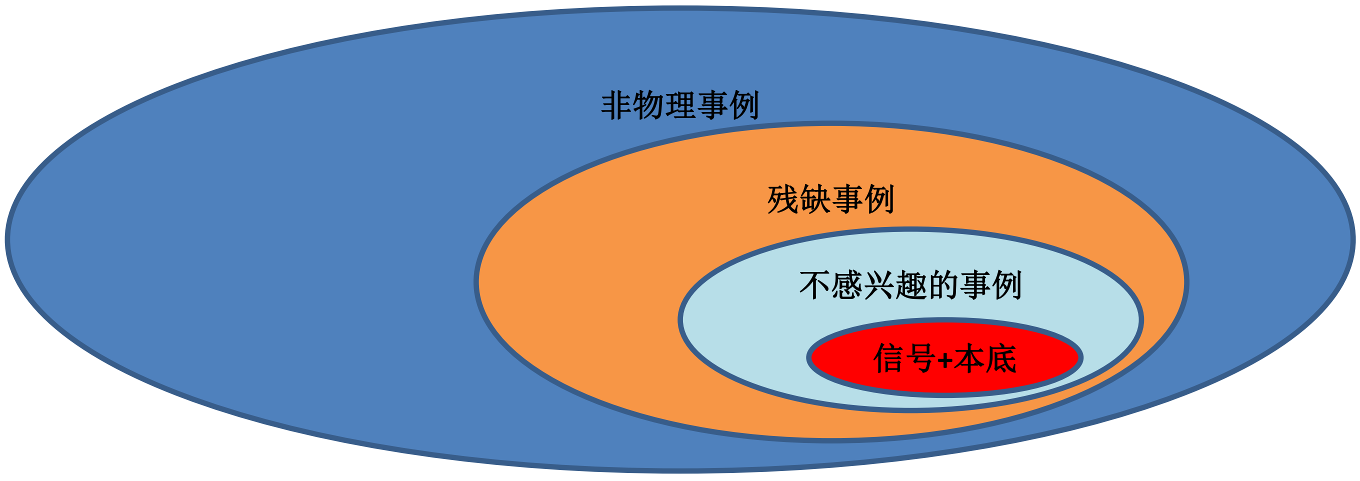
在线事例分类

傅成栋
（fucd@ihep.ac.cn）
实验物理中心
2011/9/17

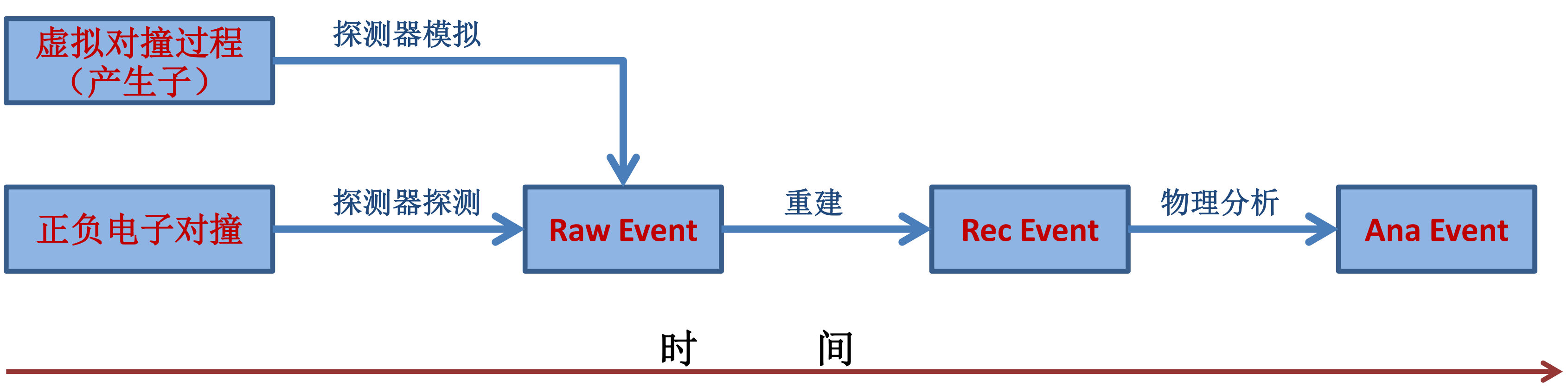
1. 事　　例

- 事例（Event）的定义
  - 百科词条：
    - 成例，可以作为依据的前事。
    - 有代表性的,可以作为例子的事情。
  - 粒子物理：
    - 常常把一次微观的物理反应称为一个事例。
  - 一个事例可以是多个事例的组合:
    - 级联反应：$e^+e^-\rightarrow J/\psi\rightarrow\rho\pi$
      $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad|\rightarrow\pi\pi$
    - 复合事例：两个及两个以上的物理反应几乎同时发生
- 事例样本（Event Sample）

- 事例类型
  - 根据物理反应来分：一个反应道（产生道+衰变道）对应一种事例类型。
  - 根据物理反应末态来分： $e^+e^-\to J/\psi\to e^+e^-$和$e^+e^-\to e^+e^-$是同一个末态事例。
  - 从BES物理分析角度
    - 物理事例：正负电子对撞反应。
    - 非物理事例：被误探测到的电子学噪声、束流相关本底、宇宙线等。
  - 根据事例的可重现性：
    - 好事例
    - 残缺事例
  - 根据每个事例在分析中的作用
    - 信号
    - 噪声/本底



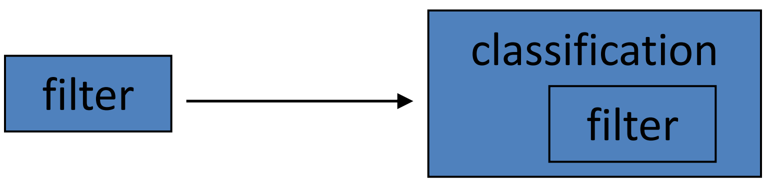
- 事例的表现
  - 模拟事例（MC Event）：
    - 在理论模型（产生子）中，事例表现为一具有确定4动量和顶点的粒子末态。
  - 原始事例（Raw Event）：
    - 在探测器中，事例表现为一系列有特征的数字信号，随着这些信号被记录下来，又成为原始数据（Raw Data）。
  - 径迹重建事例（Rec Event）：
    - 在经过重建之后，事例变成了带电径迹和中性径迹的组合。
  - 过程重建事例（Ana Event）：
    - 经过分析之后确定径迹的粒子种类，从而确定事例的末态，并进一步确定事例的反应过程，实现物理反应的重现。



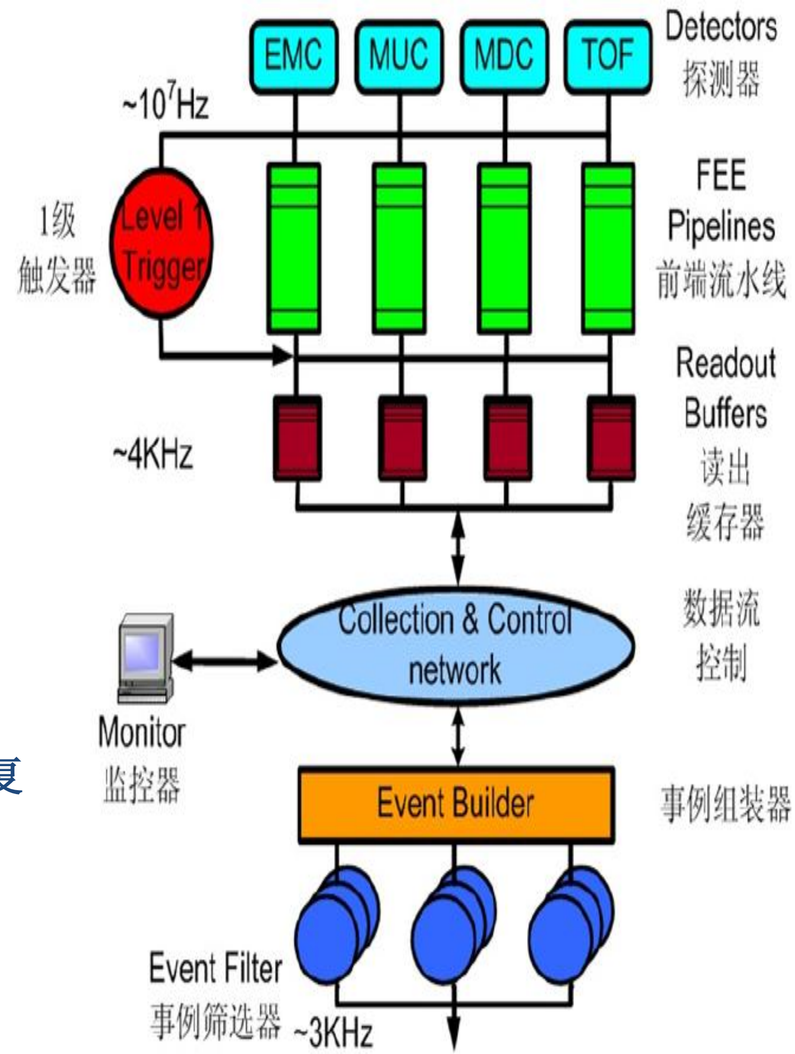
2. 在线事例分类

- 事例分类（Event Classification）
  - 利用某一事例在各个阶段的事例表现确定该事例的事例类型。
  - 研究的对象可以是RawEvent、RecEvent和AnaEvent。
- 在线事例分类
  - 在取数过程中原始数据记带之前**实时**进行的事例分类。
  - 分类的目的是利用分类结果计算亮度、快速刻度等等。

- 由来：
  - 作为**软件触发器**（Online Event Filter）
    - 降低存储的事例率
      4000Hz ⟶ 3000Hz
    - 丢弃噪声
    - 鉴别本底
  - 增加附加需求：快速刻度样本、在线亮度计算等
    - 鉴别几种典型的物理事例



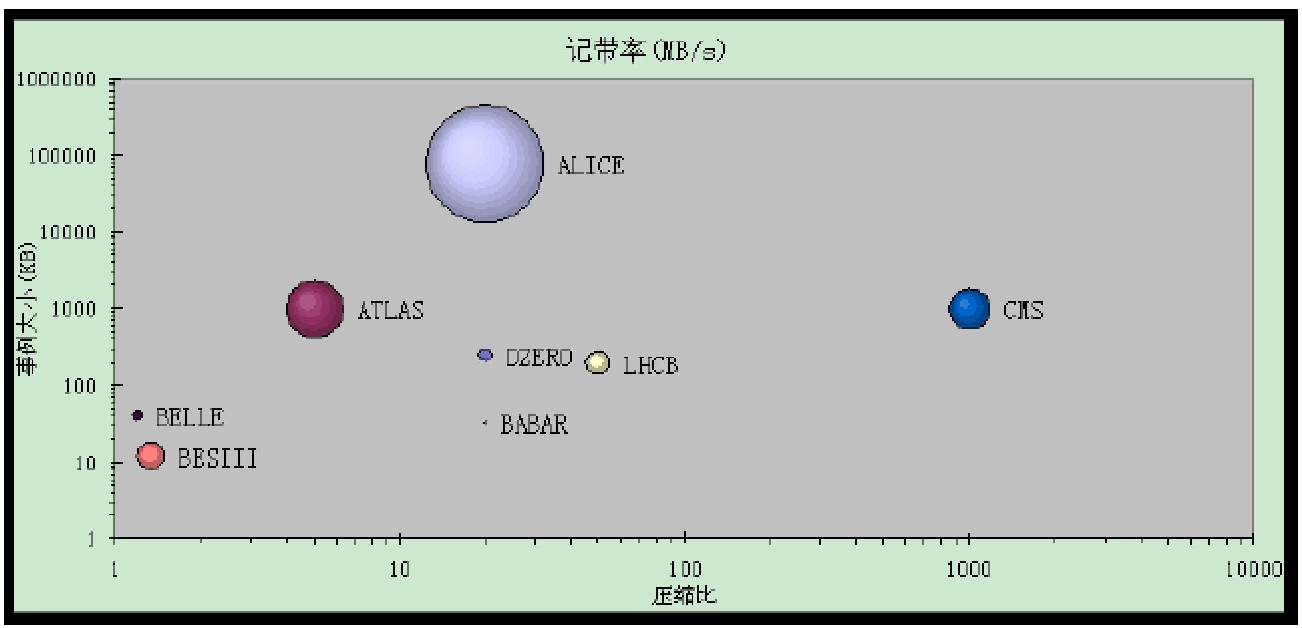
BESIII 在线事例分类系统已经成为一个较为复杂的有着众多功能和应用的子系统。



- 软件级算法（Algorithm）特点：
  - 可采用径迹级的具体算法
  - 算法配置灵活（可控性算法）

	DZERO	ATLAS	LHCb	BABAR	BELLE	CMS	ALICE	BES-III†
输入事例率（Hz）	1000	1000	10k	2000	300	100k	200	4000
输出事例率（Hz）	50	200	200	100	250	100	10	3000
压缩比	20	5	50	20	1.2	1000	20	1.33
事例大小（KB）	250	1000	200	33	40	1000	80k	12
记带率（M/s）	12.5	200	40	2.5	10	100	800	48

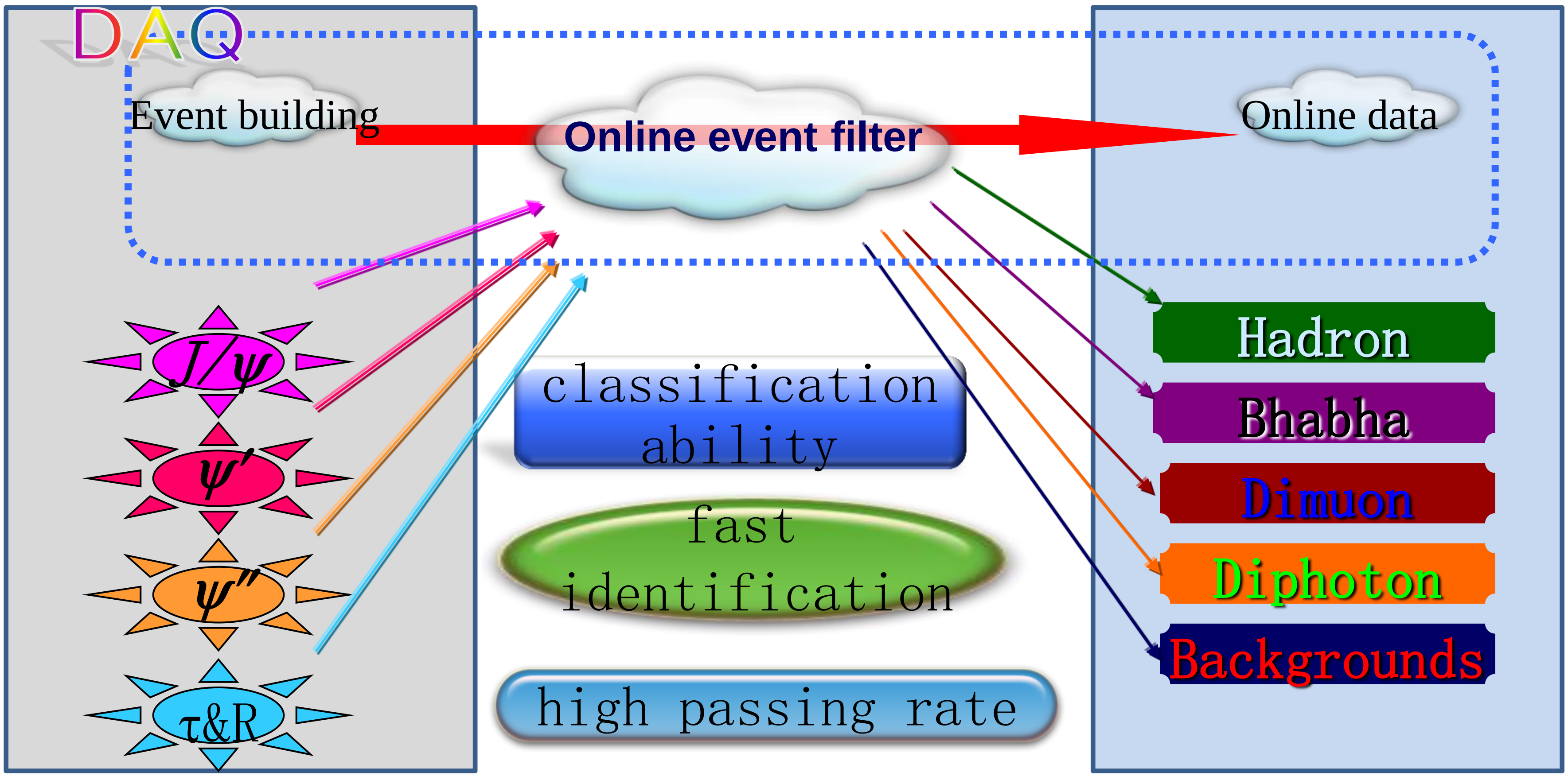
† 表中数据为J/ψ峰取数时的情况，在其它共振峰或者连续区取数时，事例率会减小，压缩比相应提高。



BESIII在线事例分类特点

- B-工厂 & LHC
  - “好”事例少: 10~100Hz
  - 单事例长度（数据记录大小）长
  - 采用大量的在线机群（online farms）, 完整地进行事例重建, 选择感兴趣的物理事例记录
- BES-III
  - “好”事例比例高： ~2000Hz (J/ψ peak)
  - 单事例长度短: ~8kB
  - **中等规模机群 (20~40), 快速重建（简单）, 利用事例拓补特征分类**

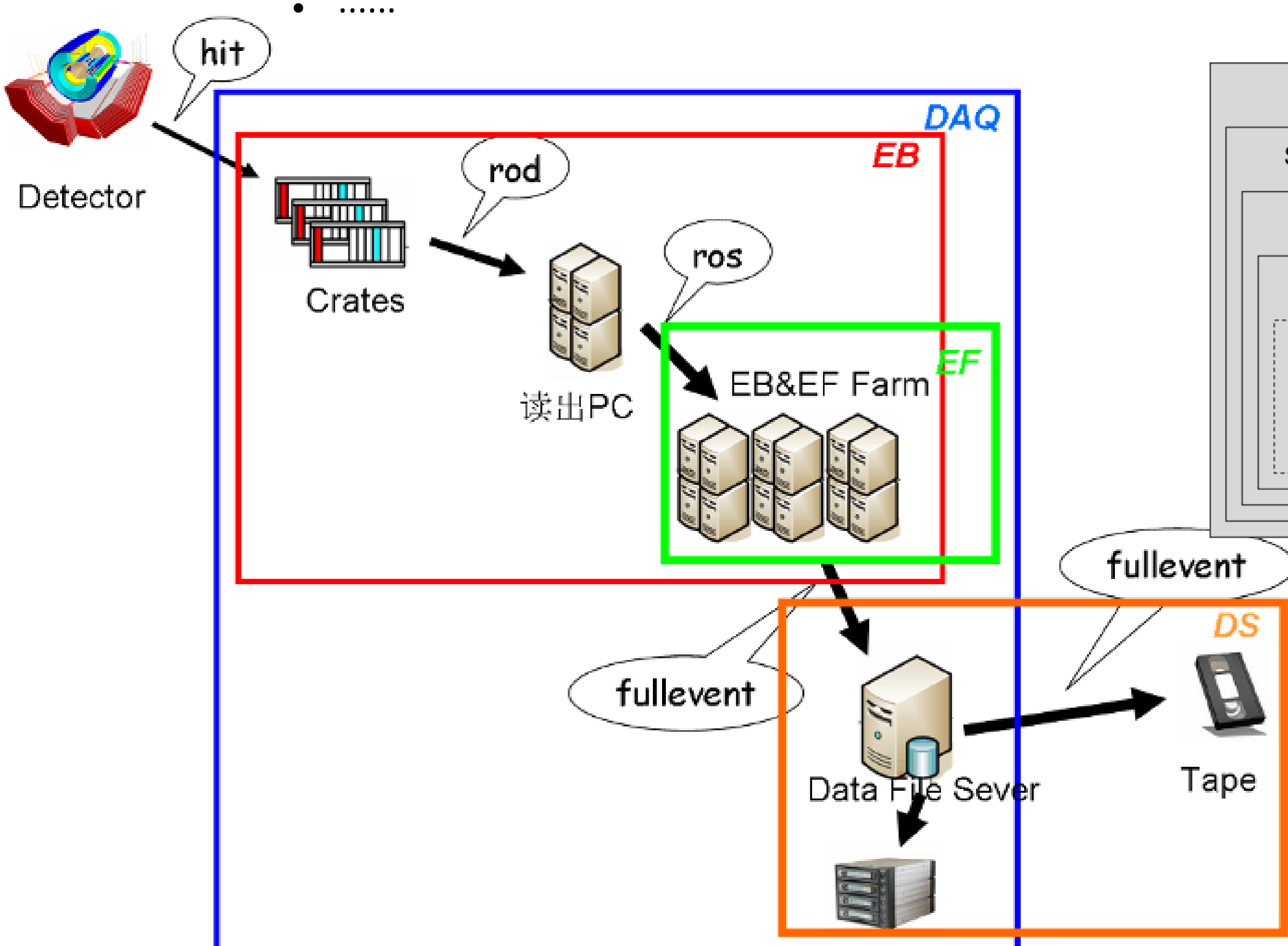
以20台规模的机群为例，在4000Hz事例率时，每个事例的分类速度必须快于5ms。

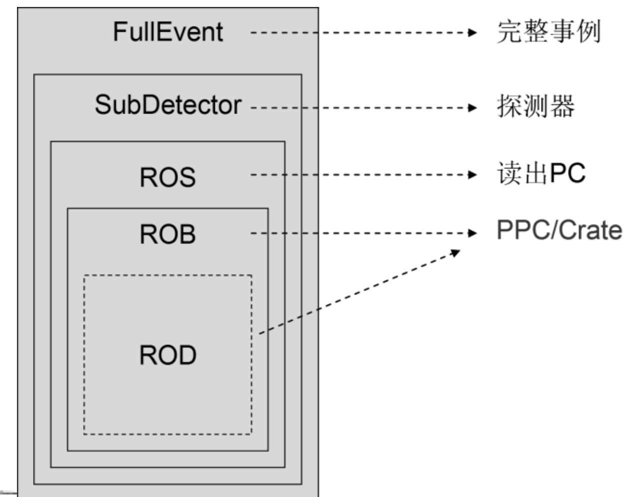


1. 软　　　　　件
2. 算　法　设　计
3. 应　　　　　用

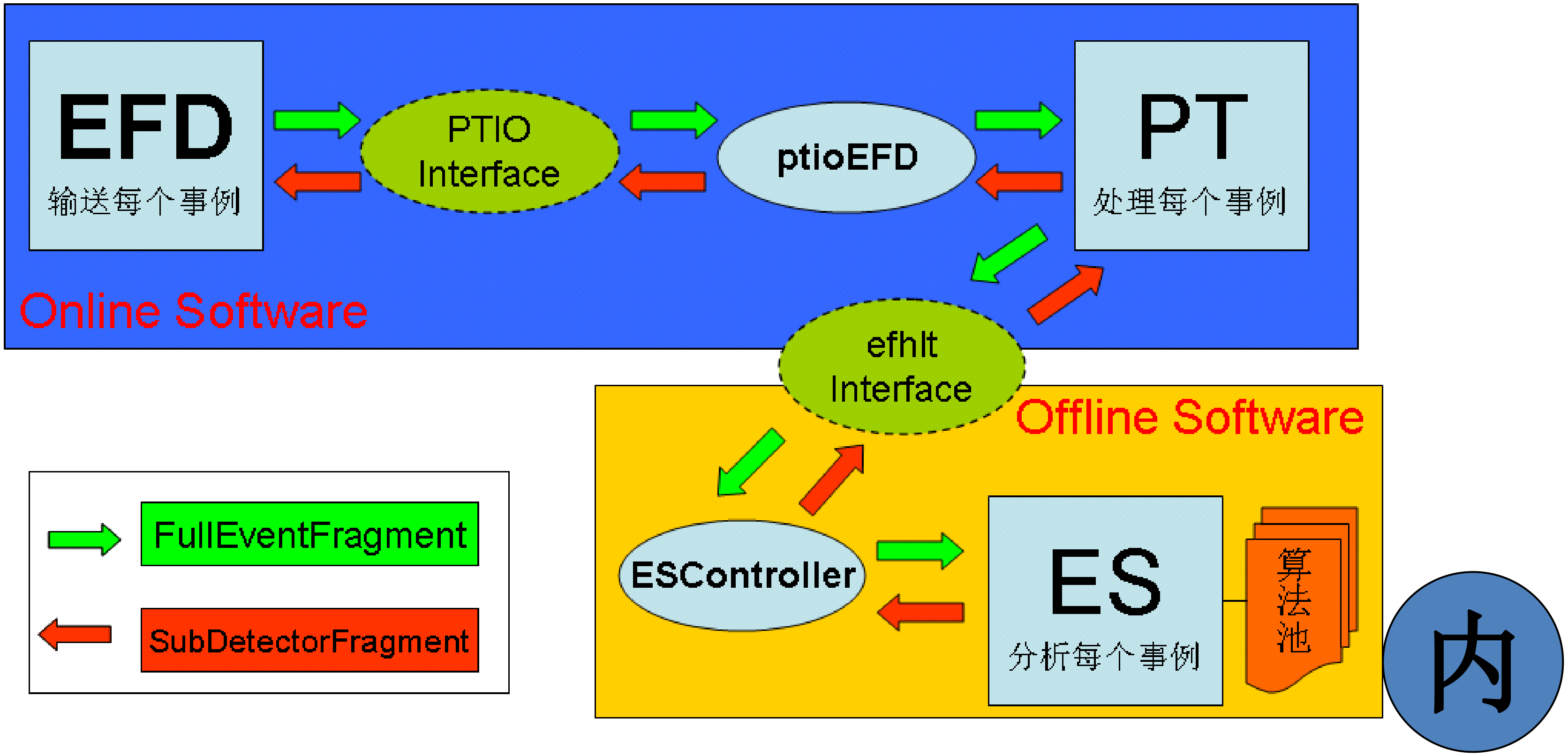
Software: 数据流（Data Flow）

- 在线事例是一段按规则排列的内存区块。
  - 事例头：固定字节
  - 子探测器片段（Fragement）
    - 片段头
    - ……





Software: 前端框架



离线软件通过**efhlt**接口由在线数据获取系统动态载入。
事例通过指针传递给内部算法，利用已有**BOSS**框架进行解析。
不同于离线的是，需要对解析的数据进行预处理（**TQ**匹配、丢弃小**EMC**信号等）以提高速度。

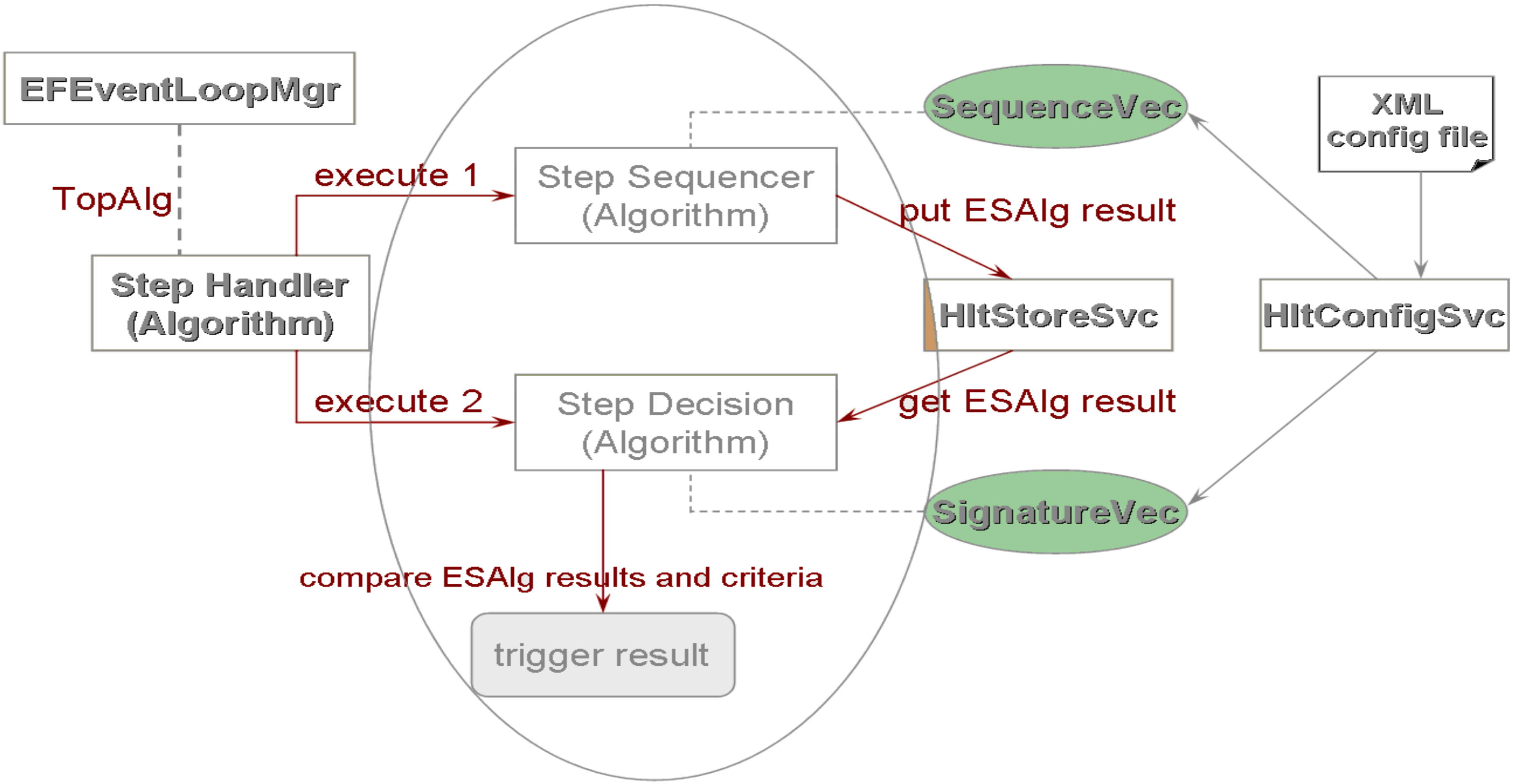
GAUDI初步

- Service
  - 主要用来数据读取、传递和存储
  - 两大基本Service
    - JobOptionsSvc
    - MessageSvc
  - Service池（vector）
- Algorithm
  - Algorithm池：ApplicationMgr.TopAlg
  - SubAlgorithm池：createSubAlgorithm**(参数)**
  - 独立Algotithm：

Software: 内部框架

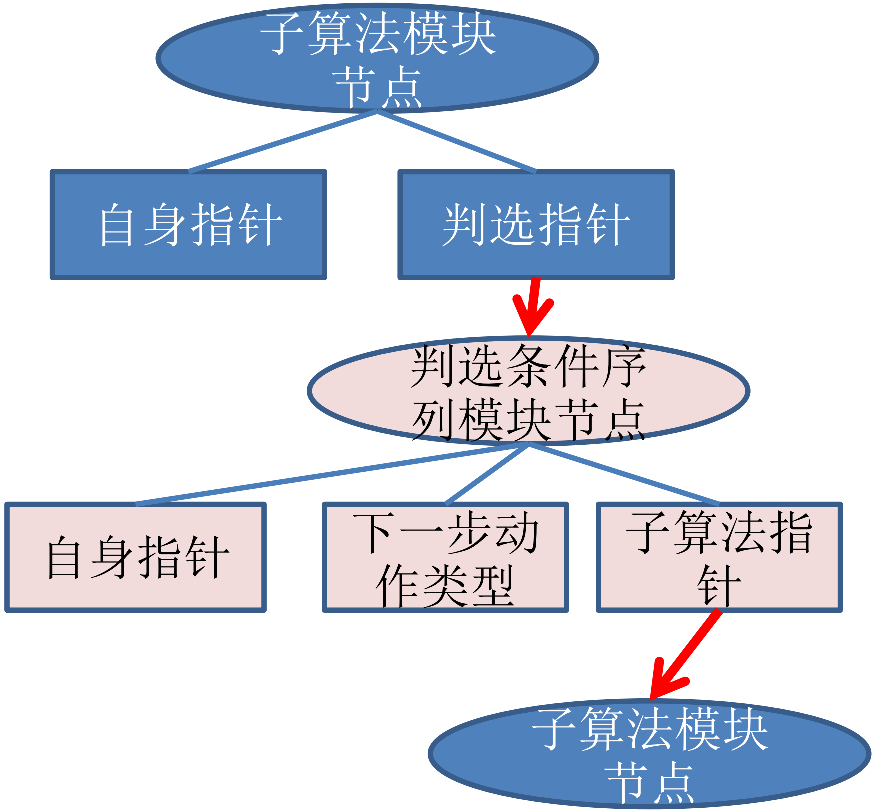
Gaudi based!

BOSS based!

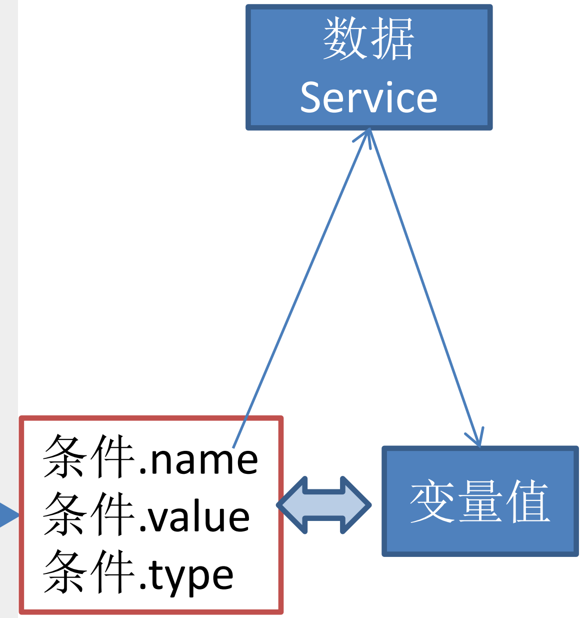


Software: 算法组织

- 子算法化: 每个子算法负责专门的物理变量计算
  - EFDetectorHits
  - EFGlobalEnergy
  - EFProcessCluster
  - EFSectorHits
  - EFChargedTrack
  - EFFlightTime
  - EmcRec
  - MdcFastTrk
- 模块化:
  - HIT: EFDetectorHits
  - EMC: EFGlobalEnergy
  - CST: EmcRec, EFProcessCluster
  - SEC: EFSectorHits
  - MDC: MdcFastTrk, EFChargedTrack
  - TOF: EFFlightTime
- 通过XML 配置文件里Keywords 控制算法:
  - Accept: finish classification for this event
  - Jump: turn to next module
  - Continue: next decision



```xml
<SIGNATURE signature_id="CST" prescale="1" forced_accept="2">
  <CRITERIA behaviour="continue" mark="LUM">
    <ITEM name="eend" value="2.8" type="MIN" />
    <ITEM name="etot" value="6" type="MAX" />
    <ITEM name="acop" value="23" type="MAX" />
    <!--ITEM name="acole" value="40" type="MAX" /-->
    <ITEM name="emax1" value="1.4" type="MIN" />
    <ITEM name="emax2" value="0.8" type="MIN" />
    <!--ITEM name="ebal" value="0.6" type="MAX" /-->
    <ITEM name="coste1" value="0.84" type="MIN" />
    <ITEM name="coste2" value="-0.84" type="MAX" />
  </CRITERIA>
  <CRITERIA behaviour="continue" mark="LUM">
    <ITEM name="eend" value="2.8" type="MIN" />
    <ITEM name="etot" value="6" type="MAX" />
    <ITEM name="acop" value="23" type="MAX" />
    <!--ITEM name="acole" value="40" type="MAX" /-->
    <ITEM name="emax1" value="1.4" type="MIN" />
    <ITEM name="emax2" value="0.8" type="MIN" />
    <!--ITEM name="ebal" value="0.6" type="MAX" /-->
    <ITEM name="coste1" value="-0.84" type="MAX" />
    <ITEM name="coste2" value="0.84" type="MIN" />
  </CRITERIA>
  <CRITERIA behaviour="jump" next="SEC">
    <ITEM name="acop" value="23" type="MAX" />
    <ITEM name="emax1" value="1.4" type="MIN" />
    <ITEM name="emax2" value="0.8" type="MIN" />
    <!--ITEM name="ebal" value="0.7" type="MAX" /-->
  </CRITERIA>
  <CRITERIA behaviour="accept" mark="Hadron">
    <ITEM name="etot" value="5" type="MAX" />
  </CRITERIA>
  <CRITERIA behaviour="accept" mark="Error">
  </CRITERIA>
</SIGNATURE>
```



Software: 数据格式

- **事例头**
  - **Event type**

0	…	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
		双光子		强子		γ对		μ对		B巴巴		E巴巴		宇宙线		束流		满天星	

  - **Algorithm process**

0	…	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
				辅助事例类型										子算法执行标记					

  - **Version**
  - **Total energy**

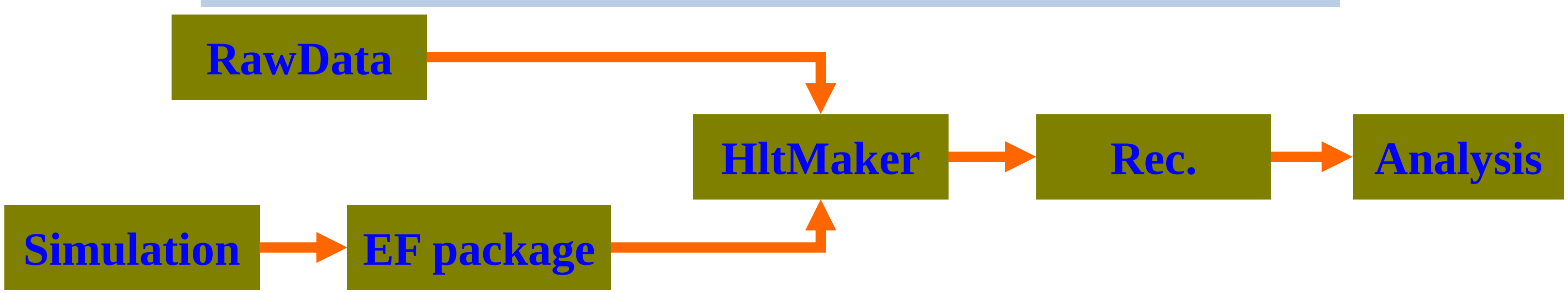
- **Sub-detector fragment**
  - **Variables +“true”ID**

1 0 3 9	1 0 0 9	1 7 0 0	1 1 1 1	0 0 0 0
Nmdc	Nmdc	Nemc	Nmdc	
Ntrk	Ntrk	E1	Ntof	
P1	P1	E2	Nemc	
P2	P2	acop	Nmuc	
Acol	Acol	Ebalance		
cosθ1	cosθ1	Ebarrel		
cosθ2	cosθ2	Eendcap		
vz1	vz1	Nmuc		
vz2	vz2			
Ntof	Nmdc			
ΔT				
Dphi				
Nmdc				

Software: Offline Module

- Event model:
  - /Event/Hlt/HltRawCol
  - /Event/Hlt/HltInf
  - /Event/Hlt/DstHltInf
- Data convertor
  - RawDataCnv: buffer $\rightarrow$ HltRawCol
  - RootCnvSvc: RootEventData
- Offline process:
  - simulation: HltRaw
  - TDS's rawdata: HltRawCol
  - HltMaker: HltRawCol $\rightarrow$ HltInf
  - User interface: HltInf

```
例子：(参考/Event/HltEvent)
SmartDataPtr<HltInf> hlt(eventSvc(),"/Event/Hlt/HltInf");
string      name=hlt->getEventName();
uint32_t type =hlt->getEventType();
```



Software: Simulation

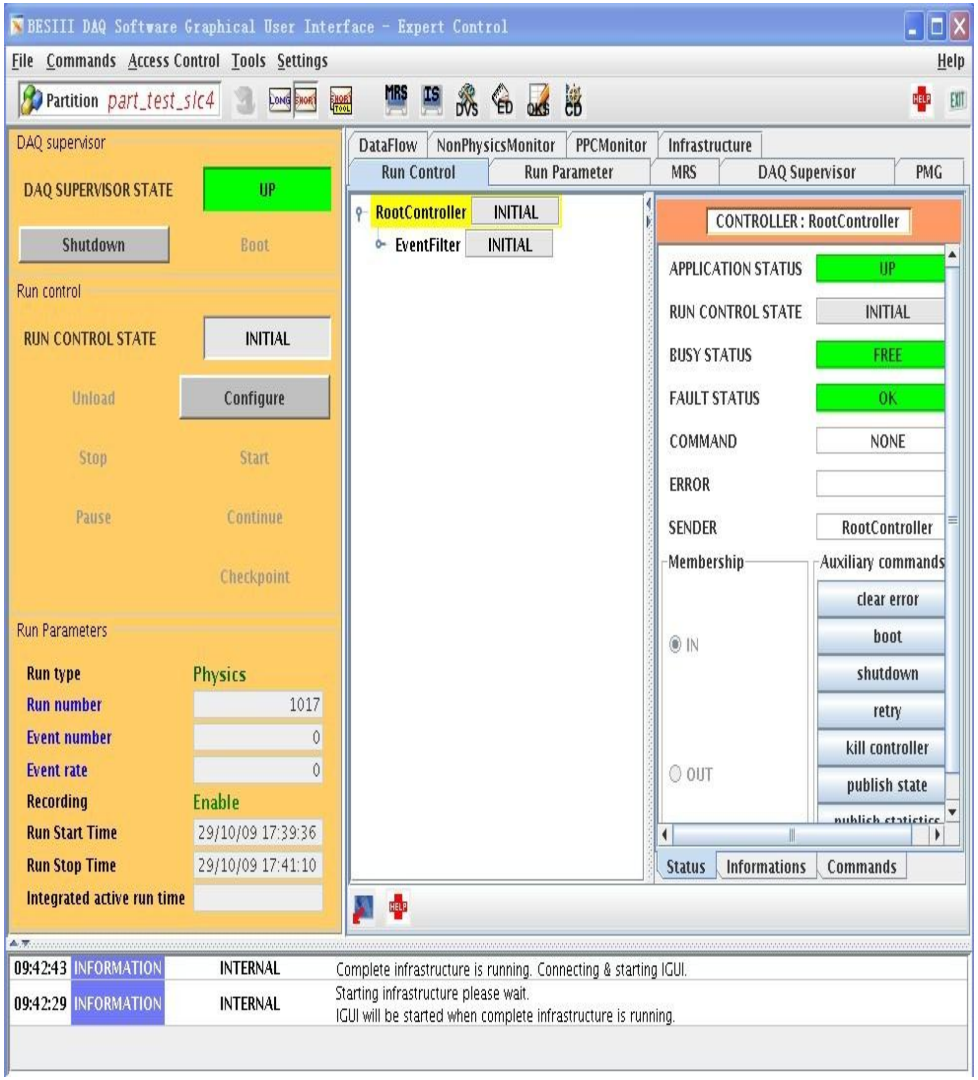
- 算法加载
  - #include “$EVENTFILTERROOT/share/jobOptions_EventFilter.txt”

- 算法配置
  - XML文件
    - 算法模块配置(xml+dtd)
    - 算法判选配置(xml+dtd)
  - 数据库
    - RUN号→efconfig_id (RunInfo/RunParams)
    - efconfig_id→4个EFfileId (configdb/EFConfInfo)
    - EFfileId→配置内容(configdb/EFfile)

HltConfigSvc.FromDB = true;

小结

- 利用现有软件改进
  - GAUDI
  - BOSS
- 围绕在线实时系统要求（快速、稳定）设计软件框架
  - 模块化
  - 可配置
- 同时兼顾在线和离线应用

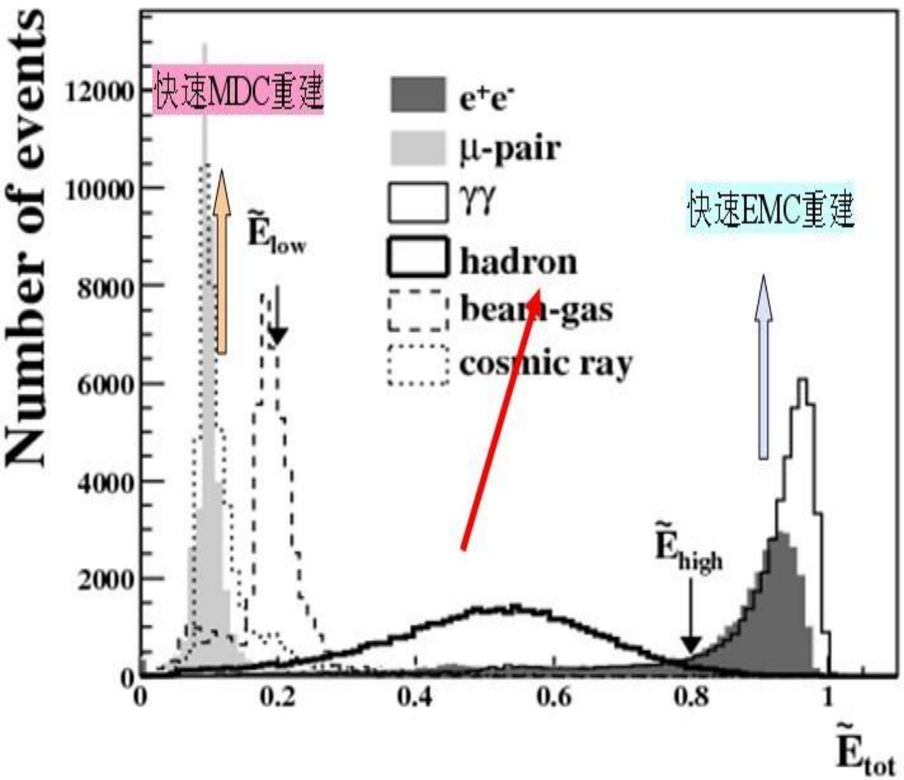


Algorithm

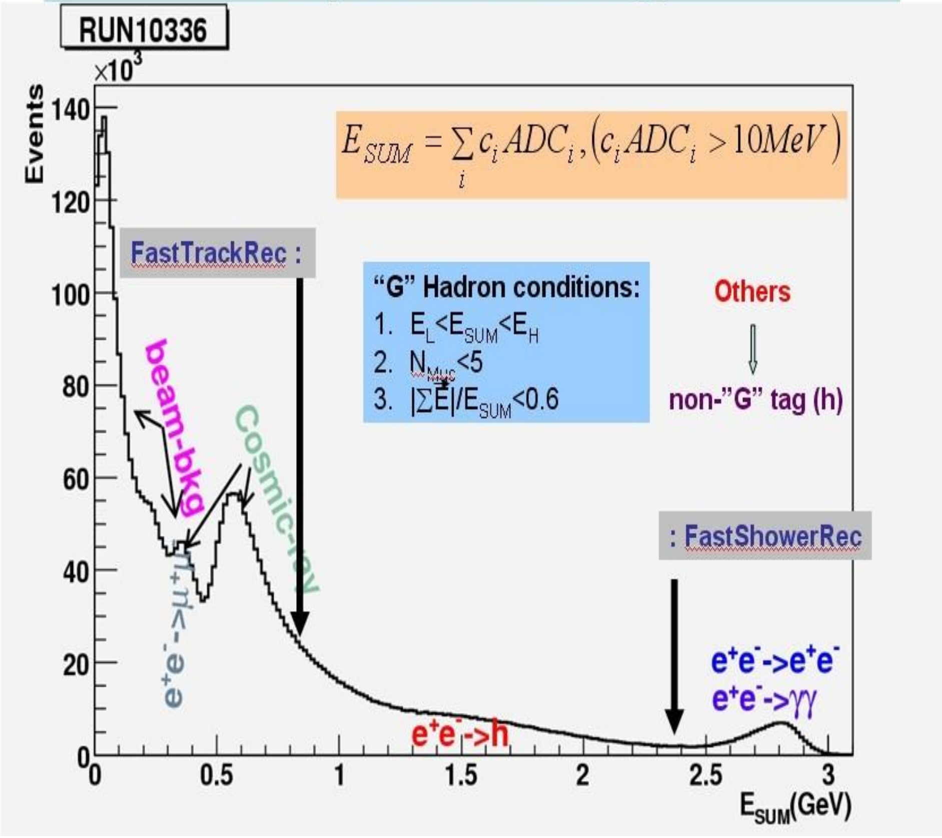
- 选择算法的对象为Digi和快速重建的MDC径迹和EMC簇射
  - <5ms/事例的处理速度要求
  - 快速量能器重建： **~1.5ms/cluster**
  - 快速带电径迹重建：**0.5~1ms/track**
- 方法：不同类型事例的拓补特征
  - Junk:
    - EMC fired fully
    - MDC fired too many
  - 亮度计算的特殊事例
    - 只用量能器信息 (no HV)
    - 不区分Bhabha和Diphoton末态：普适于无磁场情况
    - 排除束流相关本底为第一要素
- 目标：
  - Bhabha: barrel & endcap
  - Diphoton
  - Dimu
  - Hadron

Hadron快速鉴别

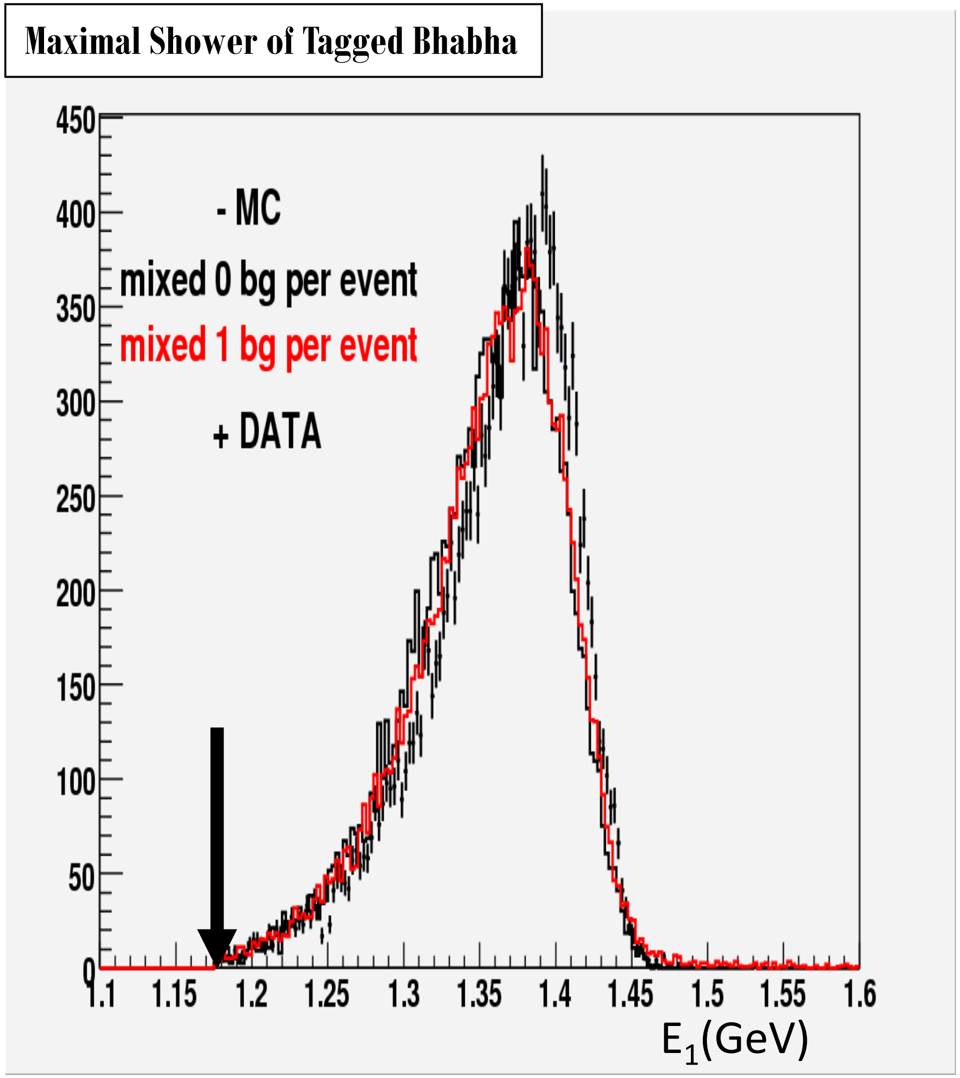
MC

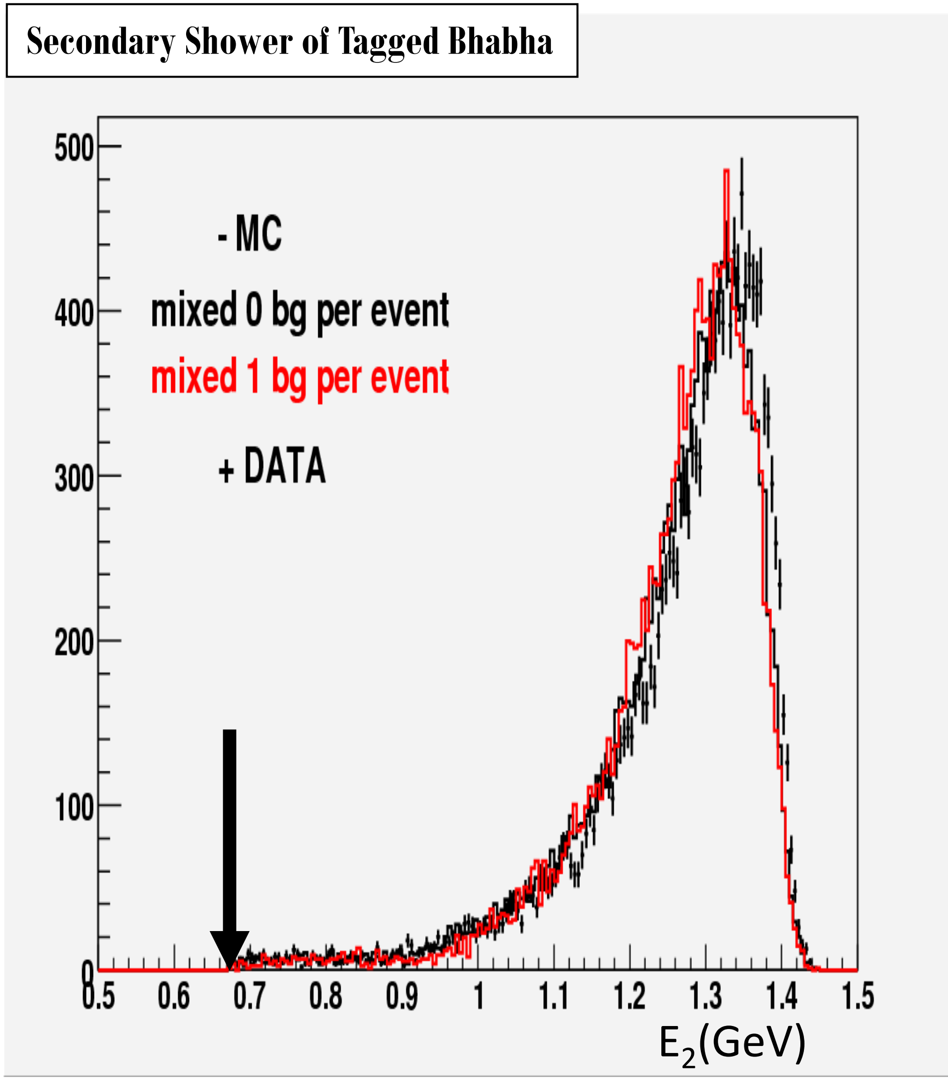


Deposited Energy

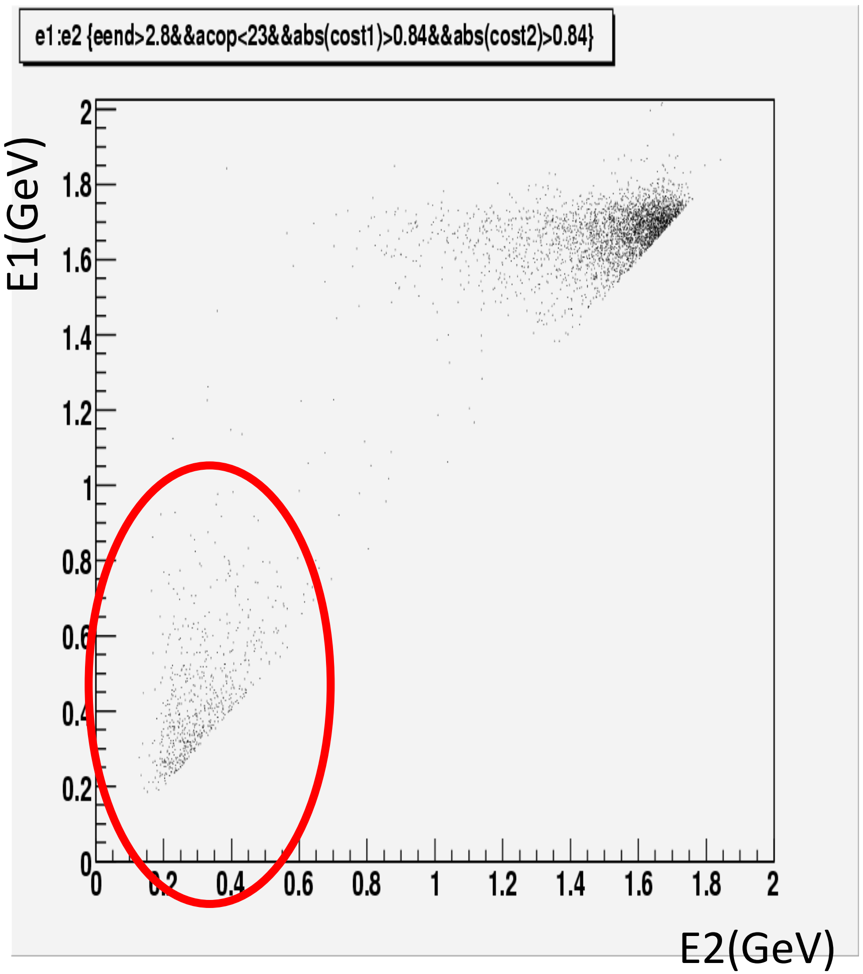


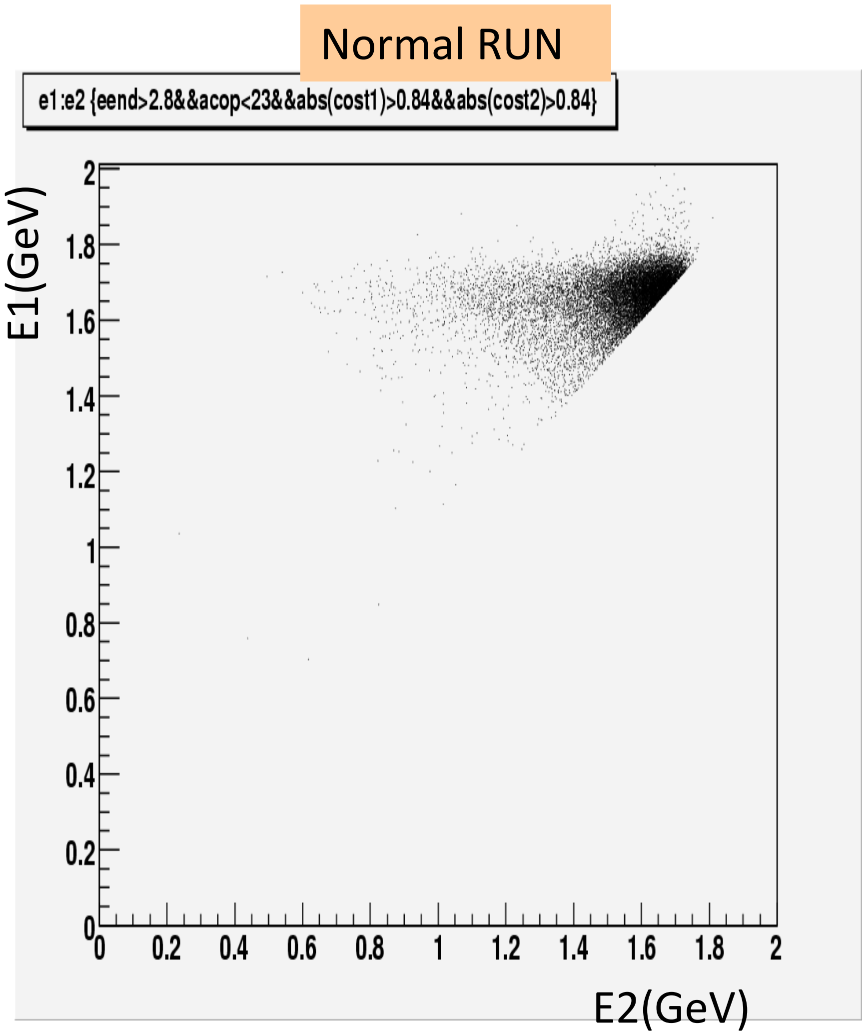
Shower Energy for $e^+e^- \rightarrow e^+e^-, \gamma\gamma$





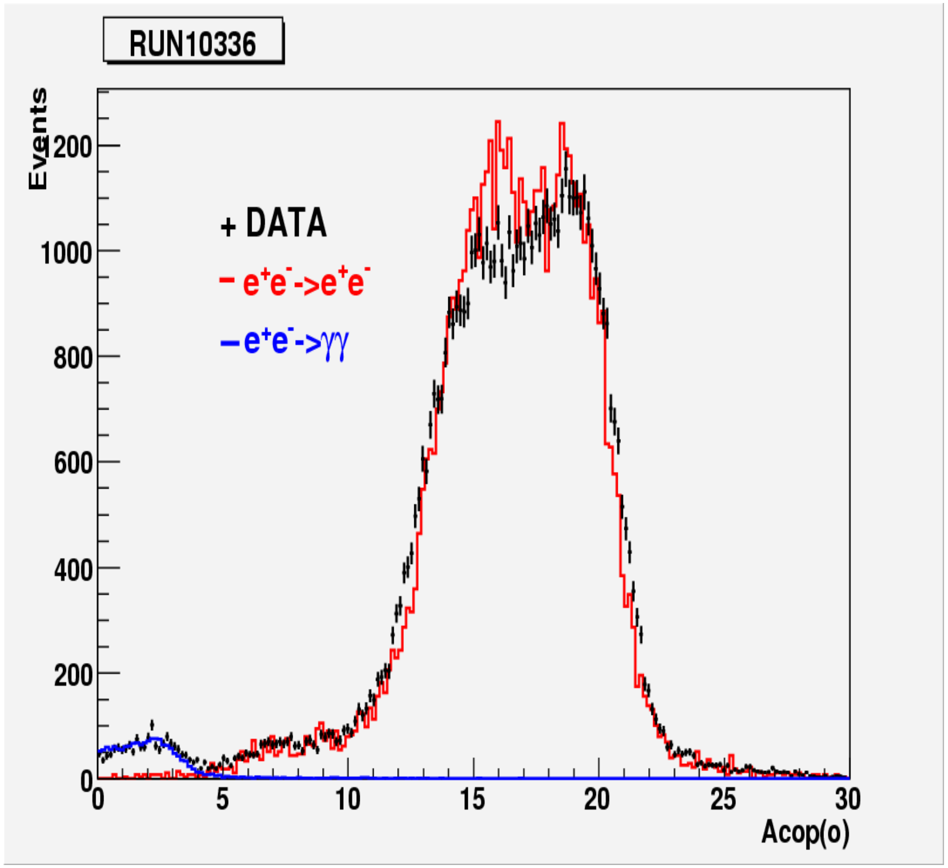
Beam-lost Rejection

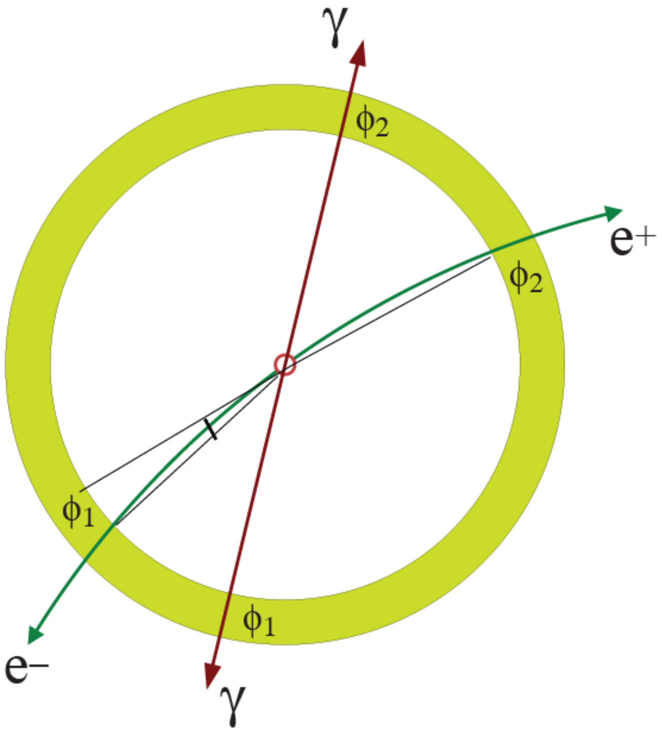




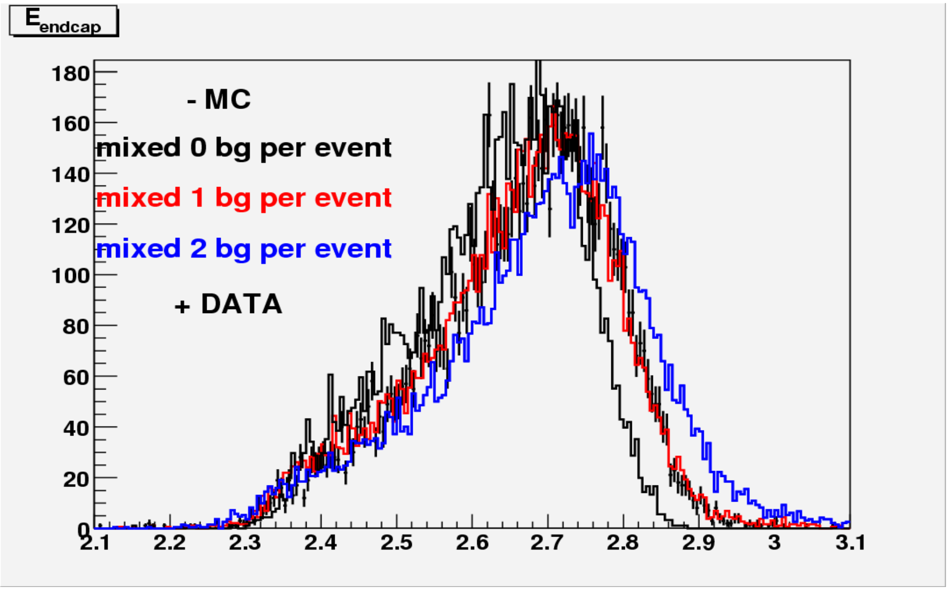
ID for $e^+e^-/\gamma\gamma$: Acoplanarity Angle

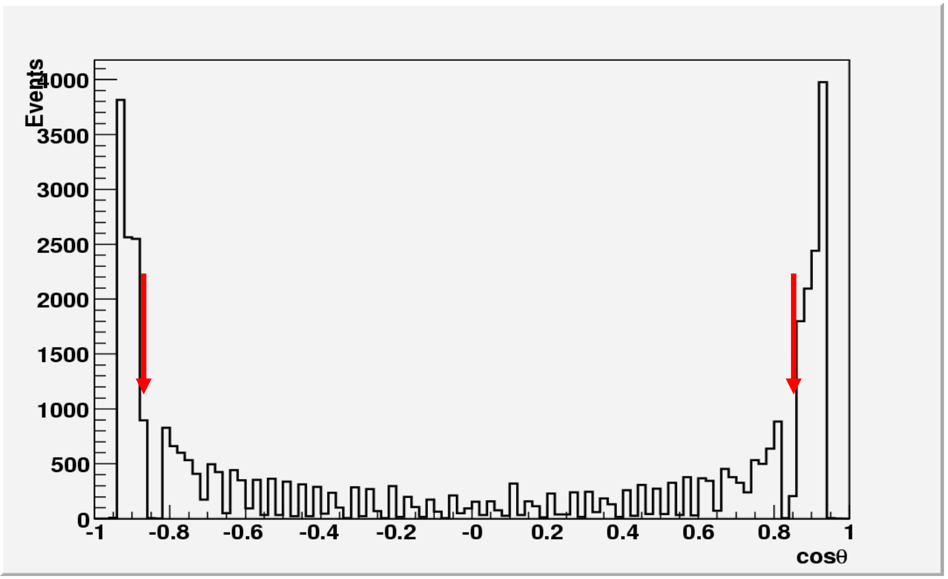
$$\cos(acop) = -\cos\phi_1 \cos\phi_2 - \sin\phi_1 \sin\phi_2$$





Selection for LUM Calculation





- $E_{end}>E_h$
- $E_1>E_{1h}$ & $E_2>E_{2h}$
- $Acop<\alpha$
- $Etot<2.6Eb$
- $|\cos\theta|>0.84$

➢ **High statistics: >1000/min**

➢ **Low resonance bkg:**

Misidentified ratio of J/ψ:

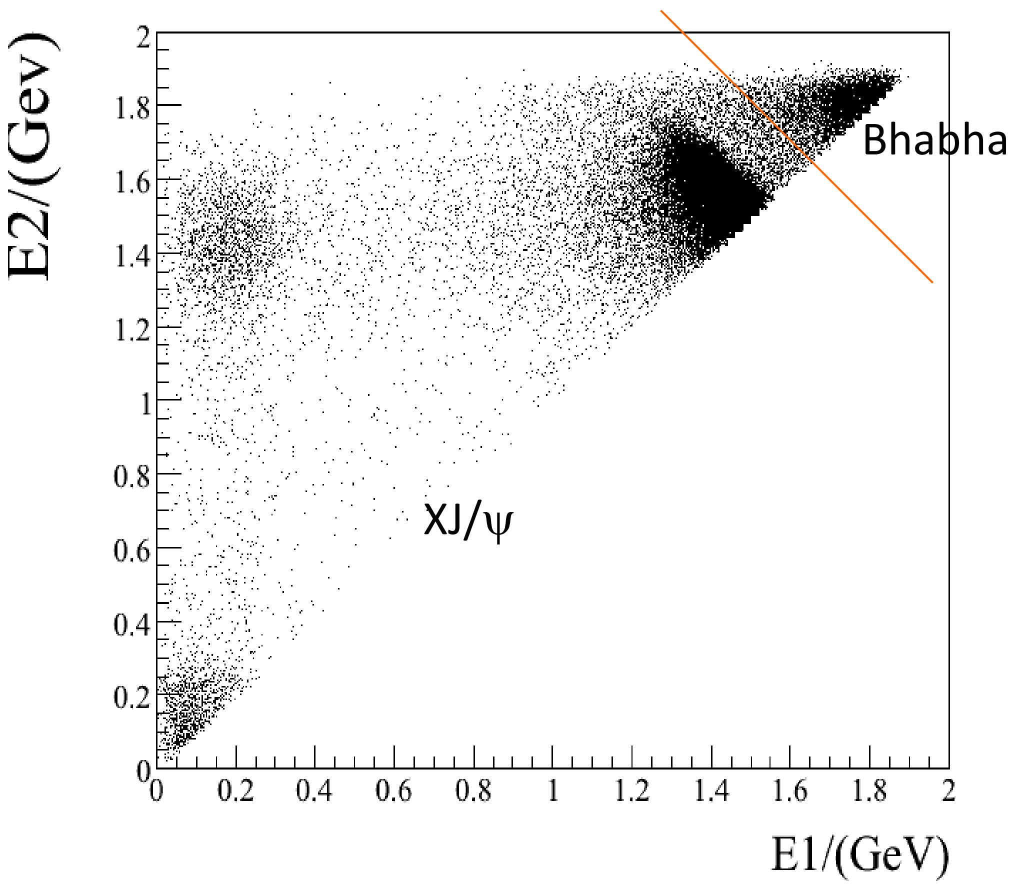
Barrel	**~4%**
Endcap	**~0.5%**

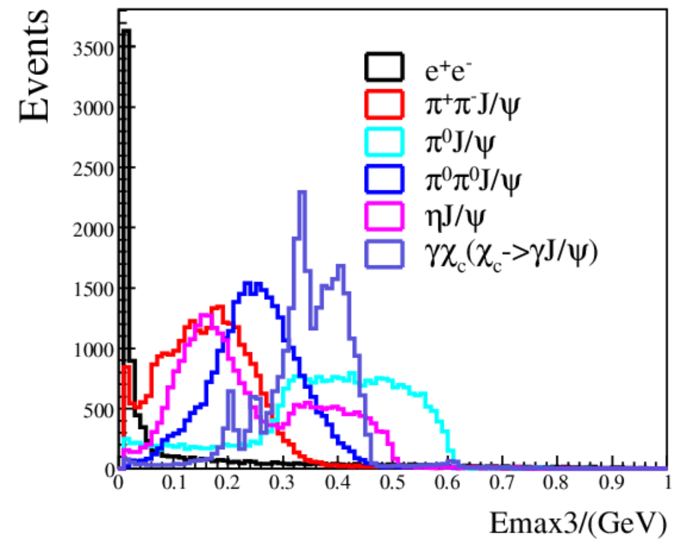
$$N_{lum}=N_{end}-\sum \varepsilon_i N_i$$

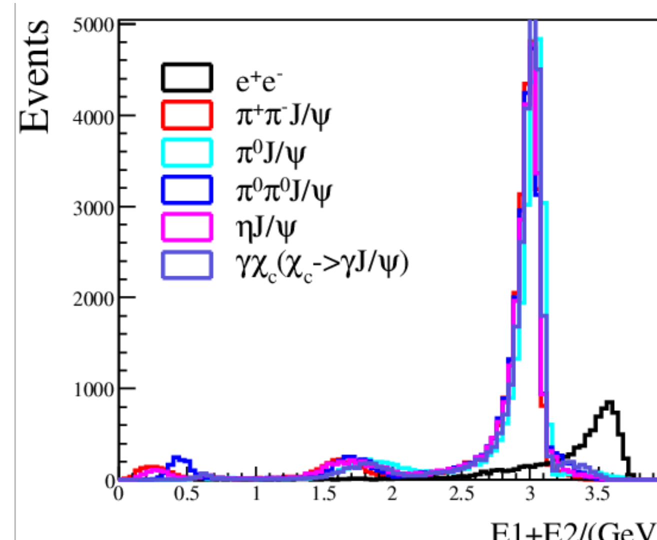
$$\int Ldt=N_{lum}/(\varepsilon_{ee}\sigma_{ee}+\varepsilon_{\gamma\gamma}\sigma_{\gamma\gamma})$$

Bhabha VS XJ/ψ

Bhabha events which are pre-scaled and stored separately must be selected strictly.

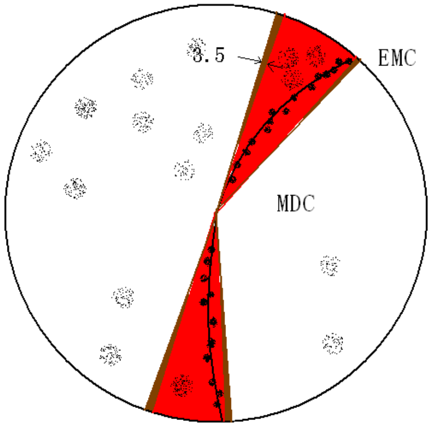


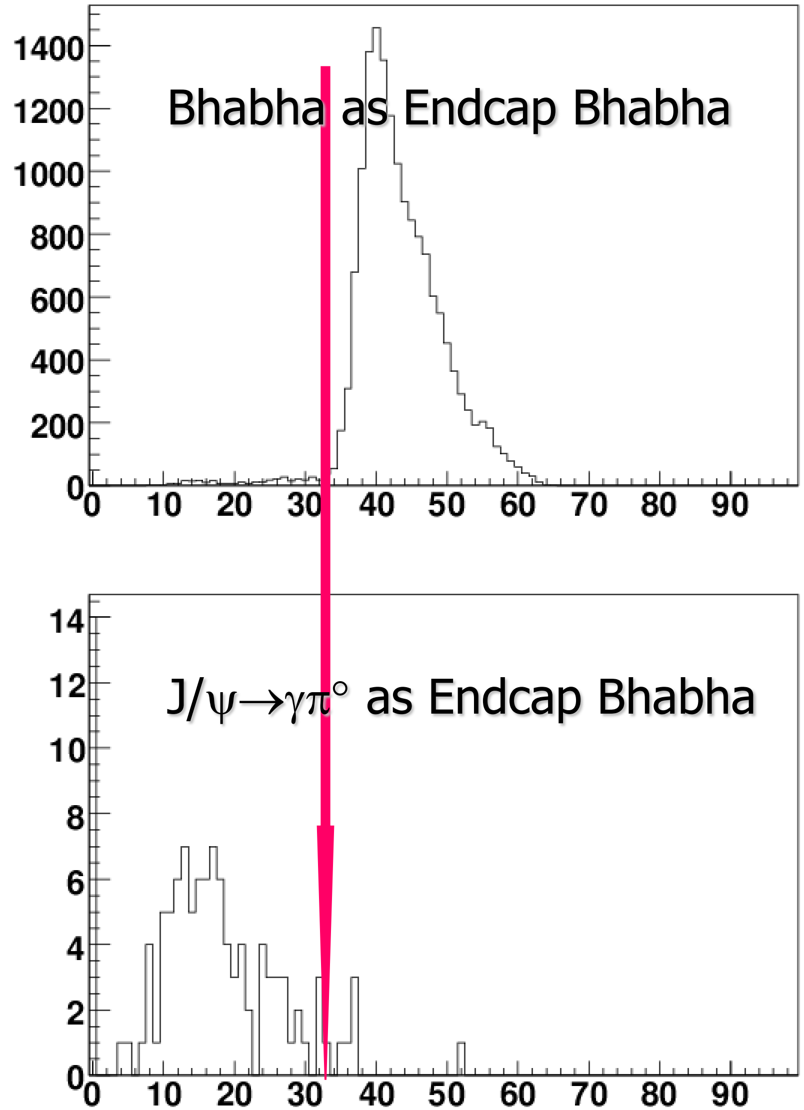




Bhabha VS Neutral Channels

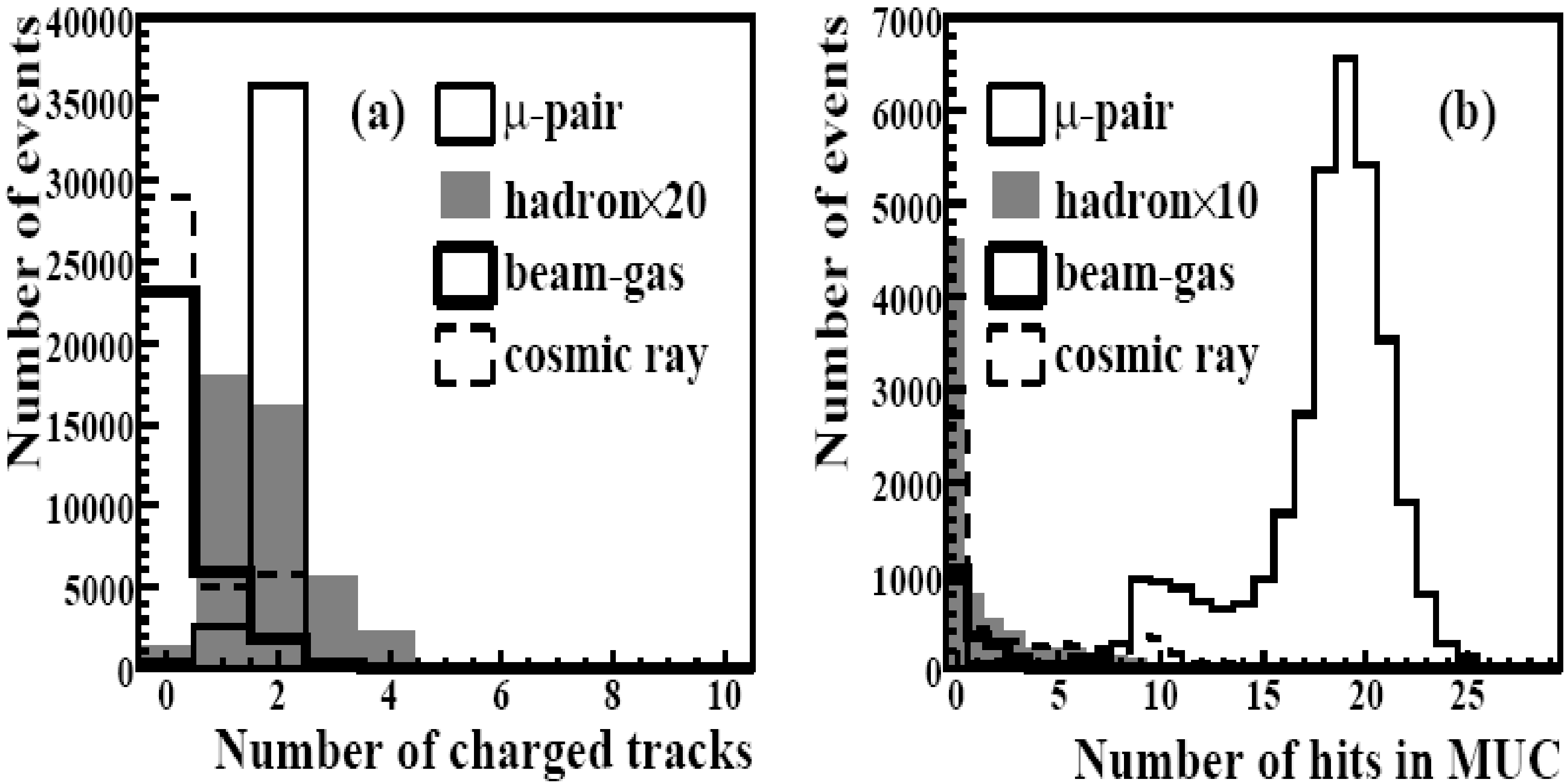
- **Hits number in sector (nsec)**
- Avoiding MDC track reconstruction to save time, at the same time, keeping low mis-identified ratio.





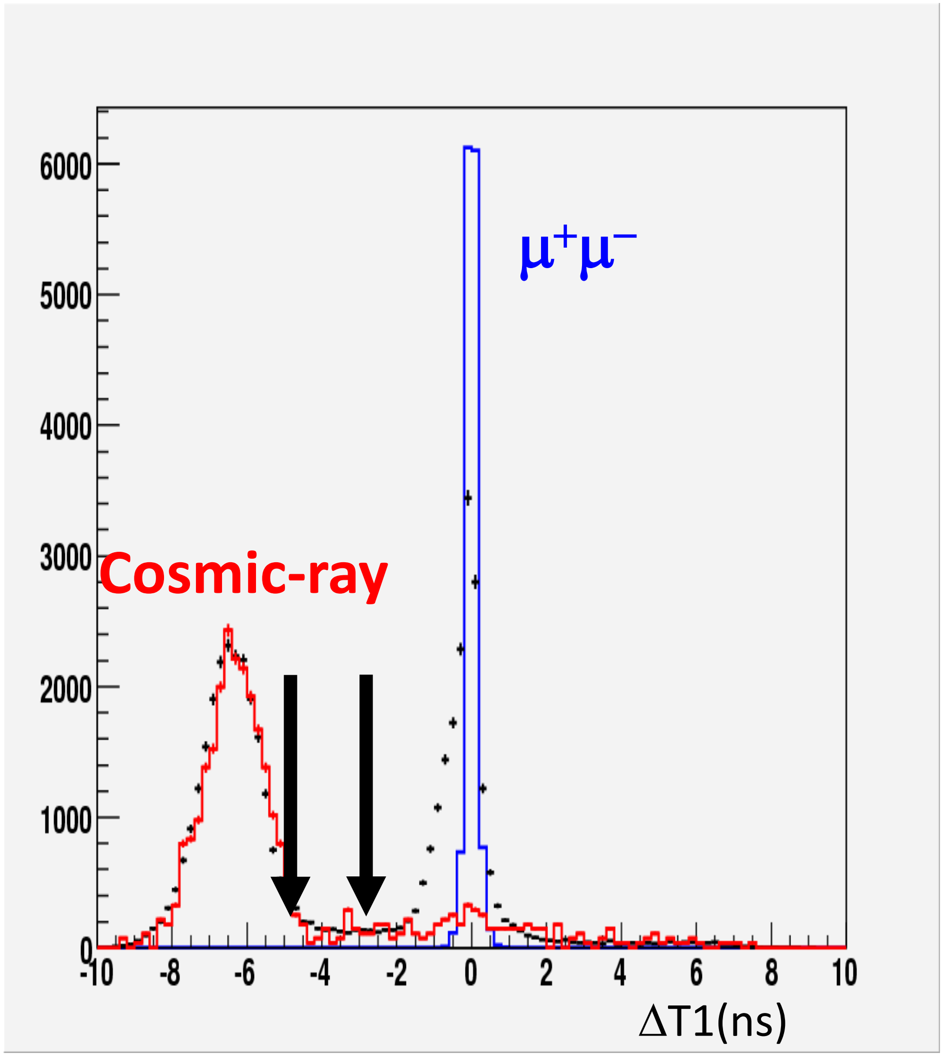
μ$^+$μ$^-$ VS Cosmic Rays, Beam backgrounds

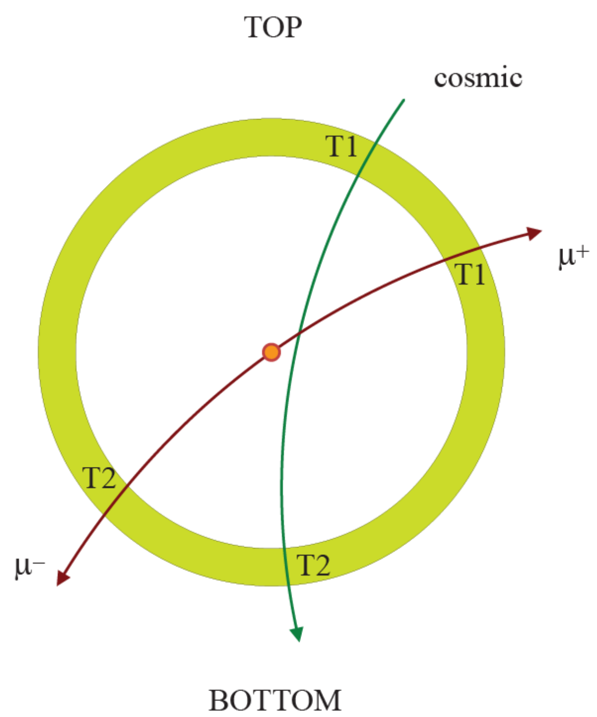
- Number of charged track
- Number of MUC hits



$\mu^+\mu^-$ VS Cosmic ray

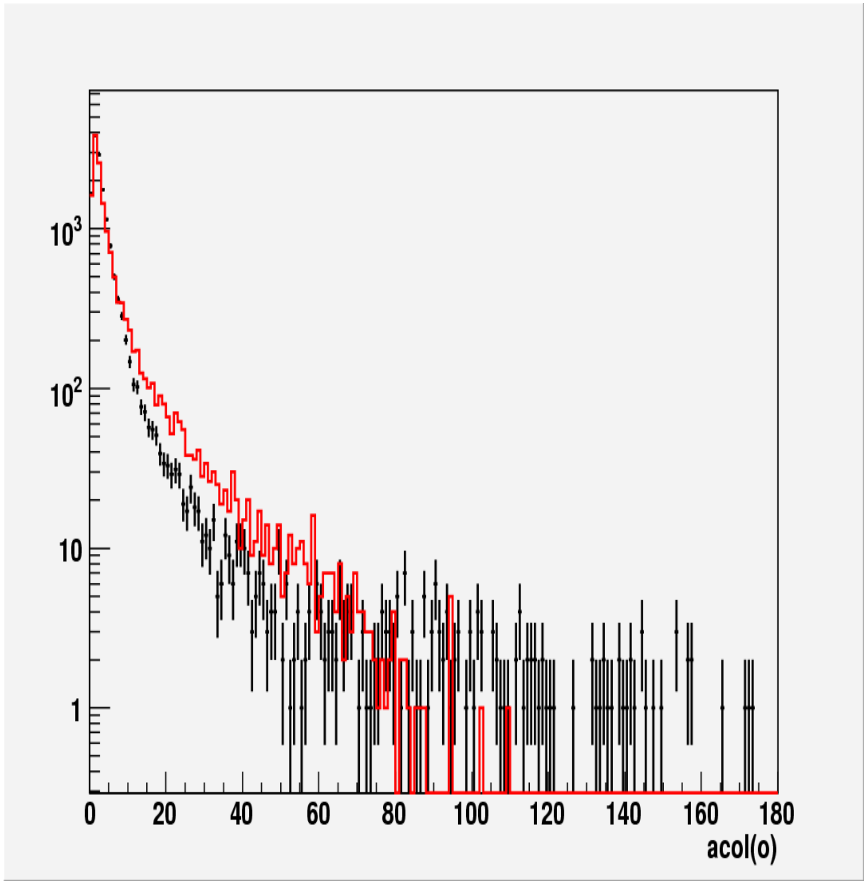
$$\Delta T = t^{top} - t^{bottom}$$





$\mu^+\mu^-$ VS Cosmic Ray

$$\cos(acol) = -\frac{\vec{p}_1 \cdot \vec{p}_2}{p_1 p_2}$$

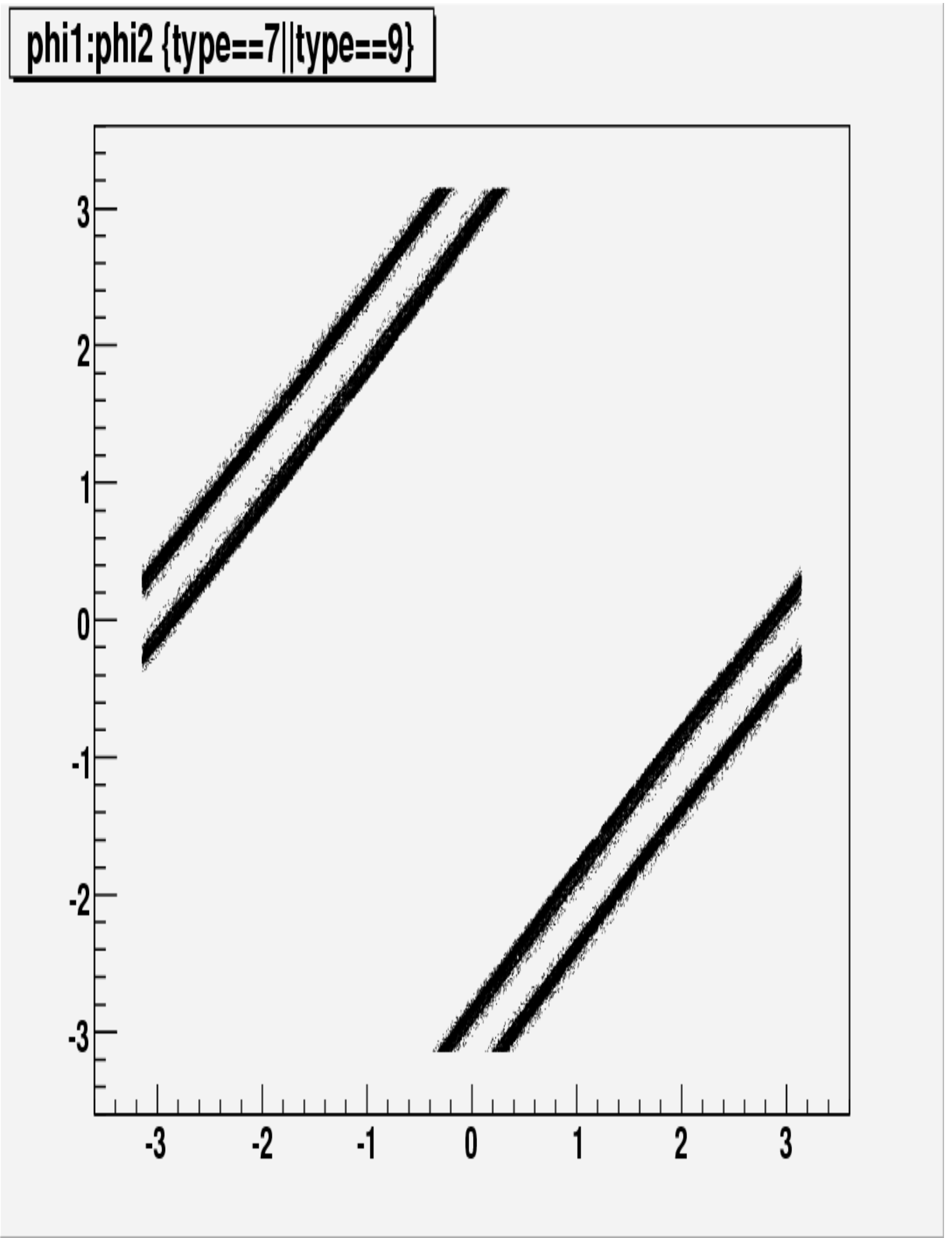


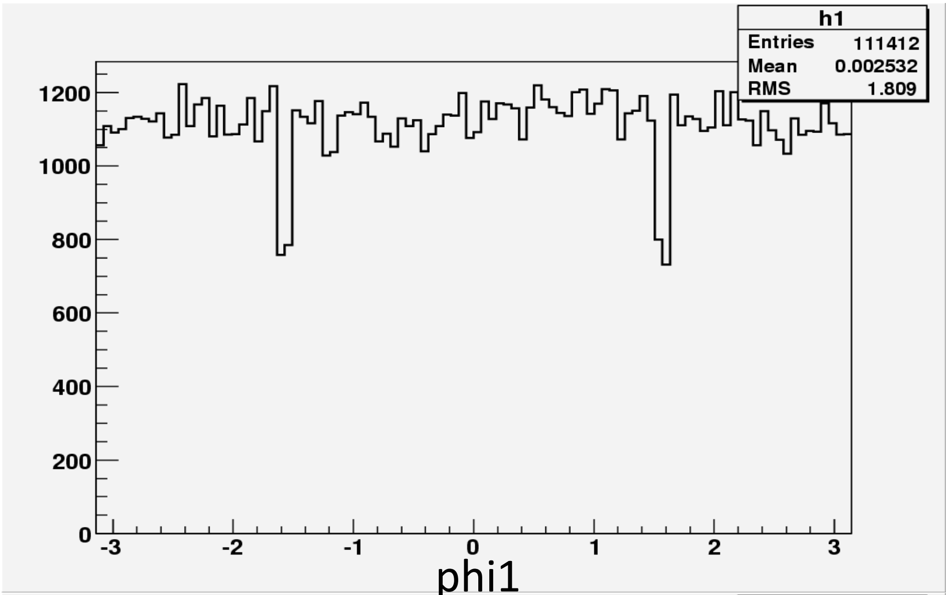
- Ntrk=2
- Ntrk=1
- Ntrk=0

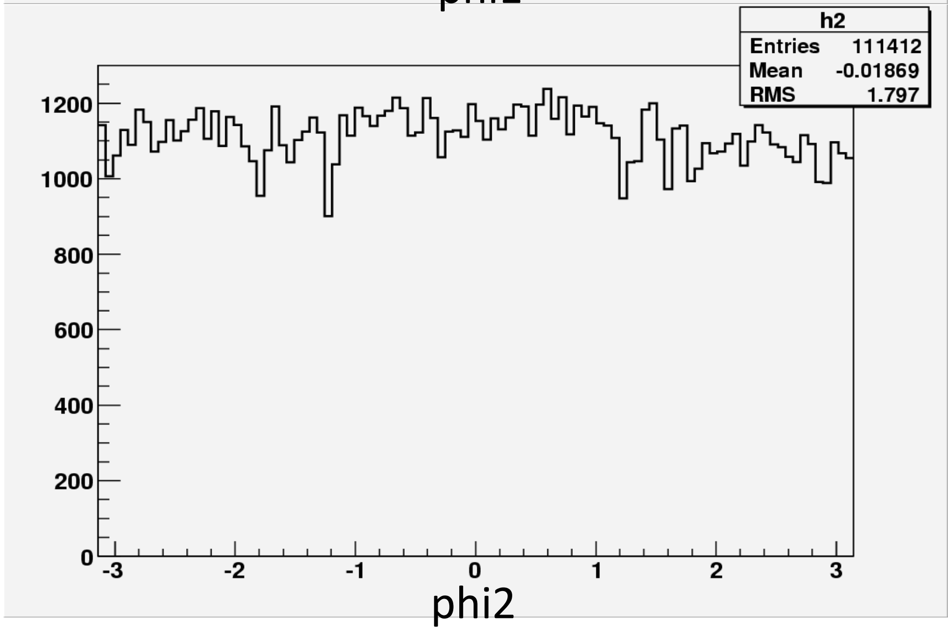
Performance Check

- 鉴别效率（Efficiency）
- 误判率
- CPU time

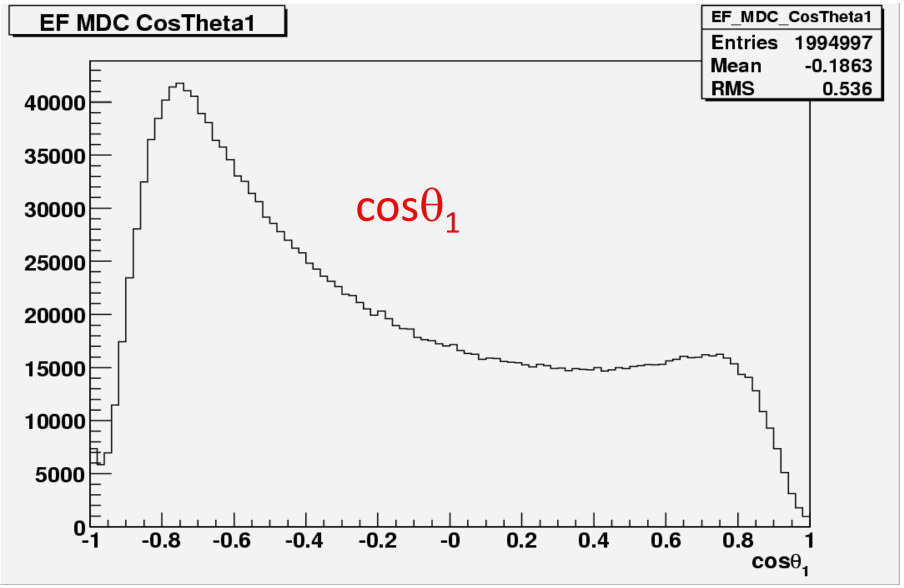
ϕ check for $e^+e^- \rightarrow e^+e^-$

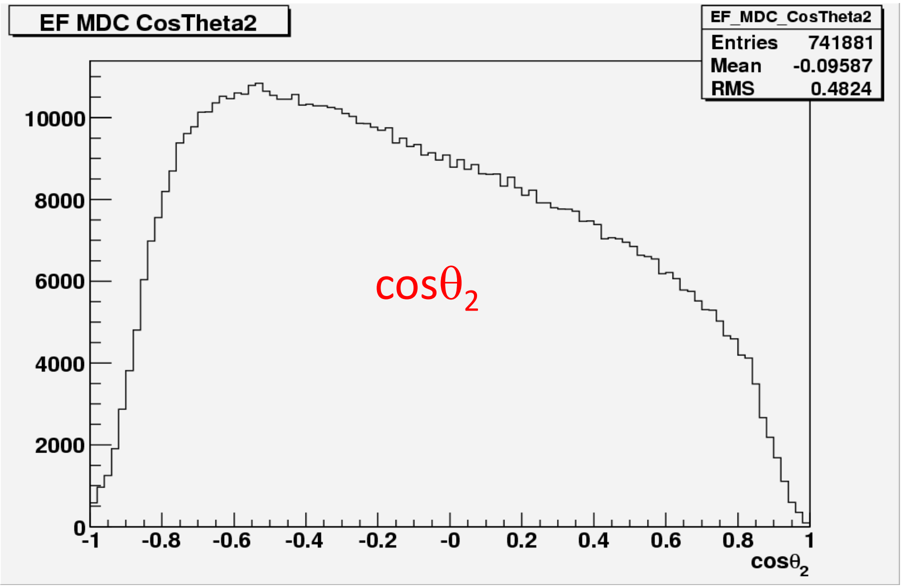




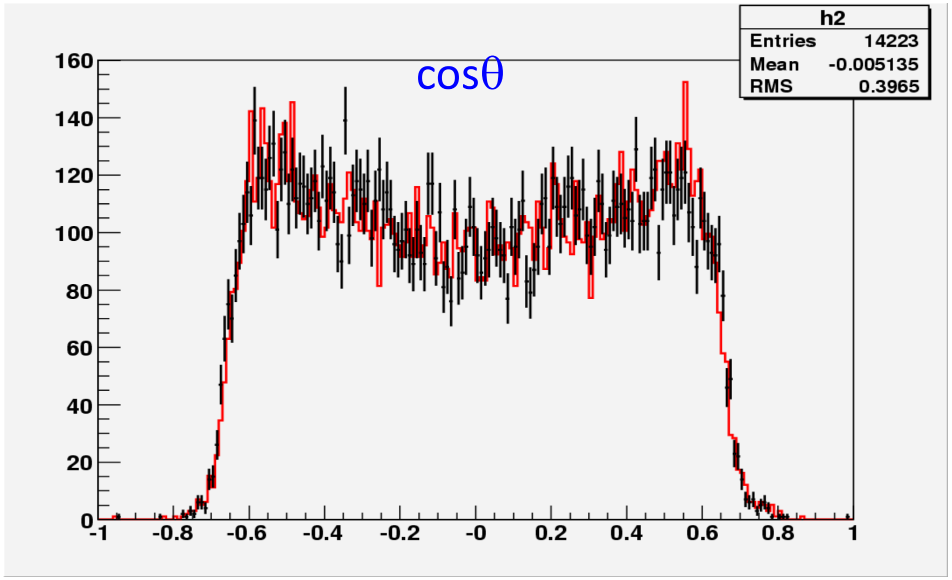


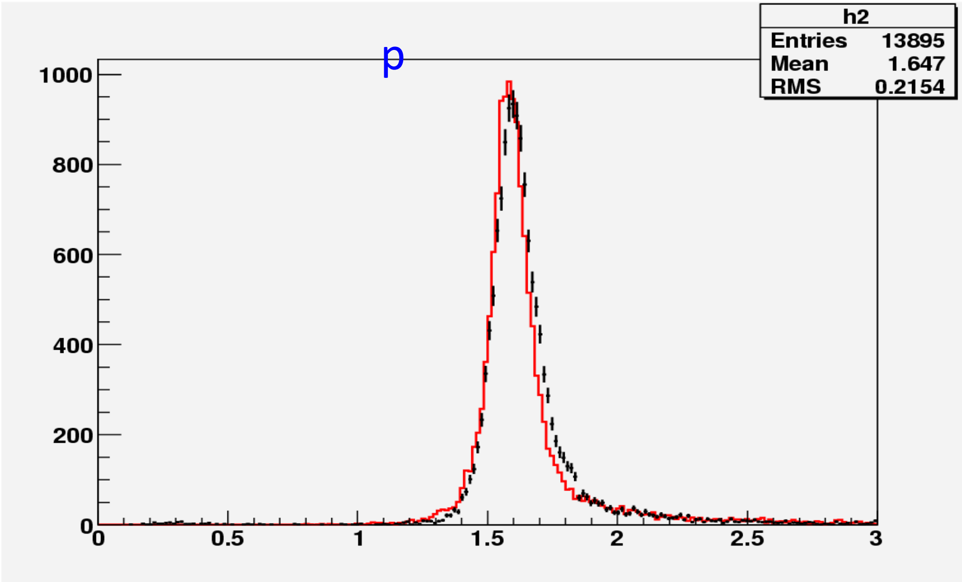
All Etot<E_L





$\mu^+\mu^-$ tag





鉴别效率

Sample 3.78GeV	Bee	Eee	μμ	γγ	hadron	Physics	Background	
							Beamgas	Cosmic ray
Barrel Bhabha	78.25%	0.04%	1.16%	0.42%	20.13%	100%	0	0
Endcap Bhabha	2.82%	71.49%	3.58%	0.21%	21.91%	99.998%	0	0.002%
dimuon	0	0	72.47%	0	26.32%	98.79%	0.97%	0.24%
Diphoton	0.95%	0.02%	0	80.33%	18.25%	99.55%	0.37%	0.07%
hadron	0.49%	0	0.002%	0.27%	99.06%	99.82%	0.04%	0.13%
Beam-gas	0	0	0	0	50.89%	50.89%	40.86%	8.25%

Exclusive $J/\psi\rightarrow$	hadron
$\gamma\eta_c$	99.93%
$\gamma\pi^0$	99.91%
KsKl	99.66%
$\Lambda\Lambda_{bar}$	99.06%
$pp_{bar}\pi\pi\pi$	99.94%
pp_{bar}	95.11%
$\rho\pi$	99.81%
$\Xi\Xi_{bar}$	99.04%

Exclusive ($J/\psi\rightarrow e^+e^-$)	hadron
$\pi^0 J/\psi$	99.95%
$\pi^0\pi^0 J/\psi$	97.13%
$\eta J/\psi$	98.54%
$\pi^+\pi^- J/\psi$	99.99%
$\gamma\chi_{ci}(\chi_{ci}\rightarrow\gamma J/\psi)$	99.10%

50000 sample, EvtGen's decay mode

本底误判率

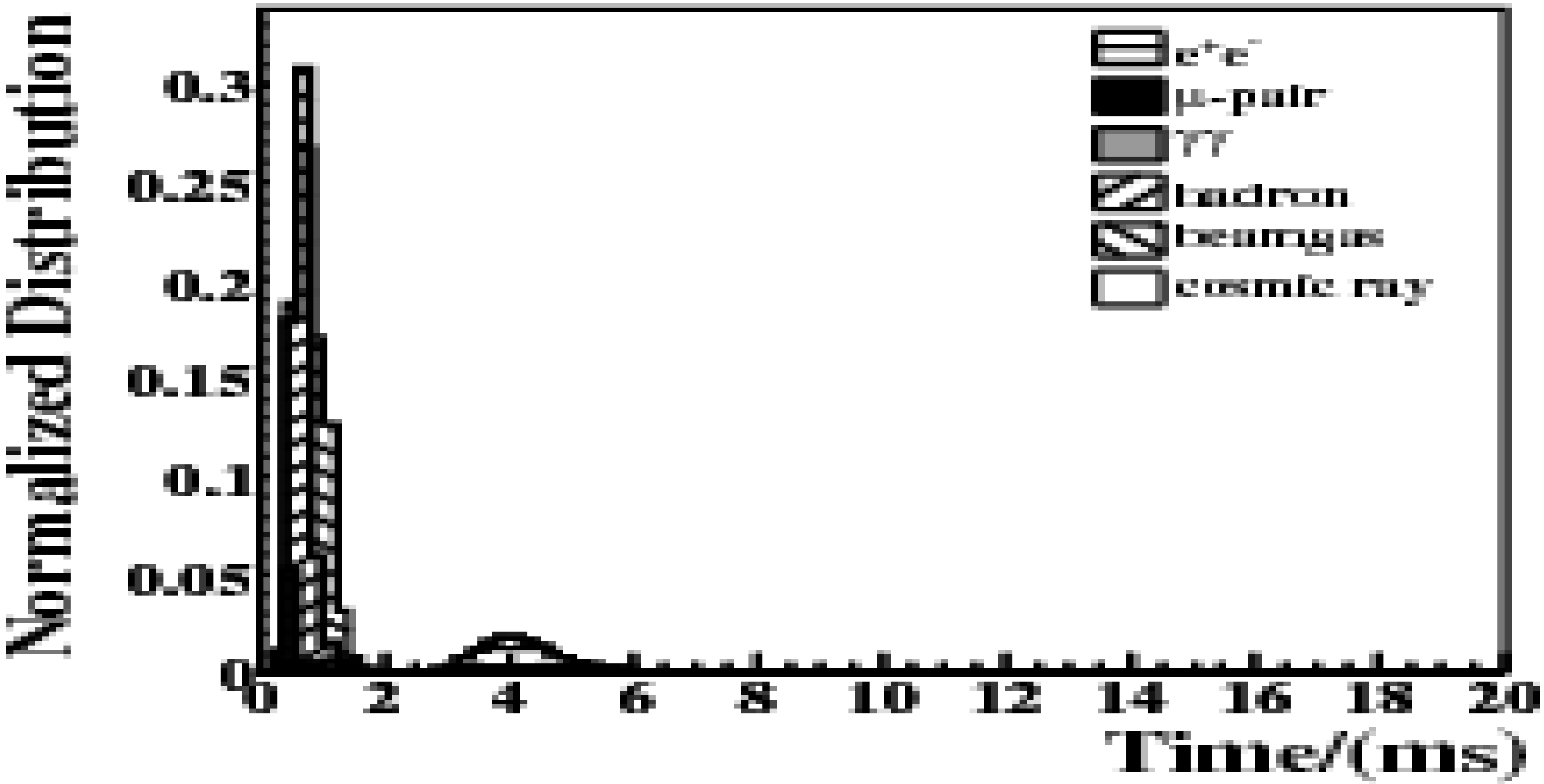
- **DQM hadronic events (Xiaodong SUN):**
  - ε_h= 85.6%
  - ε_{ee}= 6.6%
  - $\varepsilon_{beamgas}$= 0.2%
  - ε_{cosmic}= 0%
- **Offline hadronic events (Miao HE):**
  - ε_h= 71.6%
  - ε_{ee}= 18.4%
  - $\varepsilon_{beamgas}$=0.04%
  - ε_{cosmic}= 0%
- **Offline cosmic rays (Chuan ZHAO):**
  - ε_{cosmic}= 67.8%
  - ε_{uu}= 0.04%
- **Offline dimu (Chuan ZHAO):**
  - ε_{uu}= 70.4%
  - ε_{cosmic}= 0.4%
  - $\varepsilon_{beamgas}$= 0%
- **$\psi' \to$anything (MC):**
  - ε_h= 93.6%
  - ε_{ee}= 0.5%
  - ε_{bkg}= 0.3%
- **Bhabha (MC):**
  - ε_h= 29.6%
  - ε_{ee}= 51.6%
  - ε_{bkg}= 0.5%
- **Dimu (MC):**
  - ε_{uu}= 18.8%
  - ε_{cosmic}= 3.2%
  - $\varepsilon_{beamgas}$= 3.6%
- **Cosmic (MC, the parameters of trigger simulation still not best):**
  - ε_{uu}= 0%
  - ε_{cosmic}= 24.1%
  - $\varepsilon_{beamgas}$= 1.0%

bkg= cosmic+beamgas

“Good”样本检查本底

- psi'--> pi+pi-J/psi (inclusive)(from Li Gang)
  - GJunk 0%
  - Beamgas 0%
  - GBeamgas 0.00308631%
  - Cosmic 0.00881755%
  - GCosmic 0.00262734%
- psi'--> pi+pi-J/psi (mu+mu-) (from Li Gang)
  - GJunk 0%
  - Beamgas 0%
  - GBeamgas 0%
  - Cosmic 0%
  - GCosmic 0%
- psi'--> pi+pi-J/psi (e+e-) (from Li Gang)
  - GJunk 0%
  - Beamgas 0%
  - GBeamgas 0%
  - Cosmic 0%
  - GCosmic 0%
- Inclusive hadron (from Wang Zhiyong)
  - GJunk 0%
  - Beamgas 0%
  - GBeamgas 0%
  - Cosmic 0.0035%
  - GCosmic 0.0024%

CPU Time

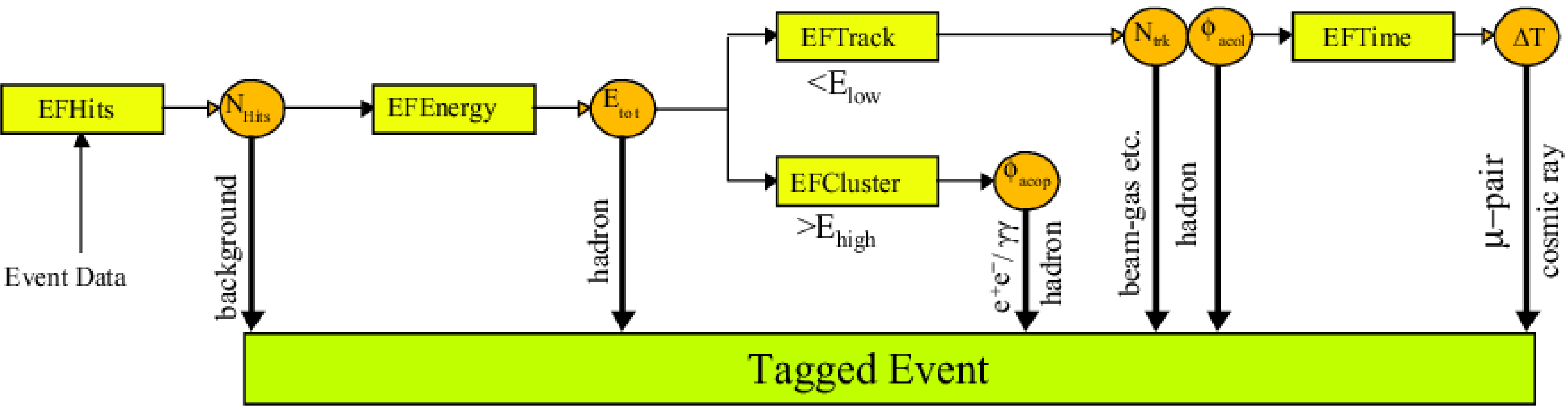


– Online: 0.8 ms/event (Pentium D 3.0GHz)
– MC: 0.9 ms/event (Xeon 2.8GHz)

小结

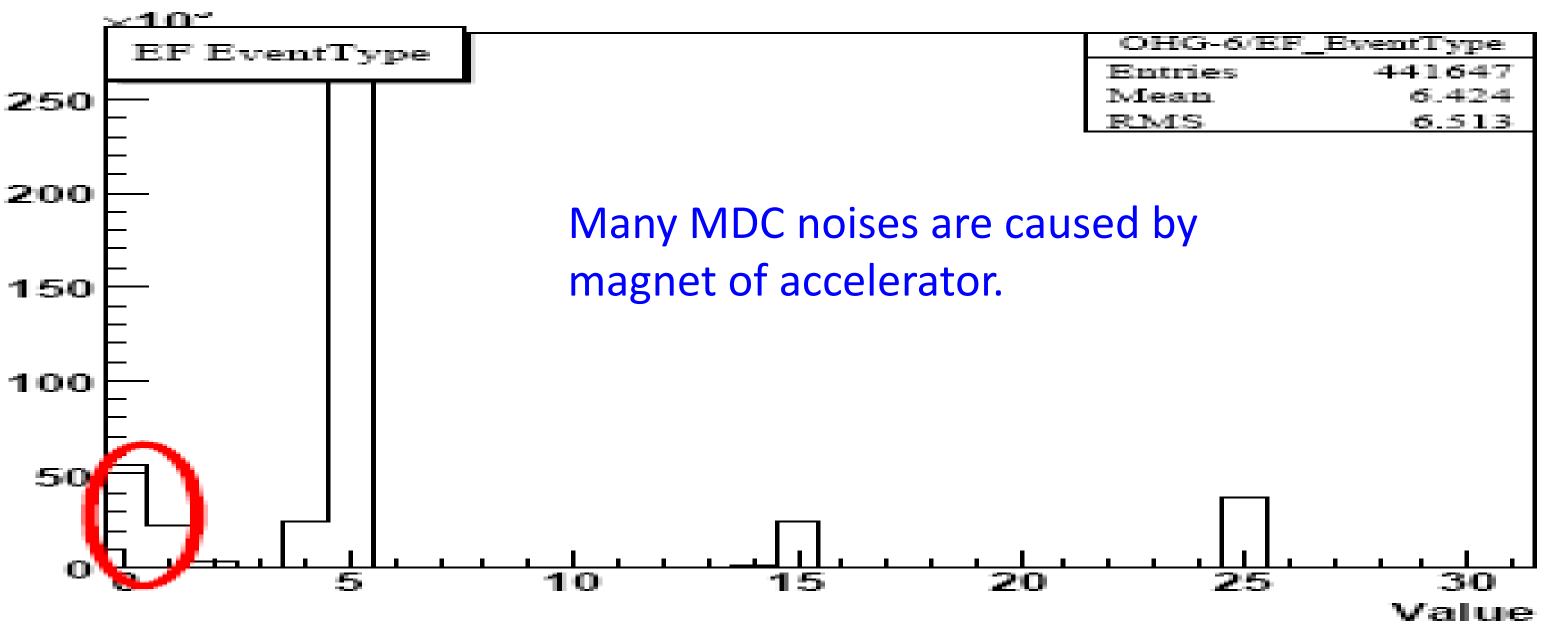
- 利用不同事例的某些变量分布截然不同，对变量值做截断完成事例的鉴别分类。根据需求决定严格度和广度。
- 不同的事例类型，经历不同的算法流（类似于硬件触发道）。
- 较高的鉴别率和较低的误判率。
- 占用较小的CPU时间，足以满足在线系统要求。

A stable, pragmatic, effective and fast algorithm is not best for all cases!



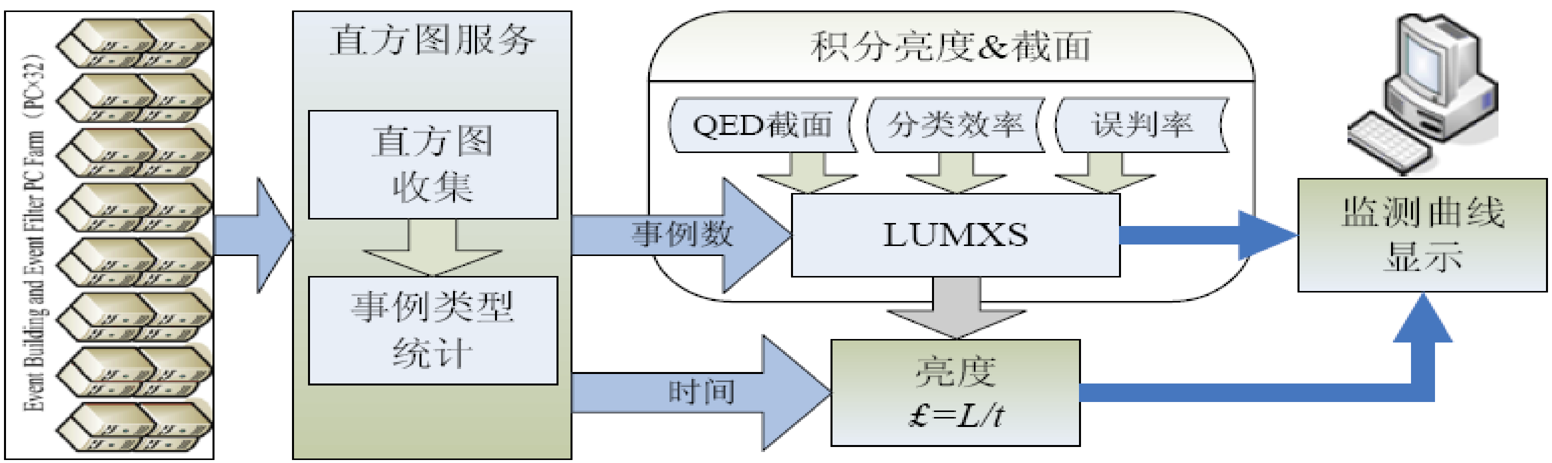
应 用

- 计算并监控在线亮度
- 反馈束流和探测器状态
- 快速刻度样本



亮度计算流程

- 事例分类结果送到专用服务器处理
- 统计各类型事例数
- 计算:
  - size of buffer: set while start run (1 min)
  - Interval time for updating buffer: 30 seconds
- 监视: 把计算结果更新到 IS 和 DIM服务。



在线亮度估计原理

- 假定亮度随时间满足指数衰减

$$\pounds\,(t) = \pounds\,(0)\,e^{-ct}$$

- 在线亮度测量的期望值

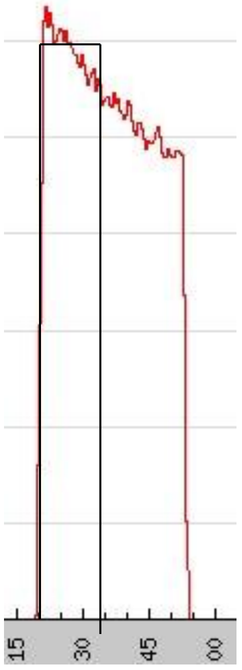
$$\pounds = \frac{1}{\Delta t}\int_{t_1}^{t_1+\Delta t} \pounds\,(0)\,e^{-ct}\mathrm{d}t = \pounds\,(0)\,\frac{e^{-ct_1} - e^{-c(t_1+\Delta t)}}{c\Delta t}$$

- 与[t1,t1+Δt]这段时间内瞬时亮度的最大偏离为

$$\pounds\,(0)\left[(c\Delta t - 1)\,e^{-ct_1} + e^{-c(t_1+\Delta t)}\right] / \,(c\Delta t)$$

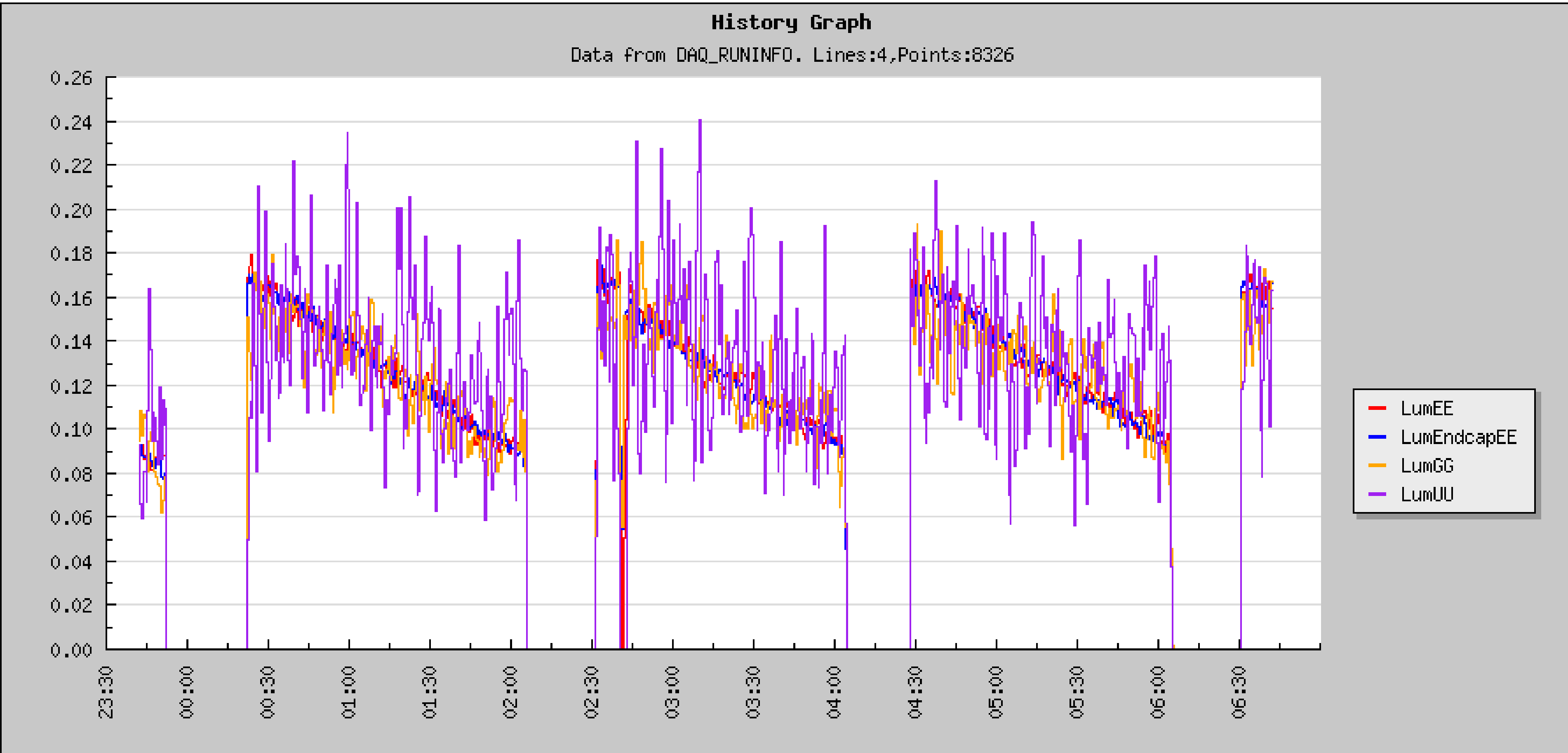
$$\left[1-\frac{1-\exp(-\Delta t/\tau)}{\Delta t/\tau}\right]\times 100\%$$

<1%



e+ life: 1.5h~2h
e- life: >2h
Lum. Life: >0.85h

Luminosity Monitor



The statistics of $e^+e^- \to \mu^+\mu^-$ is small in 1 min.

截面估计

- 亮度

$$\bar{L} = \frac{\frac{L_{ee}}{\sigma_{L_{ee}}^2} + \frac{L_{\mu\mu}}{\sigma_{L_{\mu\mu}}^2} + \frac{L_{ee}}{\sigma_{L_{\gamma\gamma}}^2}}{\frac{1}{\sigma_{L_{ee}}^2} + \frac{1}{\sigma_{L_{\mu\mu}}^2} + \frac{1}{\sigma_{L_{\gamma\gamma}}^2}}$$

- 截面

$$\sigma = N / \varepsilon L$$

- 信噪比

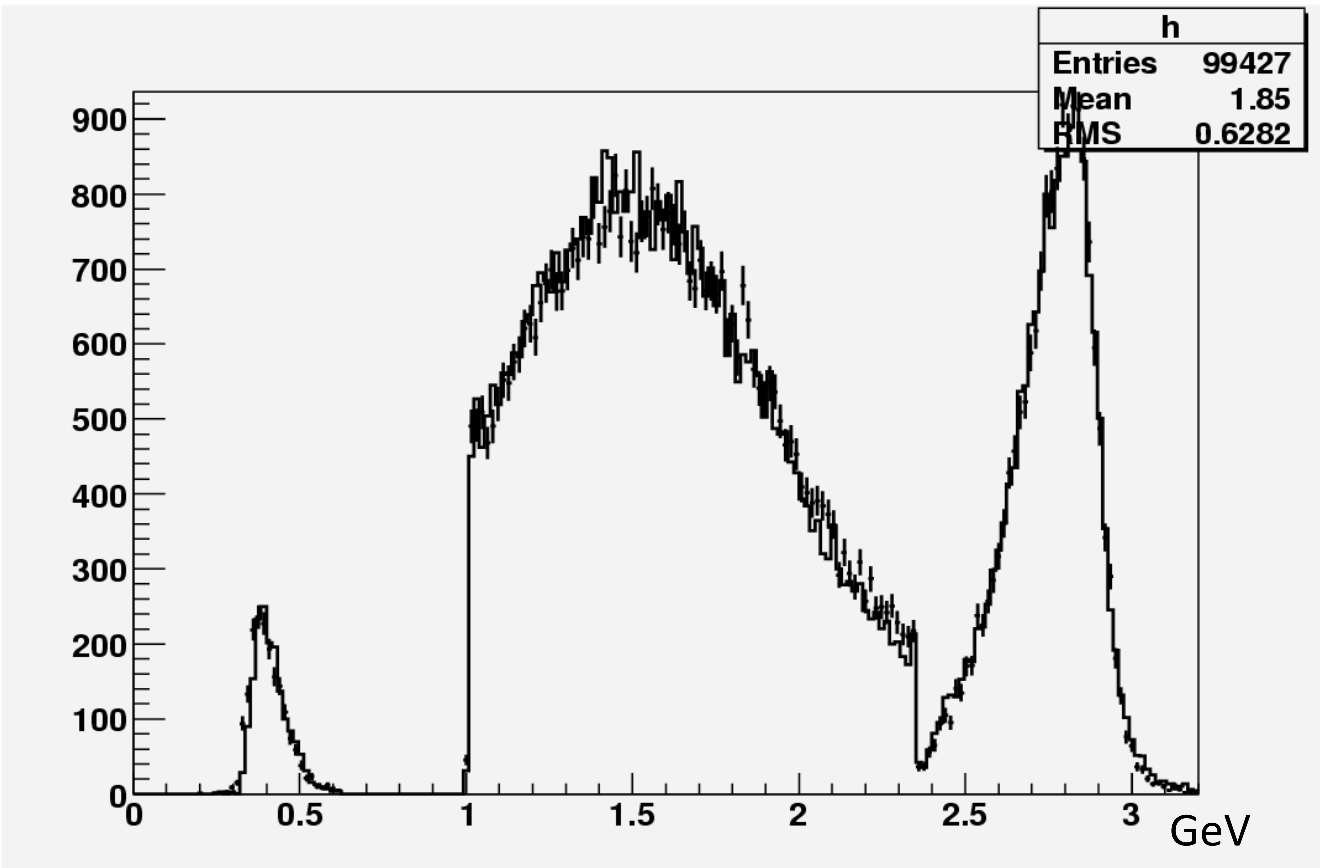
$$SNR = \frac{\sigma_0}{\sigma_0 + \sigma}$$

Fit for Cross Section

$$\chi^2 = \sum_{j=ibin} \frac{\left(N_j^{DATA} - kN_j^{MC} - k' \sum_{i=con} n_{ij}^{MC} \sigma_i\right)^2}{N_j^{DATA}}, \quad \sum_{j=ibin} N_j^{DATA} = \sum_{j=ibin} \left(kN_j^{MC} + k' \sum_{i=con} n_{ij}^{MC} \sigma_i\right)$$

Fit k to minimize χ^2

J/ψ->anything



σ_{res} for RUN10133:

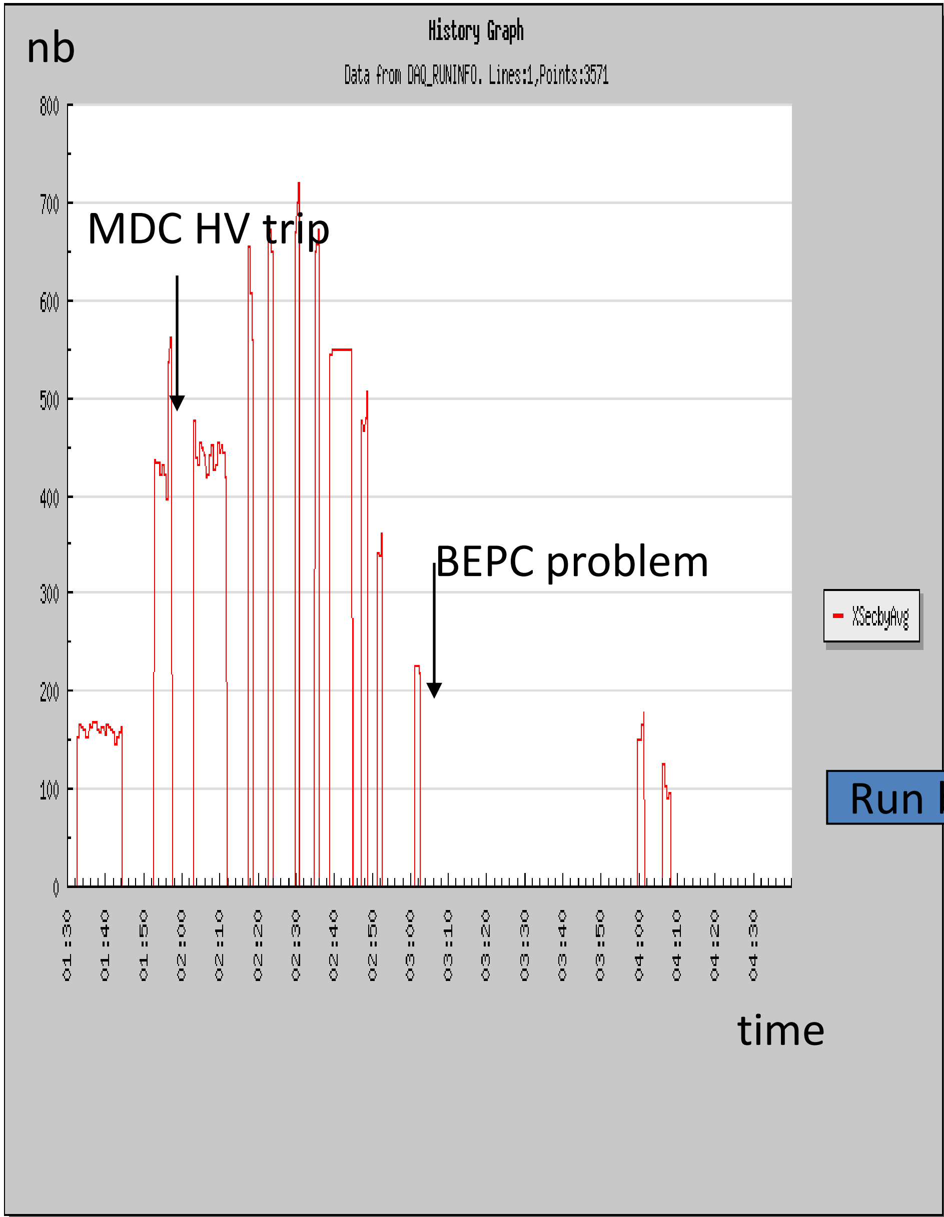
Fit result 2599nb

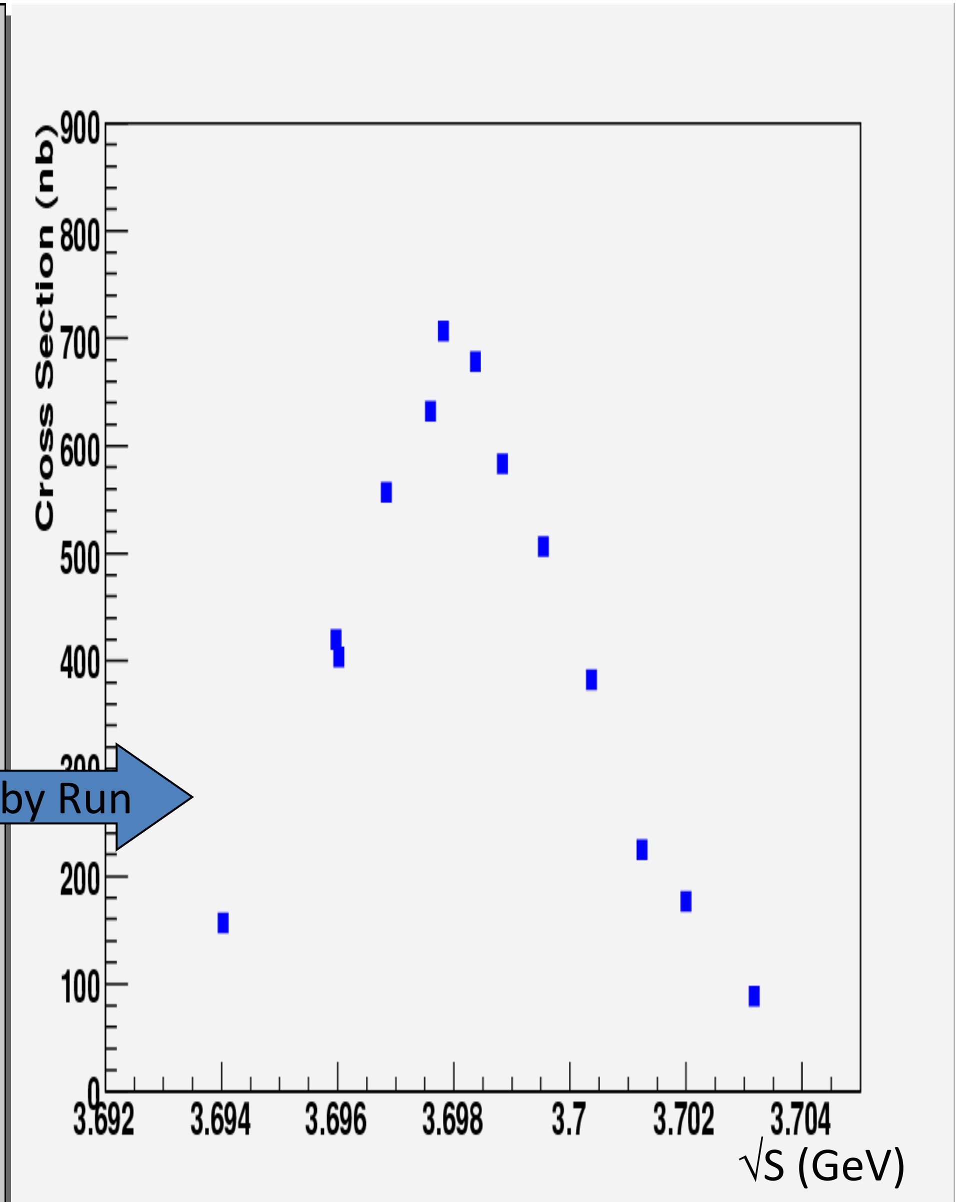
Count 2600nb

$N_{res} = N_h - \sum \varepsilon_i N_i$

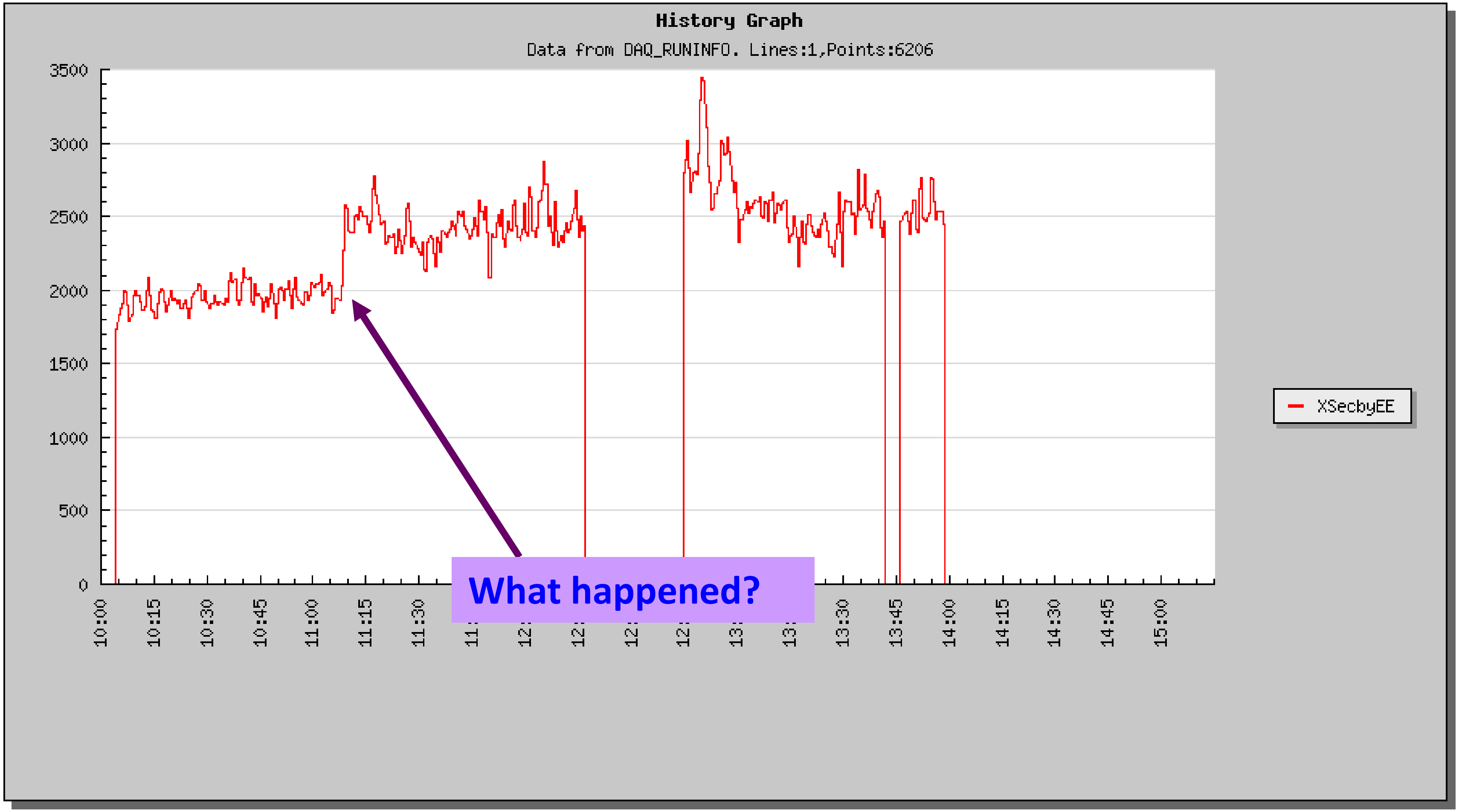
$\sigma_{res} = N_{res} / \varepsilon_{res} L$

ψ(2S) Scan Experiment

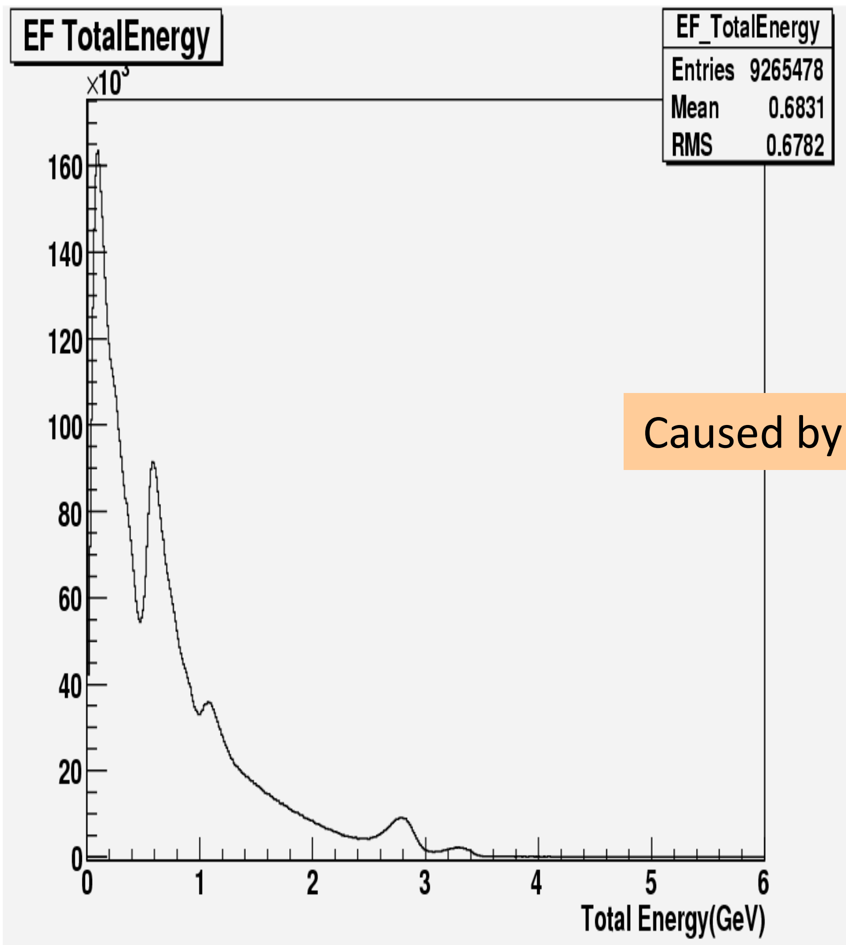




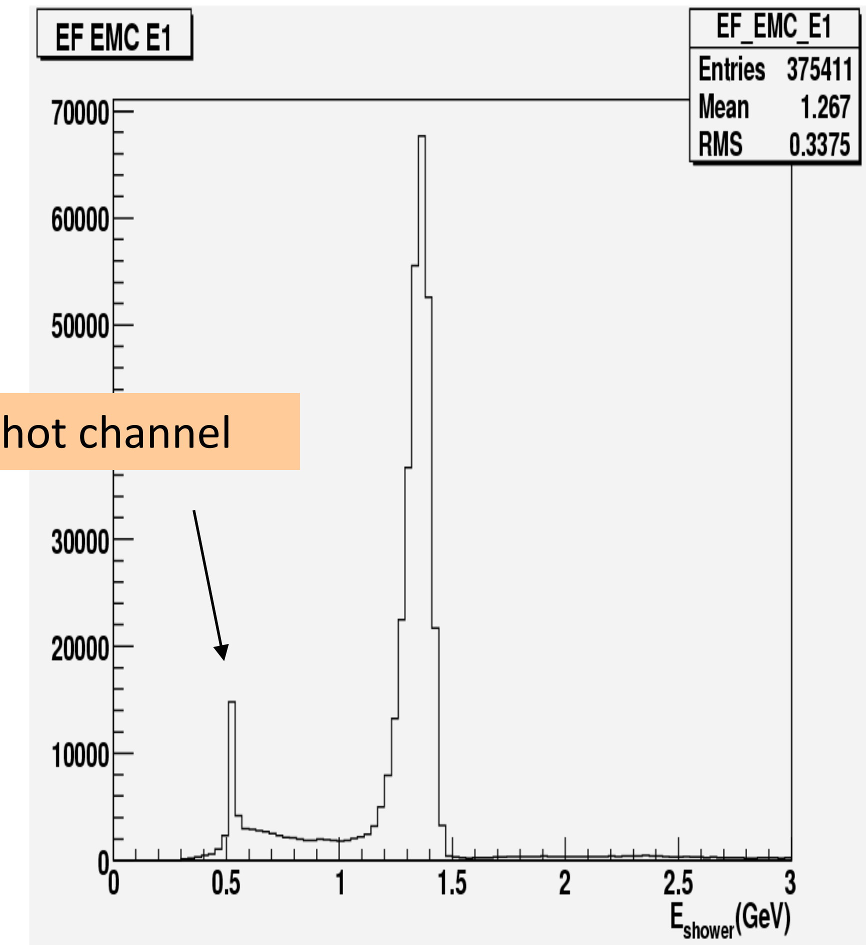
XSection Monitor



X to Cross Section



Caused by hot channel



MC results denote that XS ↑16% with one 0.5GeV hot channel; ↑6% (0.3GeV)

总结

- 在线事例分类是一个由软件实现的实时运行的简化分析系统。
- 经过多次取数的实际应用，在线事例分类系统可以稳定有效地运行。
- 对于最重要的应用——事例过滤，需要更细致的验证和改进。

谢谢!