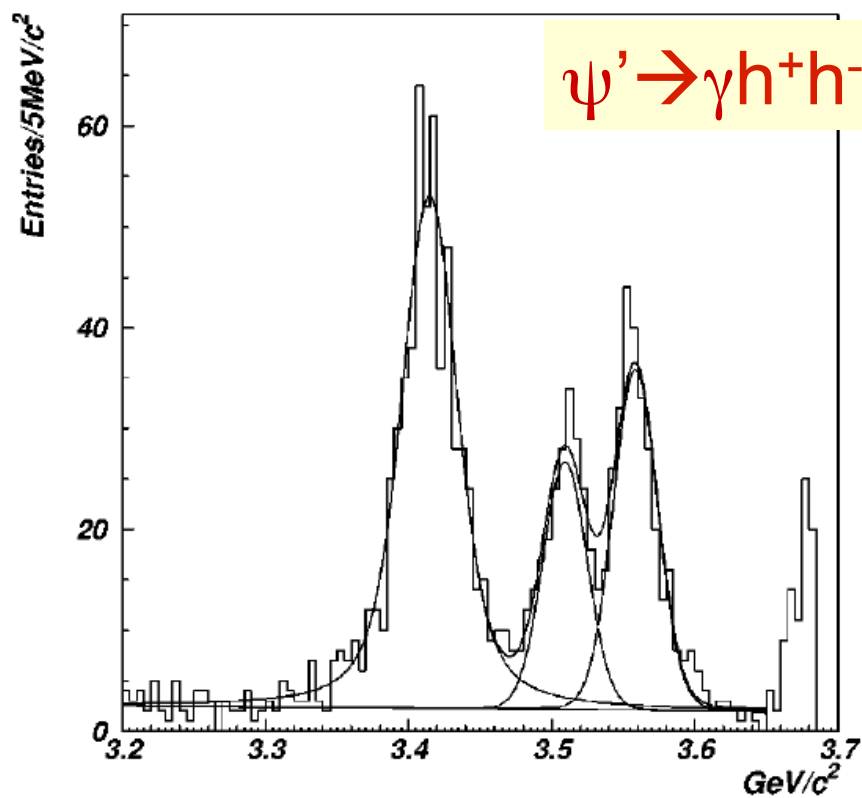
- 例如

1。Higgs质量测量

2。W的质量宽度

变通的办法

- 用精确已知的粒子刻度待测粒子。
- 例 χ_{c0} 、 η_c 质量测量。



$h^+h^-h^+h^-$ mass

宽度测量

- 衰变产物

拟合不变质量谱

- 对撞产生

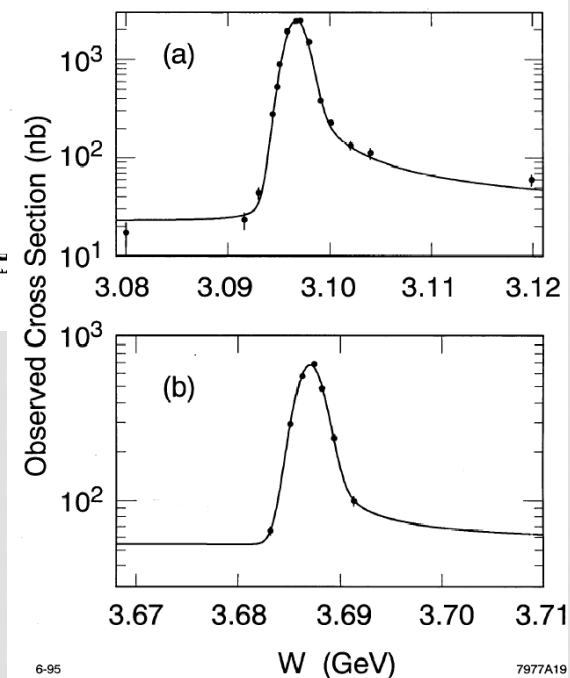
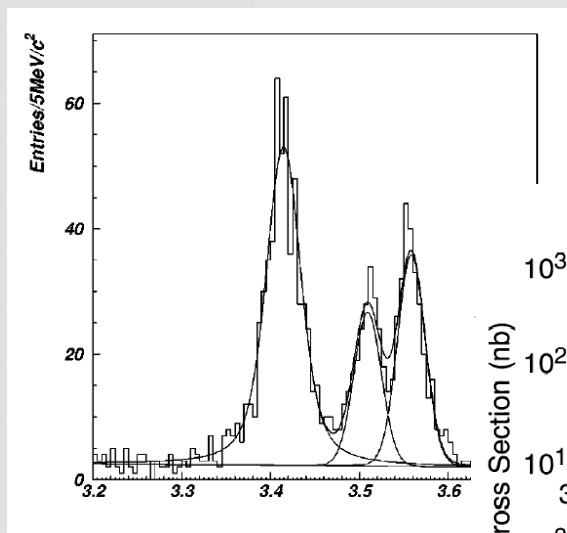
拟合产生截面

- 影响宽度测量的主要因素

1. 质量分辨或能散

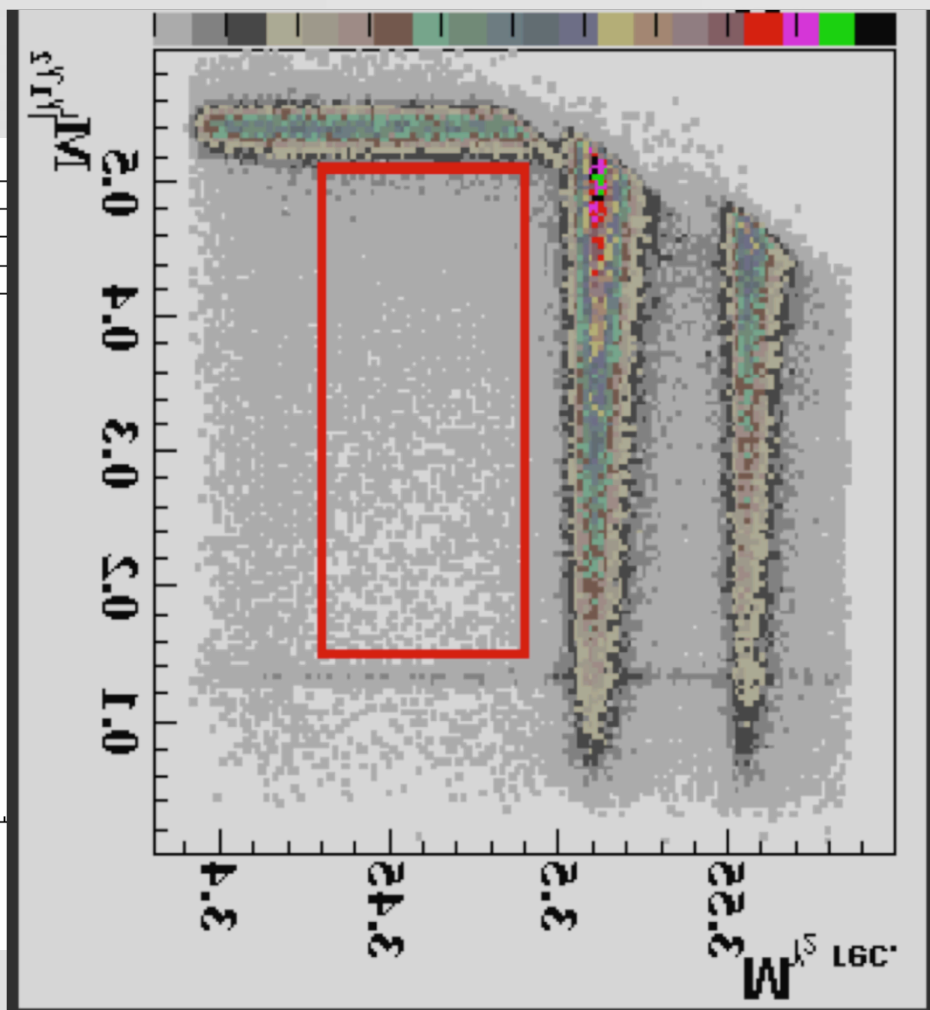
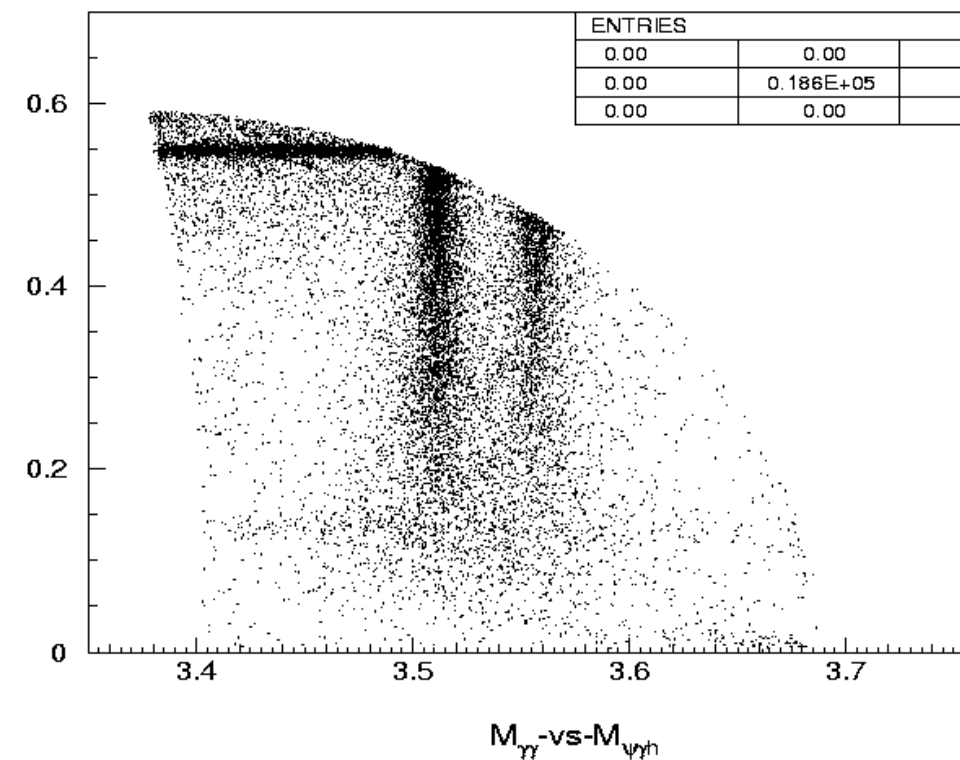
2. 本底形状

- 往往质量和宽度同时测量，但前者对质量分辨依赖很小。



非高斯形式的质量分辨

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



分支比测量

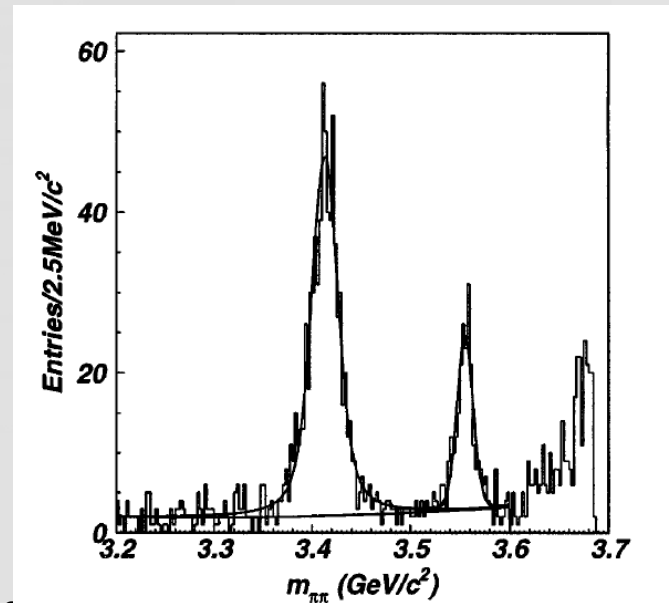
- 分支比计算公式

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

- 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

系统误差



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。

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)

误差分析

- 认识误差
- 误差的确定
- 误差的修正

Measurand

- 测量量

- 定义：可测量

- ✓ 例子： $F=GM_1M_2/r^2$

- ✓ 所以，测量前先了解下测量对象的定义很重要：误差分析已经开始了

- 一般不知道真值，也就无法严格确定其error

Error → uncertainty

- 误差：给出值与真值之间的差异（measured quantity value minus a reference quantity value）
 - ◆ 给出值：测量值，实验值，标称值，计算近似值等
 - ◆ 真值：客观值
 - ◆ 真值是未知的一一测量的意义
 - ◆ 因而误差也未知——知道误差，那还不修正？
- 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一一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；BES分析（ $J/\psi, \psi'$ 衰变）中可以忽略弱作用

uncertainty

- 意味着缺乏知识，缺乏了解，这才是误差分析的意义所在，误差分析其实是个探索的过程。
- 一般情况下：面对某种误差，无法做更多研究时，它就是uncertainty；研究其来源，了解其特点，把未知它变成已知的bias，认识深化了；修正bias → 结果更合理、可信，减小了uncertainty，提高了精度

误差的分类

- 统计误差和系统误差

- ◆ 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。

- ◆ 另一种分类：**type A** 和 **type B**

- ◆ **Type A**: 实用统计方法来估计

- ◆ **Type B**: 采用其它方法来估计

Chebyshev law of large numbers

X_1, X_2, \dots, X_n are independent random variables, and

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

$\forall \varepsilon > 0,$

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

Gaussian

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

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

利用统计的方法
进行测量的基础。

Chebyshev's inequality

X

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

$\forall \varepsilon > 0,$

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

or

$\forall k > 0:$

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

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^{-5}
10	0.01	---

即出现在不大于k倍标准偏差以外的几率不大于1/k²

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 $\hat{\sigma}_s = \frac{s}{\sqrt{2(n-1)}} \quad (\text{误差的误差})$

统计误差确定

- 物理量

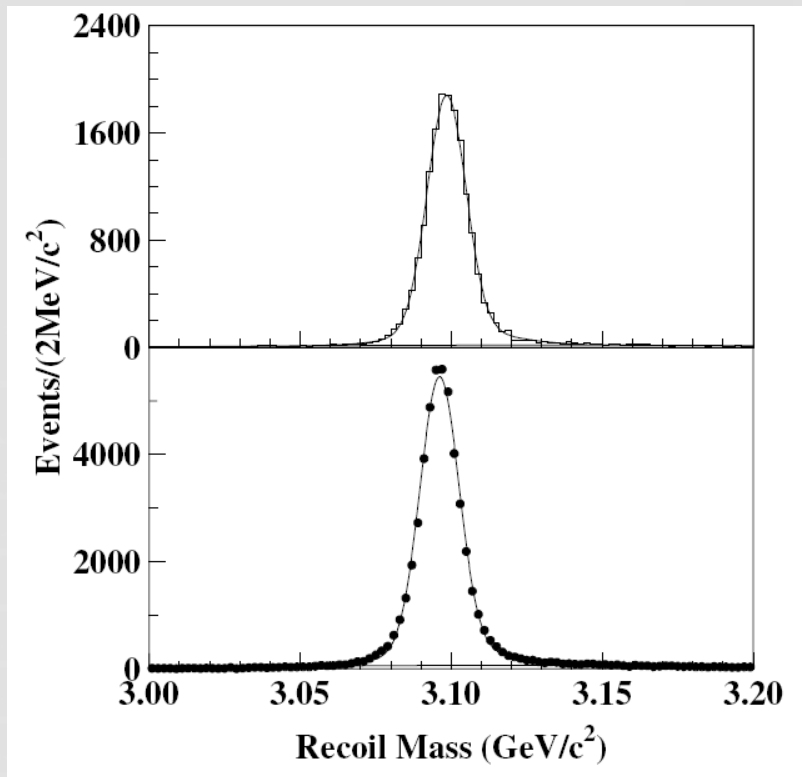
- 截面或分支比
- 质量（粒子或共振态）
- 宽度或寿命
- $I^G J^{PC}$ 或角分布
- 极化或不对称性
- ...

- 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

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



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

$$m_{J/\psi}, \sigma^2,$$

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

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

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

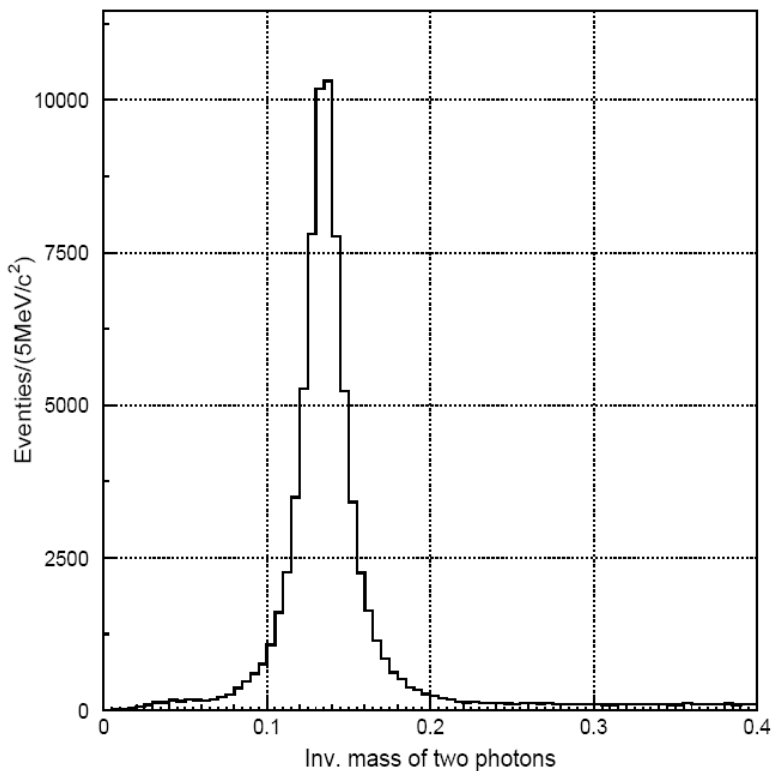
$$\hat{\sigma}^2(\bar{x}) = \hat{\sigma}^2 / n.$$

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

实际上可由程序拟合直接得到

(**root, roofit, paw, mn_fit, ... 等**)

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%/√1+p ²	0.6% @ p=1GeV	0.6% @ p=1GeV
$\sigma E/E$	22% /√E	2.5% @ E=1GeV	2.2% @ E=1GeV
PartID	dE/dx+TOF	dE/dx+TOF	dE/dx+RICH
Coverage	80%	93%	93%

计数实验

常见分布

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

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	$f(r; N, p) = \frac{N!}{r!(N-r)!} p^r q^{N-r}$ <p>$r = 0, 1, 2, \dots, N$; $0 \leq p \leq 1$; $q = 1 - p$</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)	$f(x; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp(-(x - \mu)^2 / 2\sigma^2)$ <p>$-\infty < x < \infty$; $-\infty < \mu < \infty$; $\sigma > 0$</p>	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\mathbf{x}; \boldsymbol{\mu}, V) = \frac{1}{(2\pi)^{n/2} \sqrt{ V }}$ $\times \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T V^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]$ <p>$-\infty < x_j < \infty$; $-\infty < \mu_j < \infty$; $\det V > 0$</p>	$\exp\left[i\boldsymbol{\mu} \cdot \mathbf{u} - \frac{1}{2}\mathbf{u}^T V \mathbf{u}\right]$	$\boldsymbol{\mu}$	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	$2n$

多项分布 (multinomial)

- n次试验, A_i 出现 r_i 次成功的几率

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

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

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

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

- 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。

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)

正态分布 (Normal/Gaussian)

- Poisson分布中x较大时 ($x > 20?$)

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

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

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

$N(0,1)$: 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：
 - ◆ 定义不完备
 - ◆ 实验条件实现不严格
 - ◆ 测量设备（测量不准，不稳定）
 - ◆ 实验、分析方法（运动学拟合）
 - ◆ 理论模型的应用（**lund**用在低能）
 - ◆ **Monte Carlo**的模拟（核作用的模拟）
 - ◆ 实验者的偏见（期望有新发现）
 - ◆ 理论预期的误导（标准模型总是对的）
 - ◆ 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

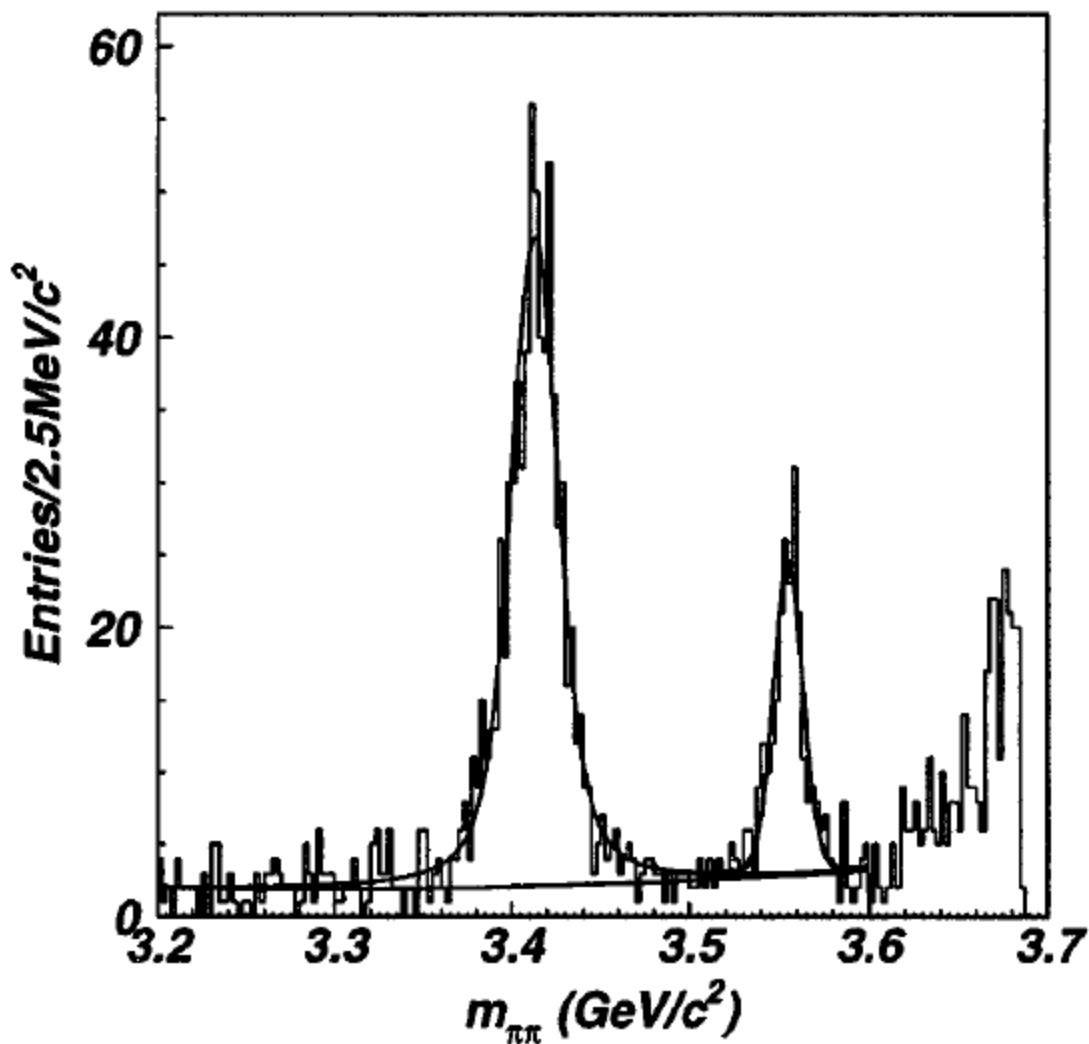
$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos\theta_{\gamma\gamma})$ 的误差与带电径迹也有关。

- 寻找变通的办法！对数据分析很重要！

系统误差的分析

- 模型误差、理论误差等较容易考虑（至少不必由实验者负责，但是你得知道它）。
- 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。

质量分辨对宽度造成的系统误差



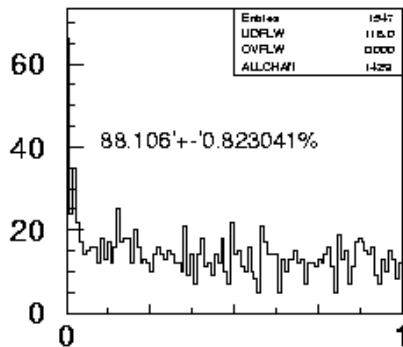
χ_{c0} 宽度测量!

用同样末态的
J/psi事例来刻度分辨

BES, PRL81, 3091(1998)

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

J/ψ-prbprp

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

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。
- 同时测量多个量时，往往需要给出相关矩阵

CEPC 测量误差初步考虑

σ , Br, m, Γ , ...

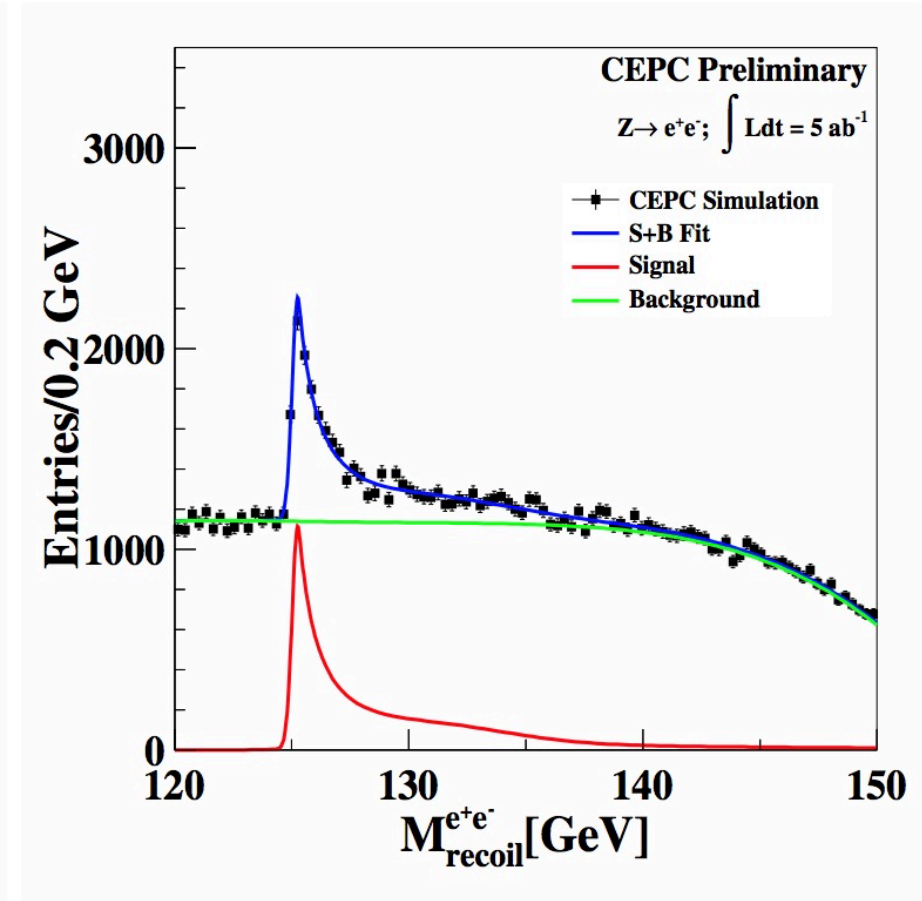
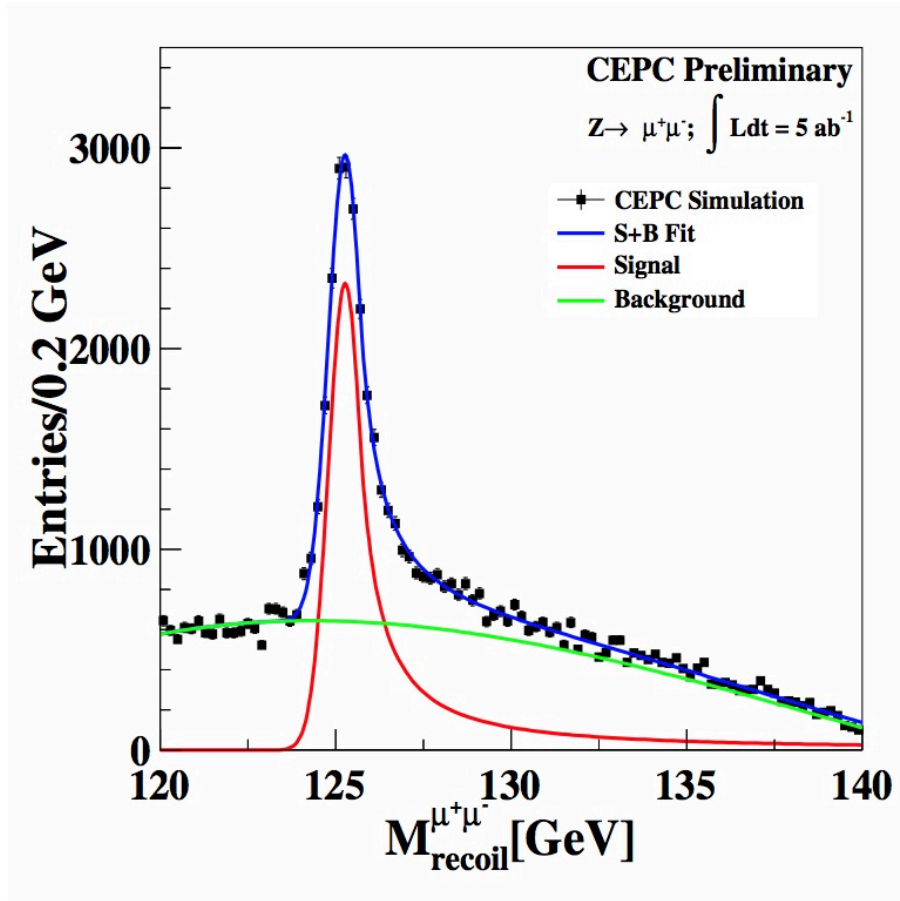
Experimental side

- ϵ_{MC} VS. ϵ_{DT}
 - Tracking & Pid(Flavor tag)
 - Difference between data and MC
 - Momentum/energy calibration
 - Beam measurement, beamstrahlung effect
 - Trigger efficiency
- Signal yields determination
- Integrated luminosity

Theoretical side

- ISR/FSR simulation
- Model of MC (Dynamics, higher order contribution)
- Hadronization model
- Multi-jets
- Correlation among jets/hadronization
- Interference between different intermediate states

Goal in CEPC preCDR: $\sim 6\text{MeV}$



Fit to the $ee/\mu\mu$ recoil mass spectra: only experimental uncertainties

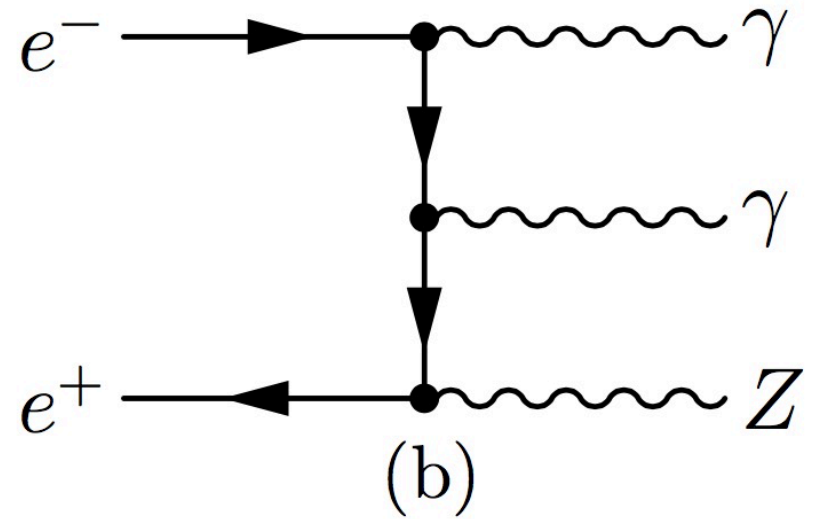
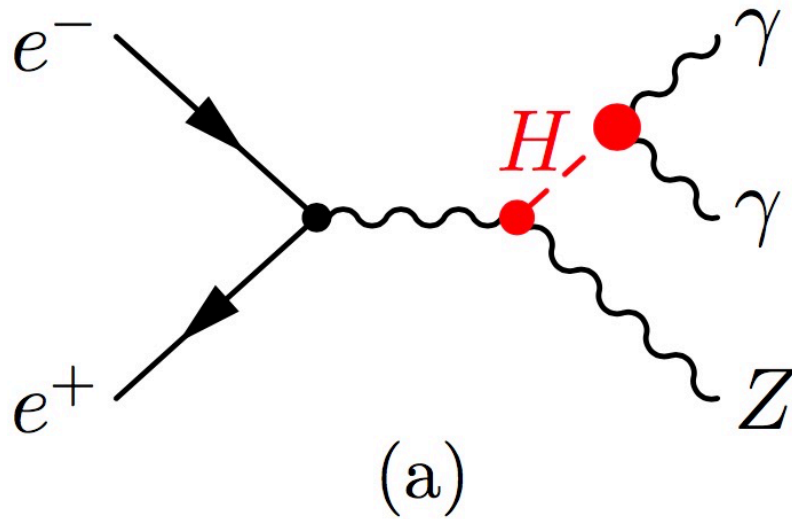
Let's take a simple example

$$e^+e^- \rightarrow H(\gamma\gamma)Z$$

[arXiv:1503.07830](#)

[arXiv:1505.06981](#)

Feynman diagrams at ee collider



experiment, it also causes the theoretical uncertainty through interference

For most cases except the interference effect: narrow width approximation adopted

$$\frac{1}{(s - M_H^2)^2 + M_H^2 \Gamma_H^2} = \frac{\pi}{M_H \Gamma_H} \delta(s - M_H^2)$$

Because $\Gamma_H \approx 4.2 \text{ MeV} \approx (3.4 \times 10^{-5}) M_H$, this is usually fine.

The cross section can be parameterized as

$$\frac{d\sigma_{ee \rightarrow Z\gamma\gamma + ZH(\gamma\gamma)}}{dm} = C(m) + \frac{1}{D(m)} [S(m) + (m^2 - \Gamma^2)I(m)]$$

Where

m = the diphoton invariant mass

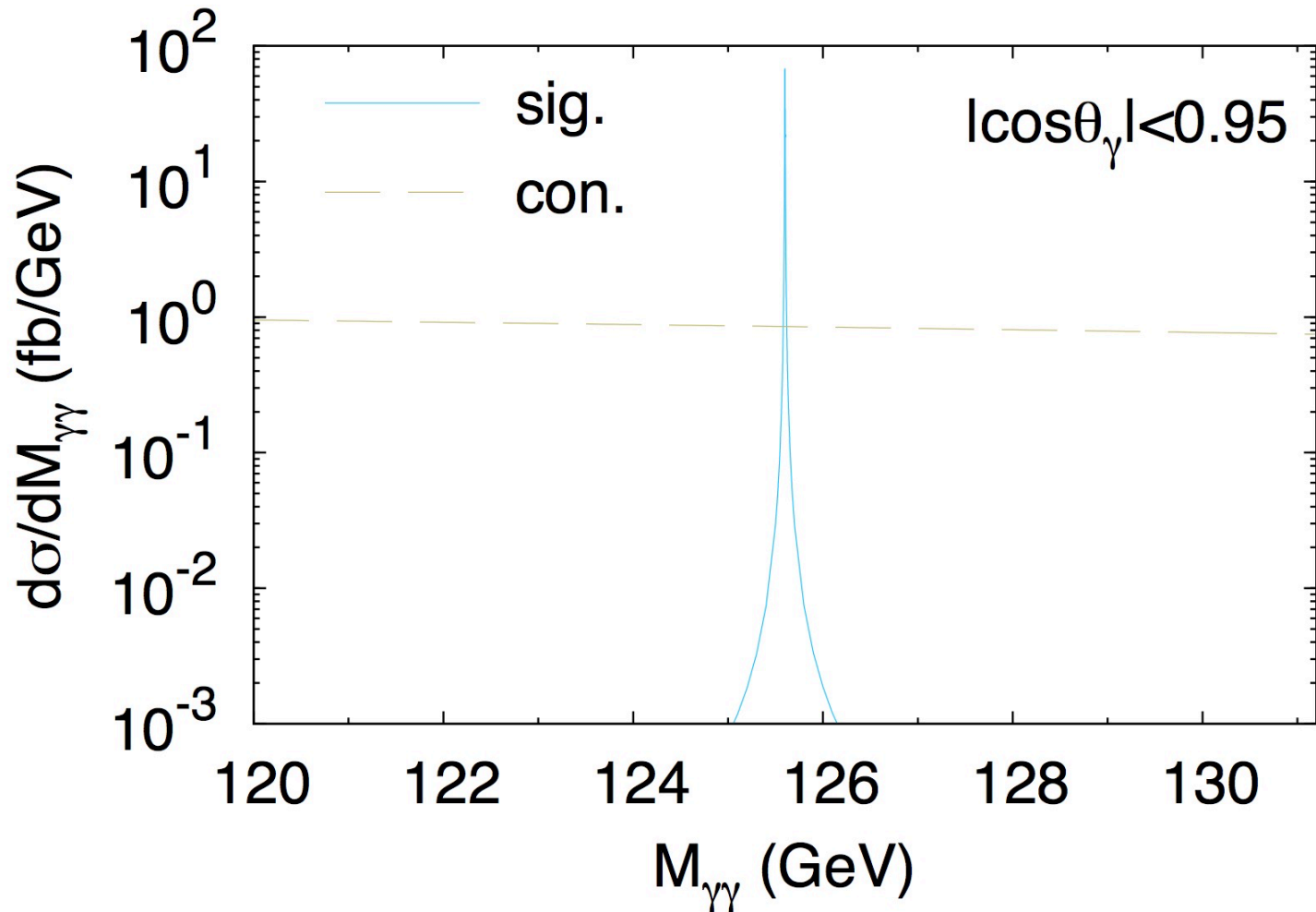
$D(m) = (m^2 - M_H^2)^2 + M_H^2 \Gamma_H^2$

$C(m)$ = background

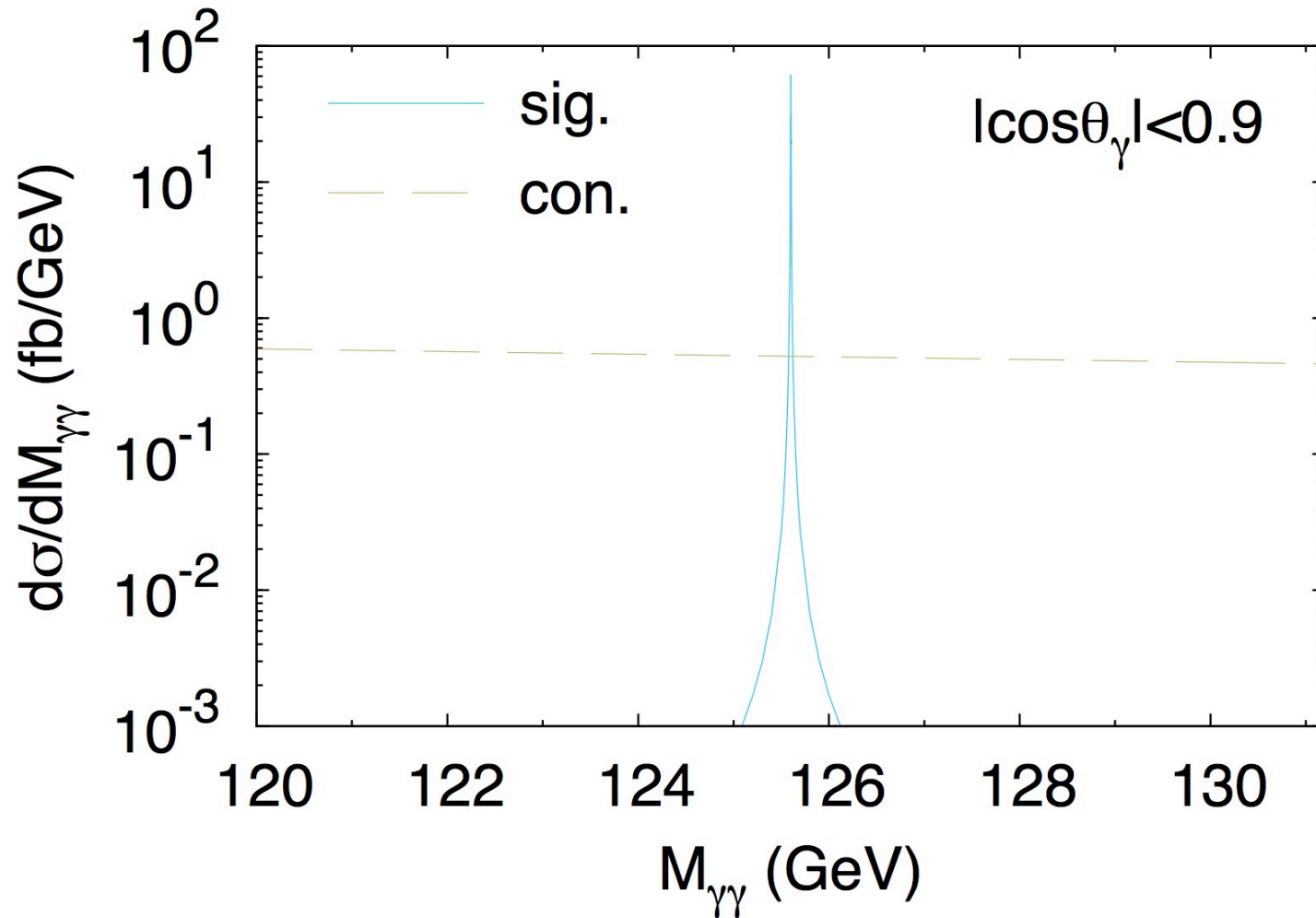
$S(m)$ = signal (with small contribution from interference)

$I(m)$ = pure interference term

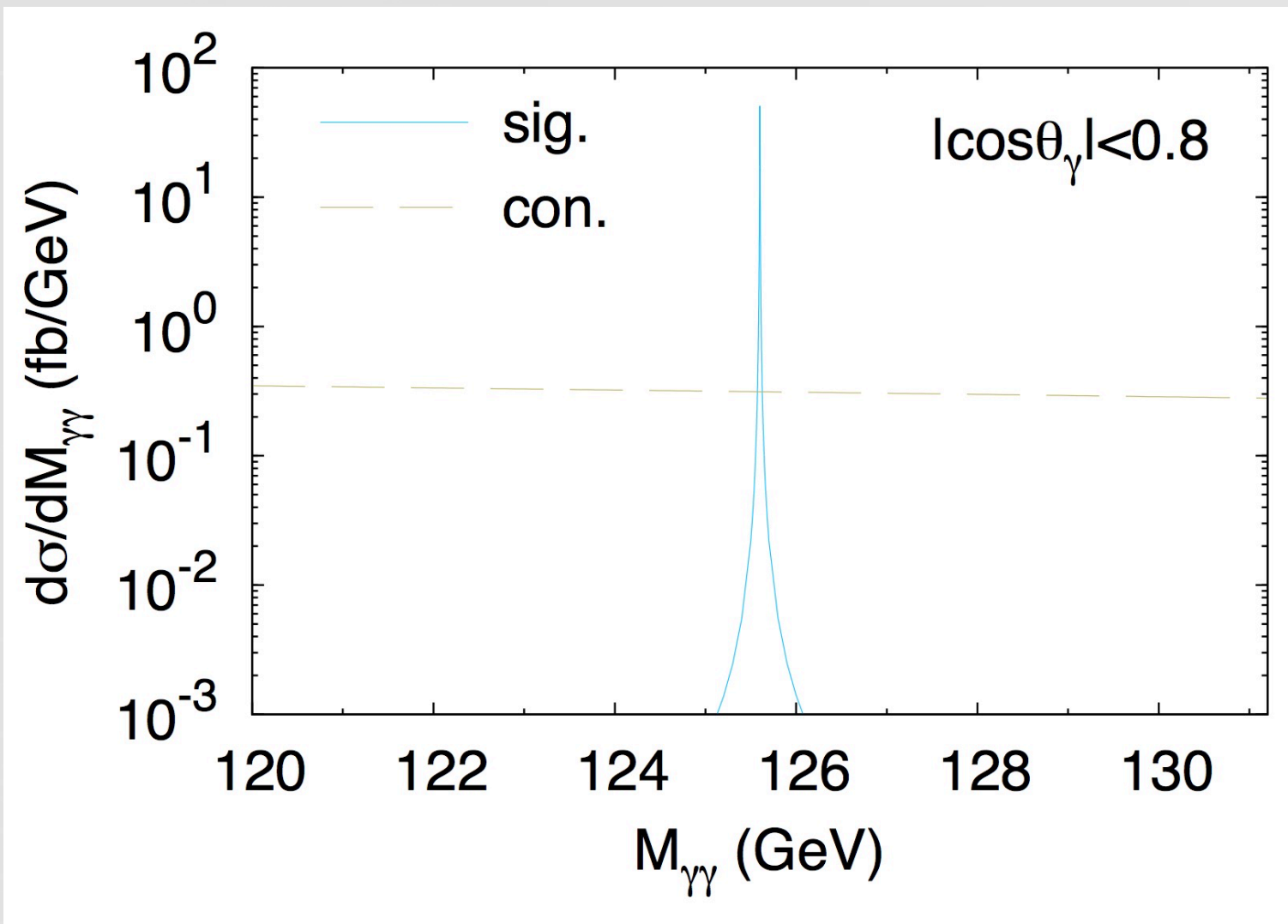
$ee \rightarrow Z \gamma\gamma$ sensitive to the polar angle cut

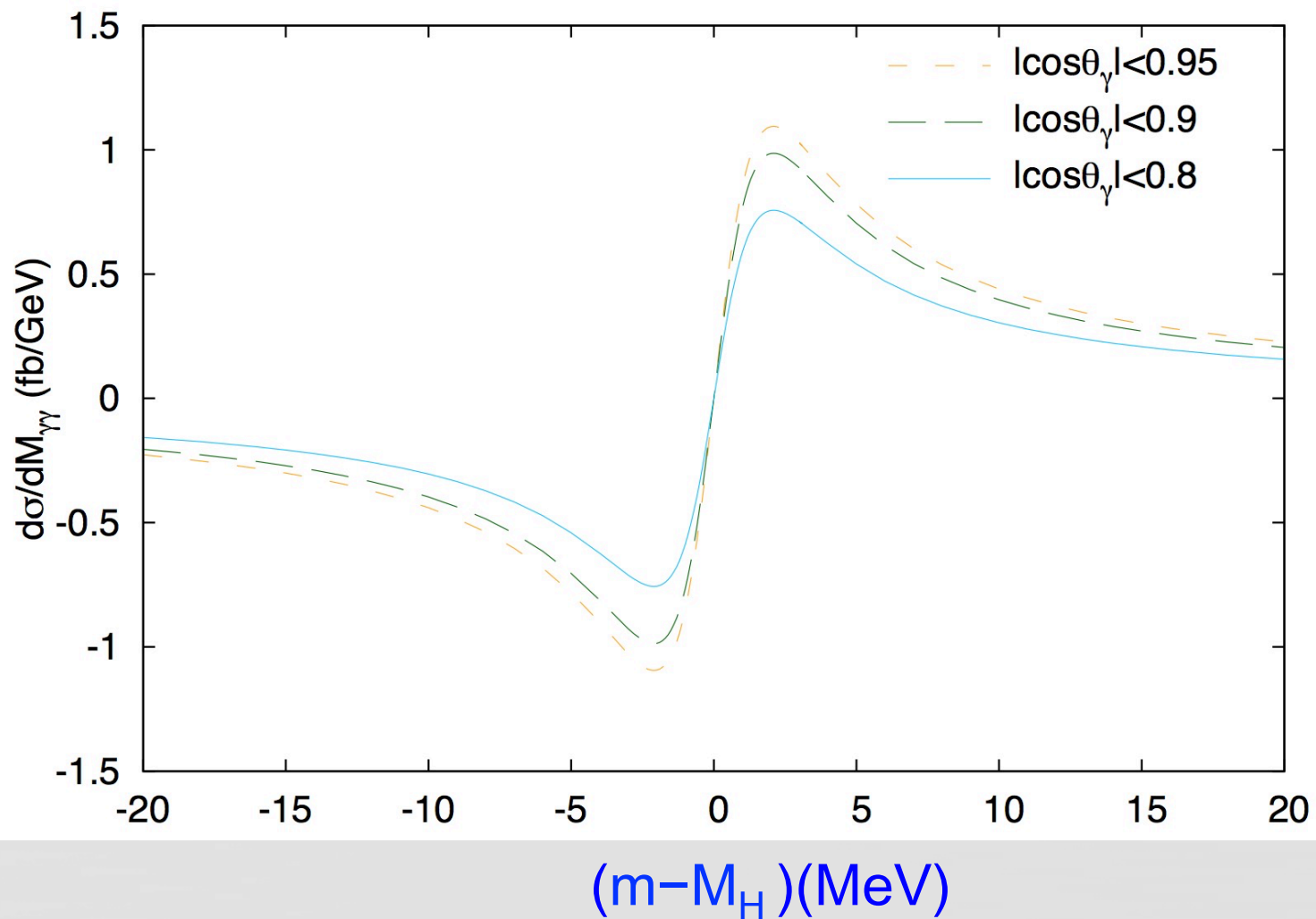


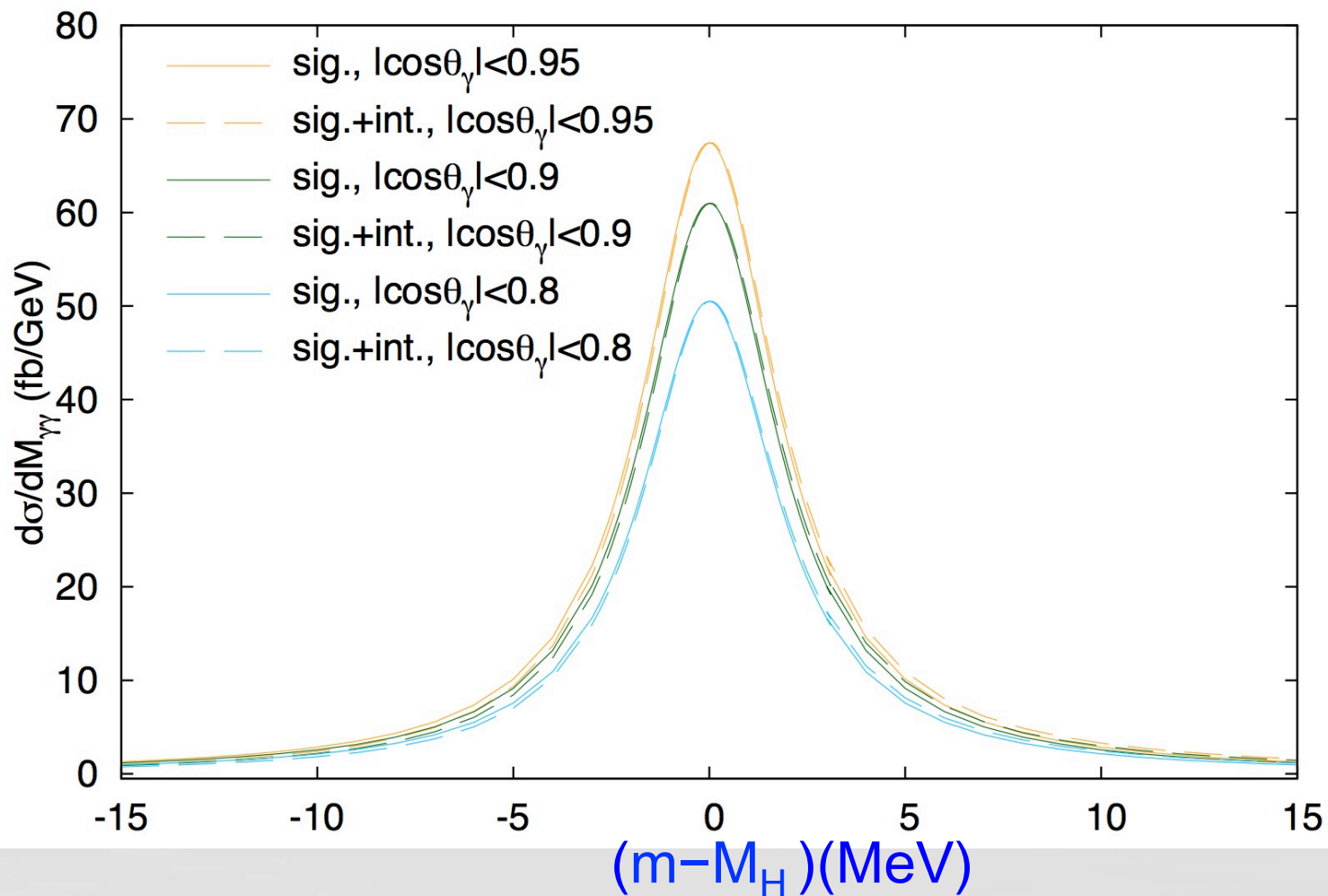
$ee \rightarrow Z \gamma\gamma$ sensitive to the polar angle cut



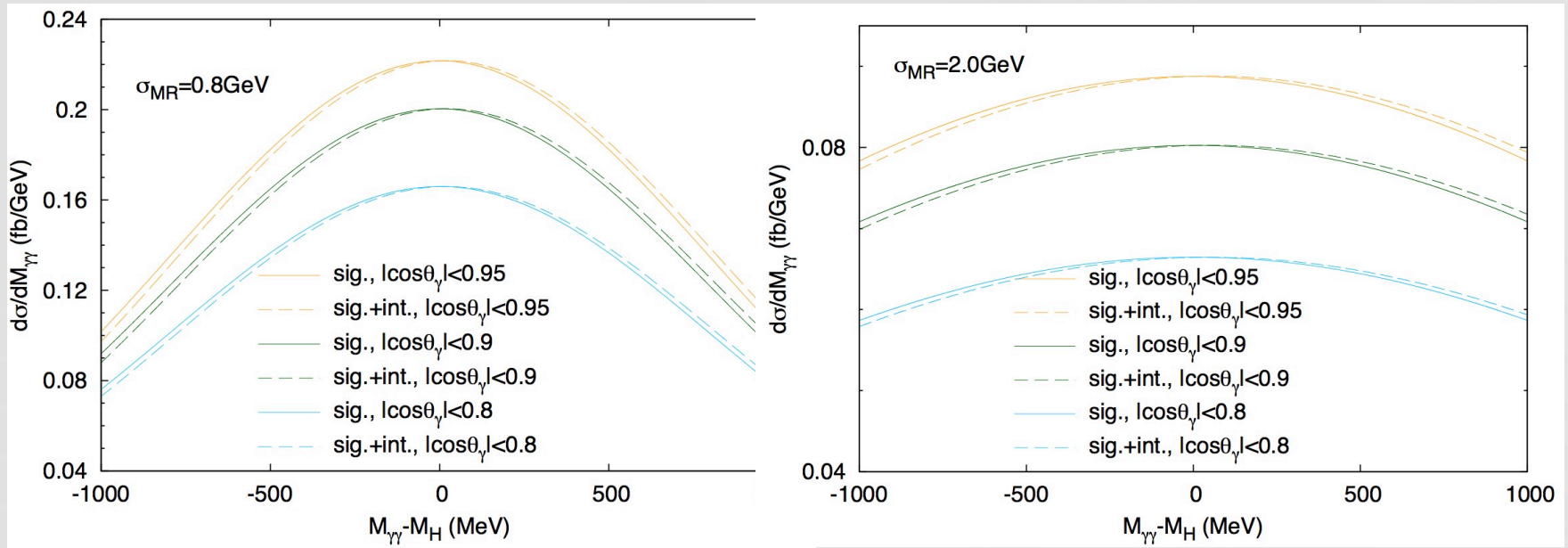
$ee \rightarrow Z \gamma\gamma$ sensitive to the polar angle cut





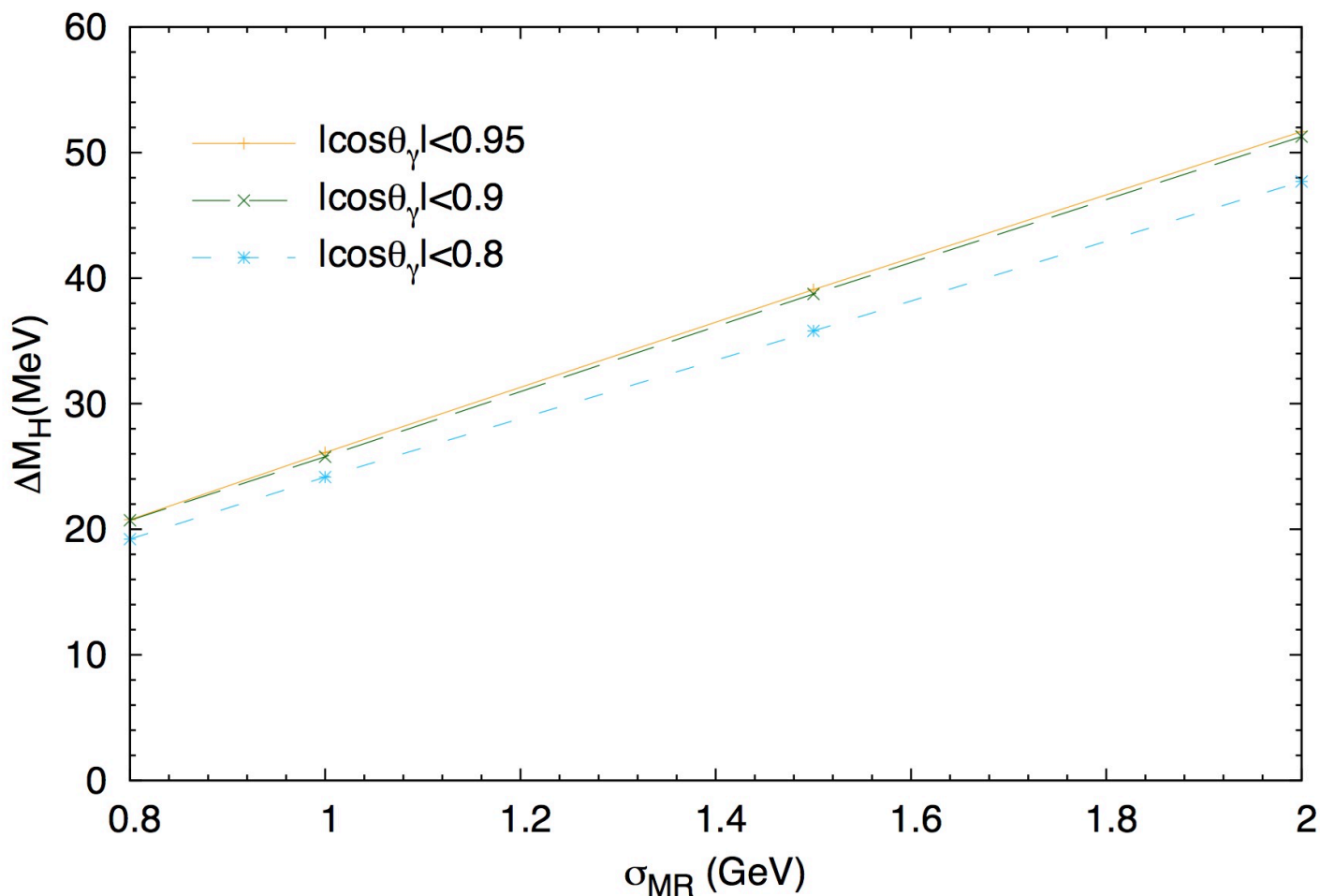


A Gaussian used to model the detector resolution, varying from 0.8-2.0GeV



Sizeable shift observed and increases with resolution

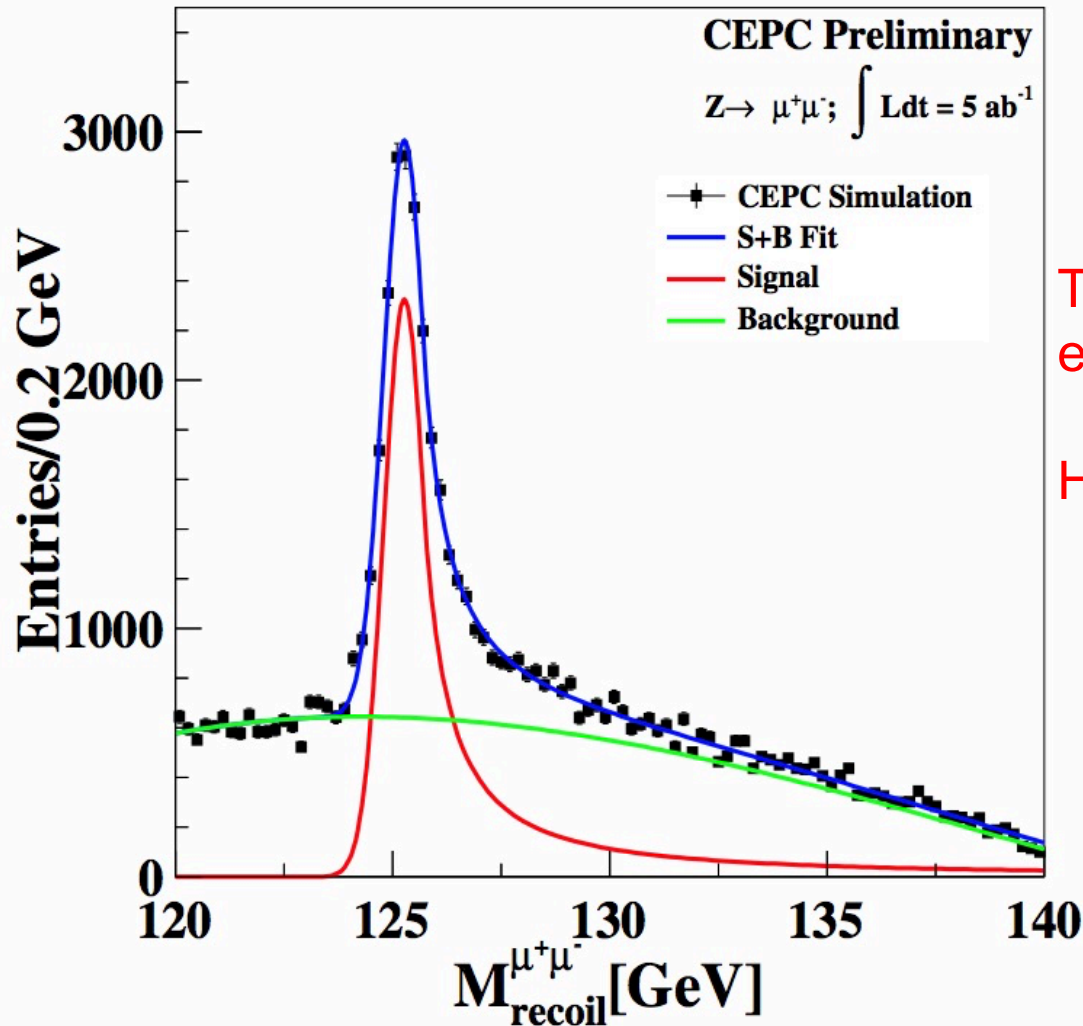
The detector resolution affects the mass shift



Discussion on $H \rightarrow \gamma\gamma$

- $\delta M \sim 50\text{MeV}$ due to interference when $\sigma=2\text{GeV}$
- Bad resolution amplifies the shift
- Polar angle cut reduces the background level, so does the mass shift of Higgs
- The $B(H \rightarrow \gamma\gamma) \sim 0.23\%$, less than 1000 can be selected for the further Higgs mass measurement \rightarrow
 $\delta M = 100\text{--}200\text{MeV}$.
- Interference effect not a urgent issue here. If LX10 or more?
- But in case of the model independent measurement : ee and $\mu\mu$ recoil mass

$\mu\mu$ channel: 22K signals, $\delta M=6.5\text{MeV}$



The potential interference effect not been considered yet .

How?

- Precision mass measurement is important to Higgs study and to new physics
- The interference effect has to be taken into account when M_H is measured with di-photon final state when statistics increase
- The same effect should be evaluated carefully in case of $O(10\text{MeV})$ precision

小结

- 测量过程和结果的关键都在于**修正**：修正和改进已有的知识
- 测量是一个有大量知识输入的高技巧的活动：合作组的协作
- 丰富的经验会对结果价值、过程、及工作量有比较好的预期
- 误差分析是一个迭代过程，在其中会不断加深我们对测量量的认识
- 其中充满了挑战和乐趣

谢谢大家！