

LHAASO-WCDA簇射事例重建与噪声过滤方法的研究

LHAASO 合作组

王晓洁

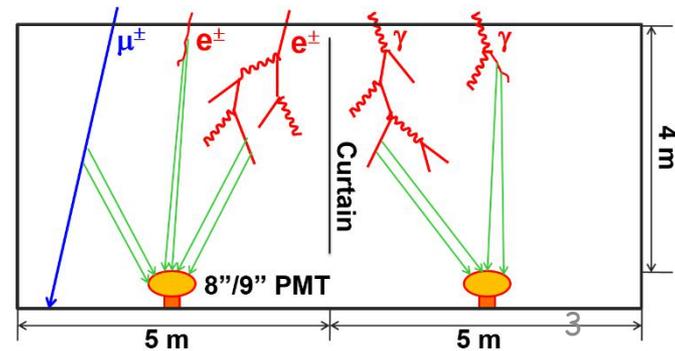
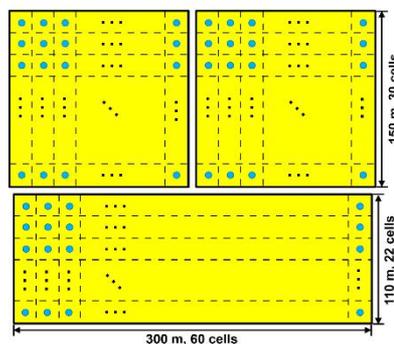
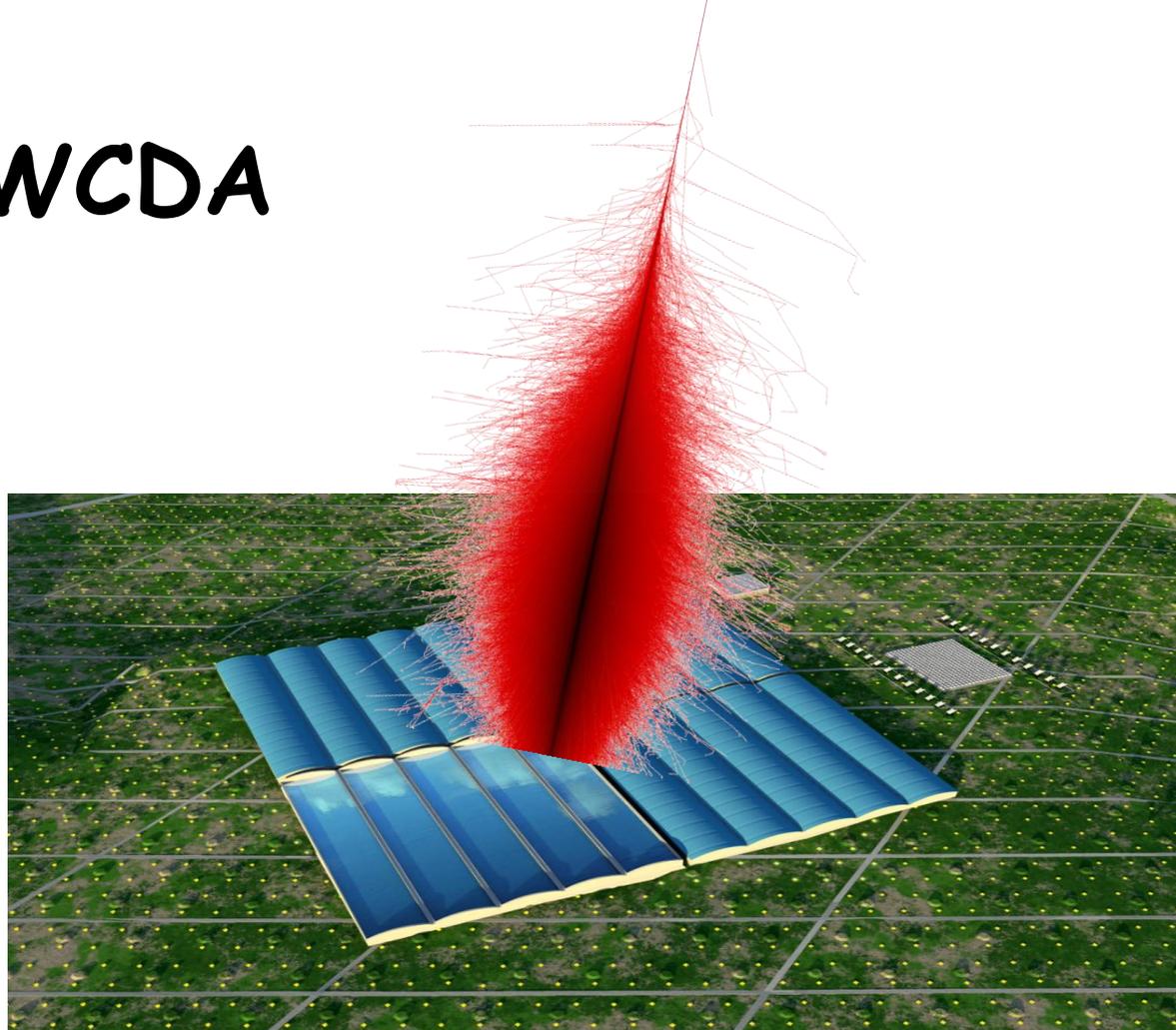
2016/08/15

目录

- ◆ LHAASO-WCDA
- ◆ 簇射事例重建
- ◆ 在线快速重建方法
- ◆ 算法优化
- ◆ 总结

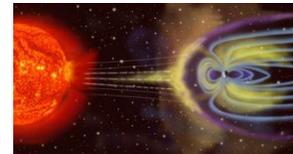
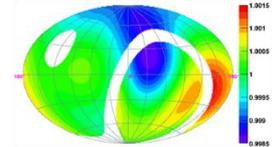
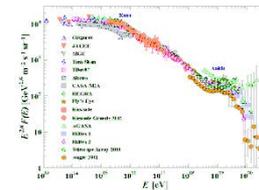
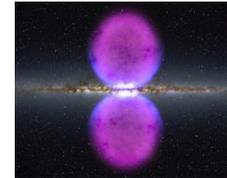
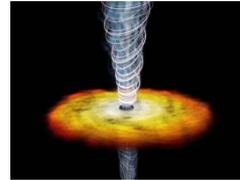
LHAASO-WCDA

- ◆ 占地78,000 m²;
- ◆ 3120个单元;
- ◆ 有效水深4m。



物理目标

- ◆ 甚高能伽马天文观测 (100 GeV-30 TeV):
 - 河外源 (尤其是AGN耀变信号) 的全天普查;
 - GRB的高能发射信号的探测;
 - 河内源;
 - 弥散伽马射线的探测.
- ◆ 宇宙线的测量 (10 TeV-1 EeV):
 - 甚高能宇宙线天文;
 - 宇宙线能谱;
 - 强相互作用模型.
- ◆ 其他:
 - 暗物质信号的探测;
 - 太阳风暴和行星际磁场的测量.



簇射事例重建（1）

◆定义：

- 对于每个触发的宇宙线事例，通过各个探测器记录的时间和电荷信息，可以重建出原初粒子的方向、芯位和能量。

◆重要性

- 事例的重建是物理分析的第一步，也是最重要的一步，其精度会直接影响到物理结果；
 - 方向的重建精度直接影响探测器对源的灵敏度；
 - 芯位的重建精度直接影响簇射方向和能量的重建精度。

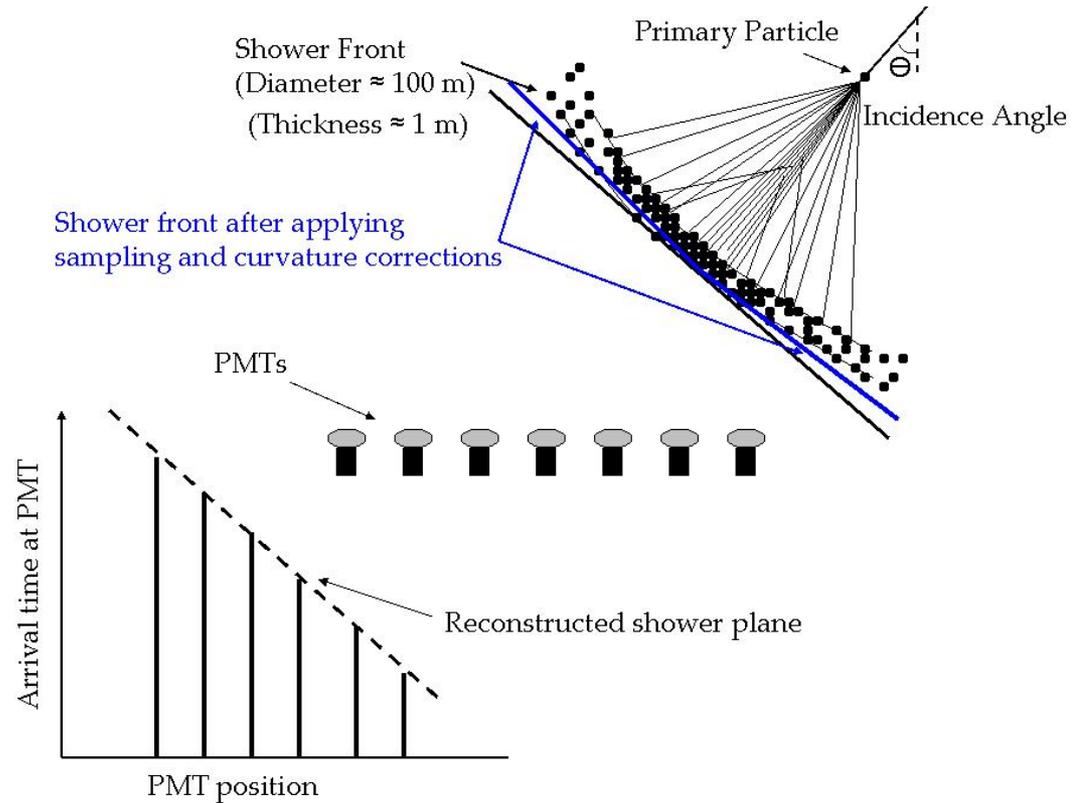
方向重建

◆ 时间修正：

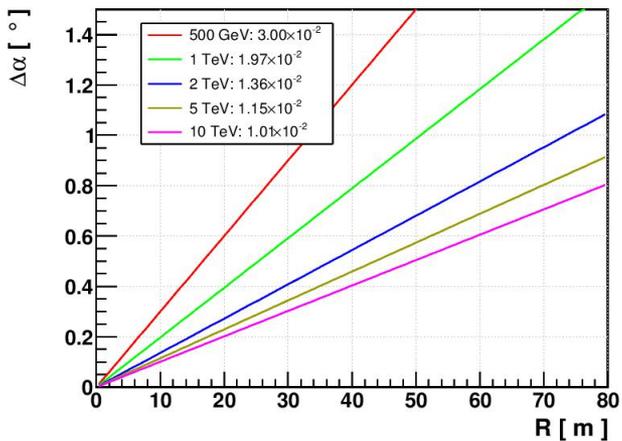
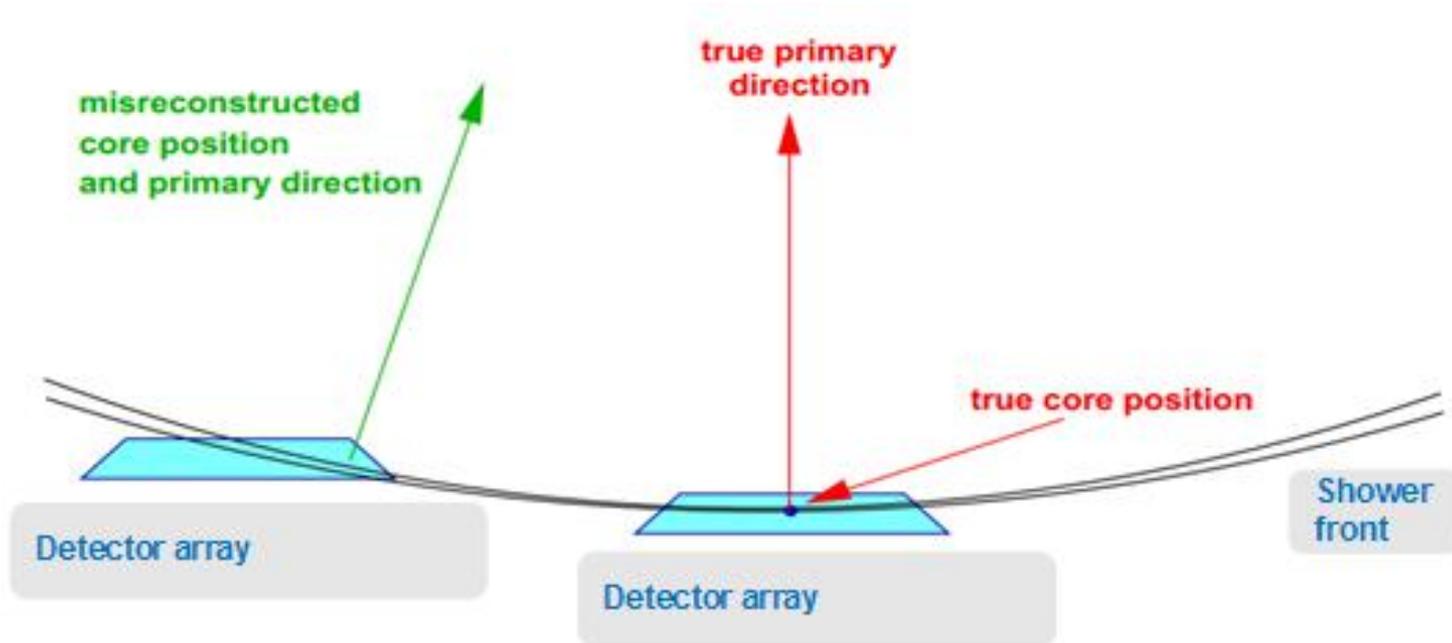
■ 采样修正；

- 簇射前锋面的厚度及其涨落
- 粒子的击中位置（重建时我们认为是单元中心）
- 入射角度
- 粒子数的多少

■ 曲面修正；



芯位重建



- ◆ 重心法；
- ◆ 树形分析法：MTree法、Tree法、Dtree法；
- ◆ NKG函数拟合；
- ◆ 2D-Gaussian 拟合；

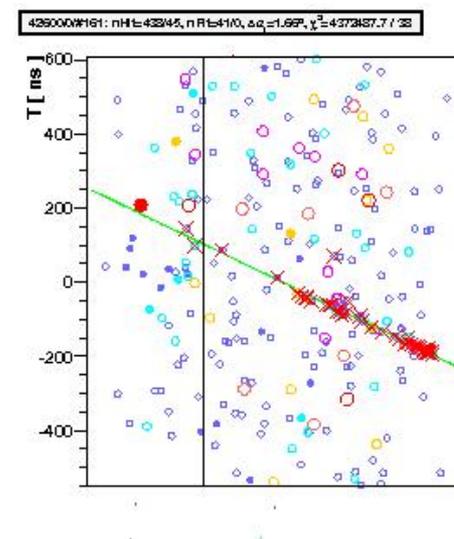
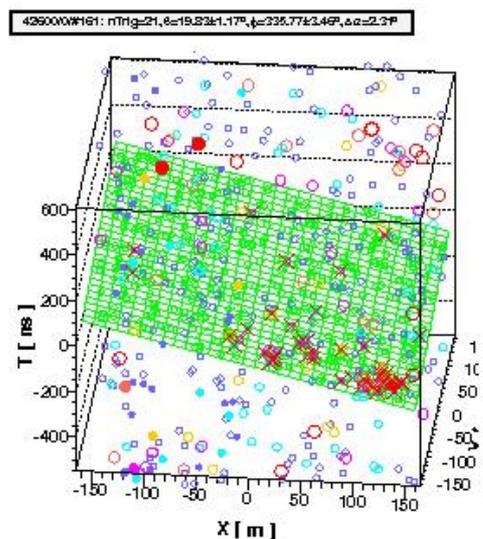
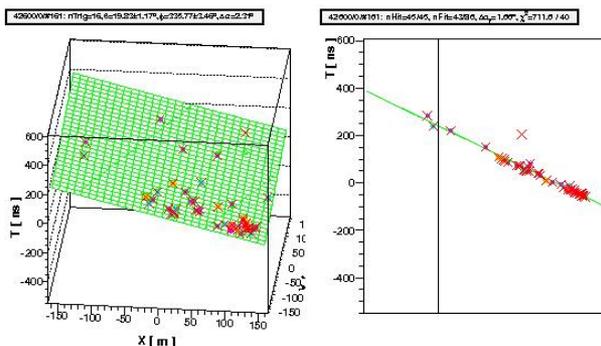
簇射事例重建 (2)

◆重建困难

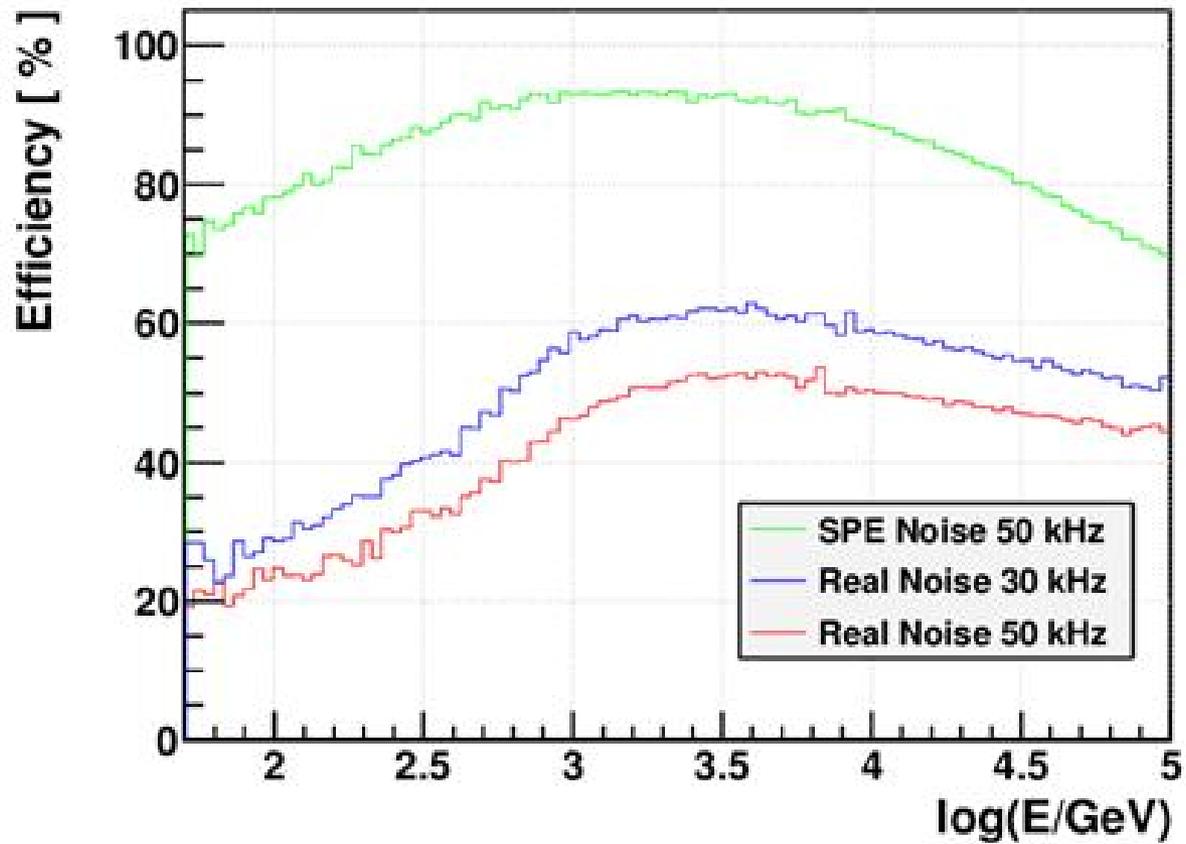
- 高噪声率 (~50kHz, 多光电子信号)
- 数据量大 (触发率高, 需考虑在线重建, 压缩数据)
- 前鋒面修正 (非简单锥面)
- 低能簇射重建 (无明显簇芯)

◆需要创新的重建方法

- 提高效率, 过滤噪声
- 快速



重建效率图



簇射事例重建（3）

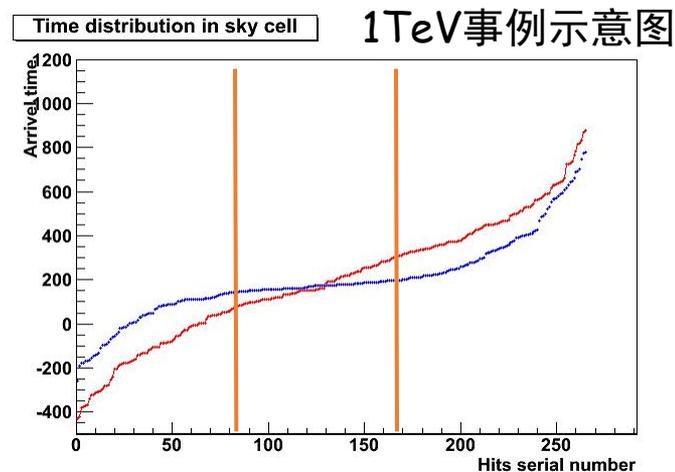
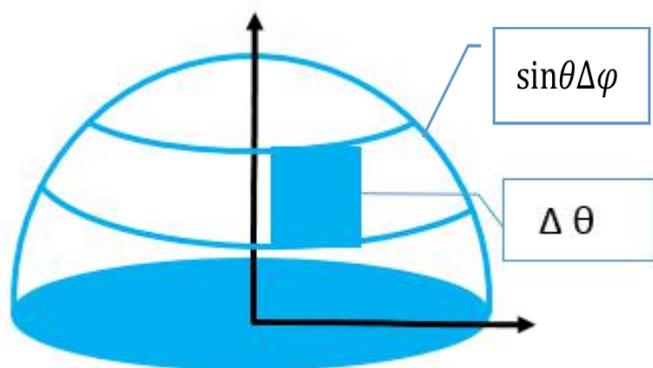
◆重建工作分为两步：

- 在线快速初步重建，主要用于排除大部分噪声，要求速度快，节省资源和成本；
- 在线高精度重建，包括芯位重建、方向重建、簇射前锋面曲面修正、能量重建等。

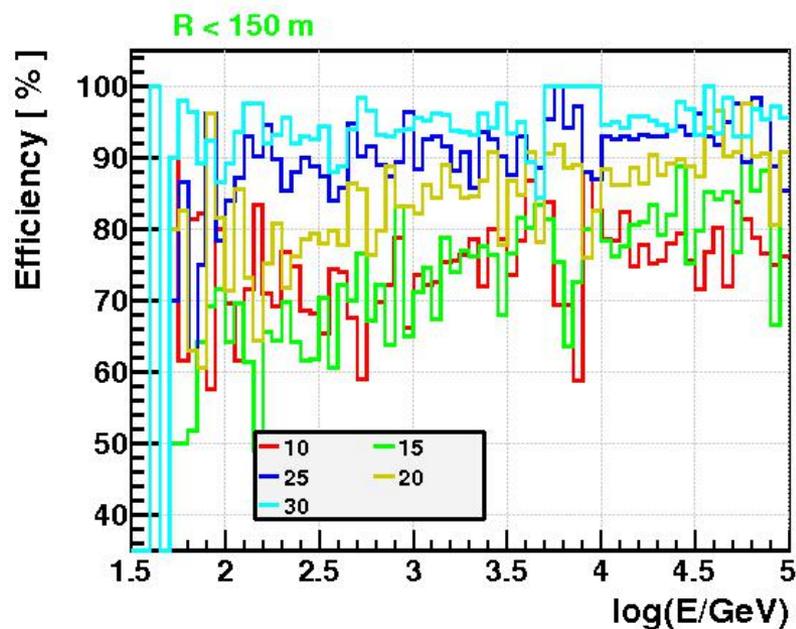
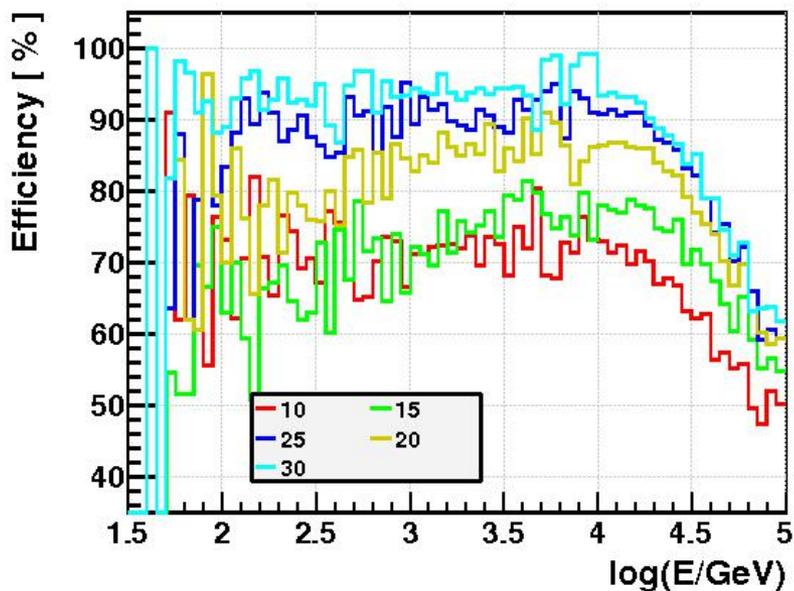
在线快速重建方法的研究

◆ WCDA自己发展的快速遍举法

- 天区分区（分bin）；
- 逐个扫描，排除噪声。
 - hits时间信息转换到特定方向上的前锋面上排序



优化前效率图



$$\eta = \frac{\text{真实信号保留比} > 90\% \text{的触发事例数}}{\text{触发事例总数}}$$

程序优化简介

◆ 数据

- 速度：cosmic-ray；效率：gamma；

◆ 要求：保证重建效率、提高程序运行速度

- 在当前算法下，找出最耗CPU的部分进行优化改进
- 算法优化

◆ 参数介绍

- X_σ ：事例筛选
- N_θ ：速度相关
- 排序窗口：速度相关
- 数据保留窗口：效率相关

参数介绍

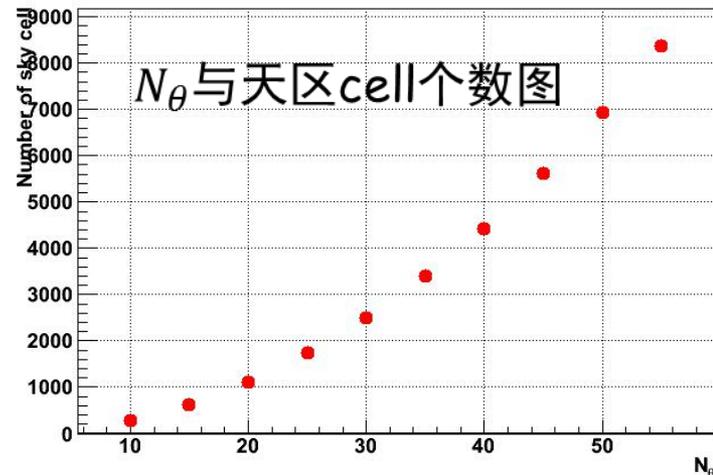
◆ X_σ

■ 事例标准 $\frac{|t_{k+1}-t_k|}{\sqrt{(\sigma_{t_{k+1}}^2 + \sigma_{t_k}^2)}} < X_\sigma$

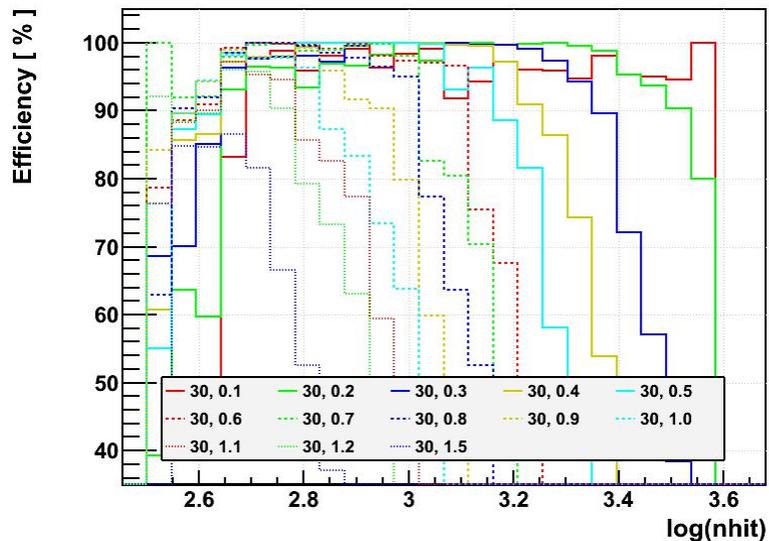
◆ N_θ

■ 决定了天区cell个数，速度相关

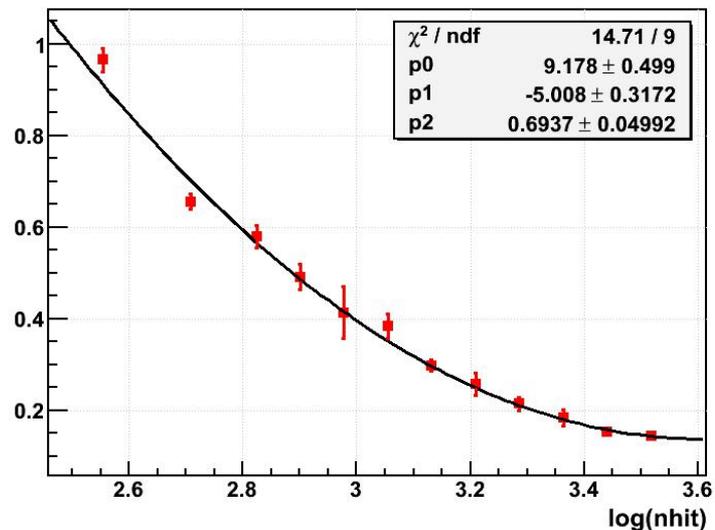
Graph



R < 150 m

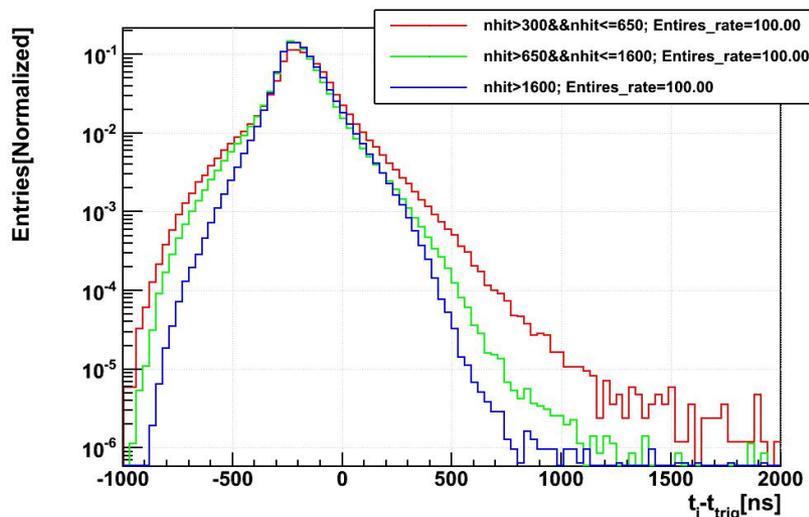


Opt xsigma

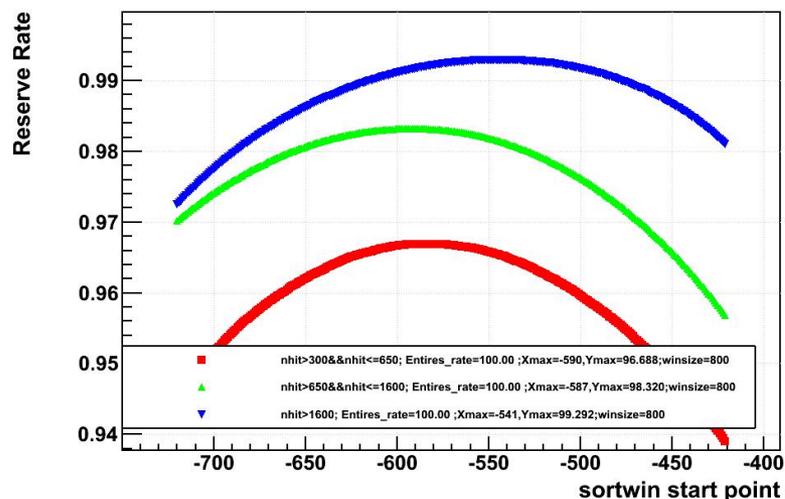


时间窗口优化---sort窗口

归一化事例分布图



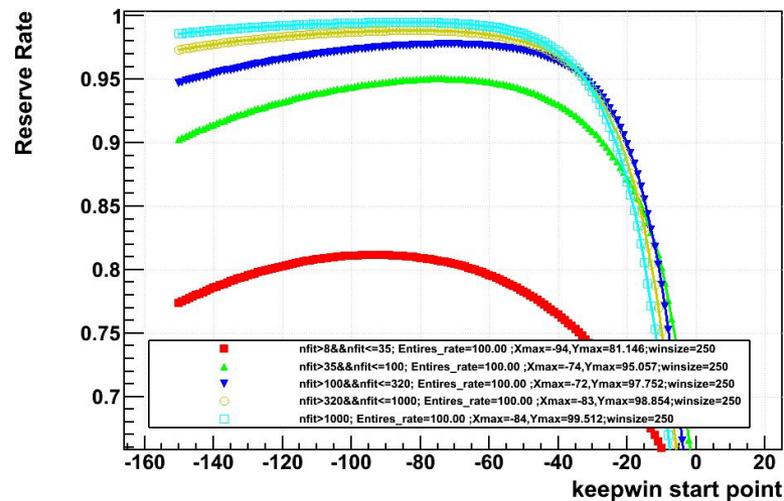
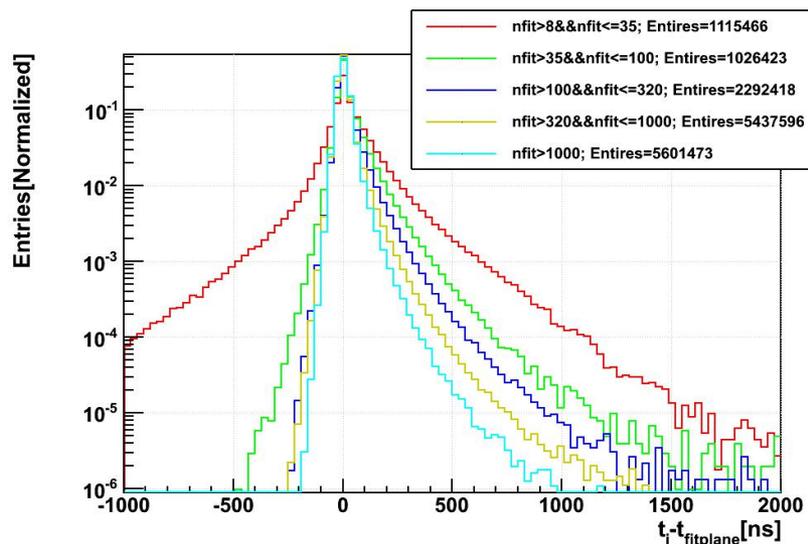
hit保留比与窗口起点关系图



◆ 要求窗口窄、hit保留率高

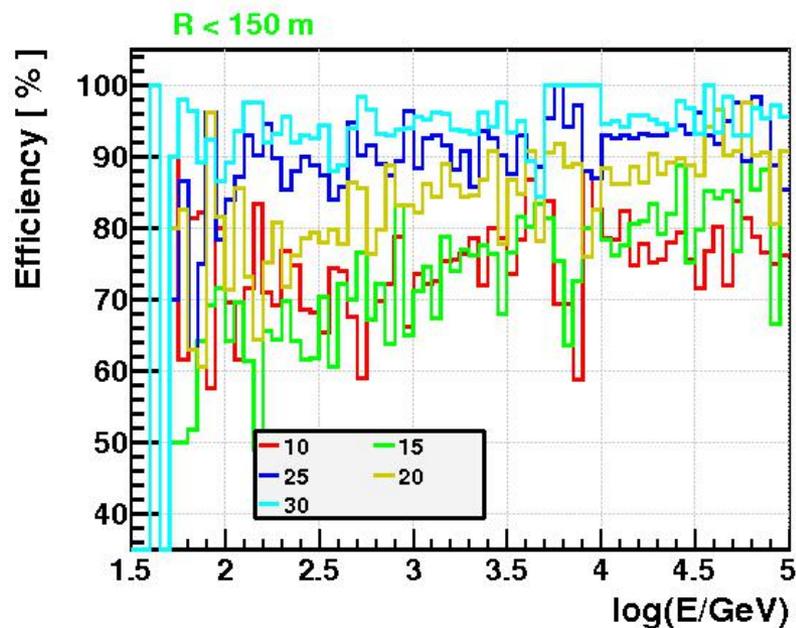
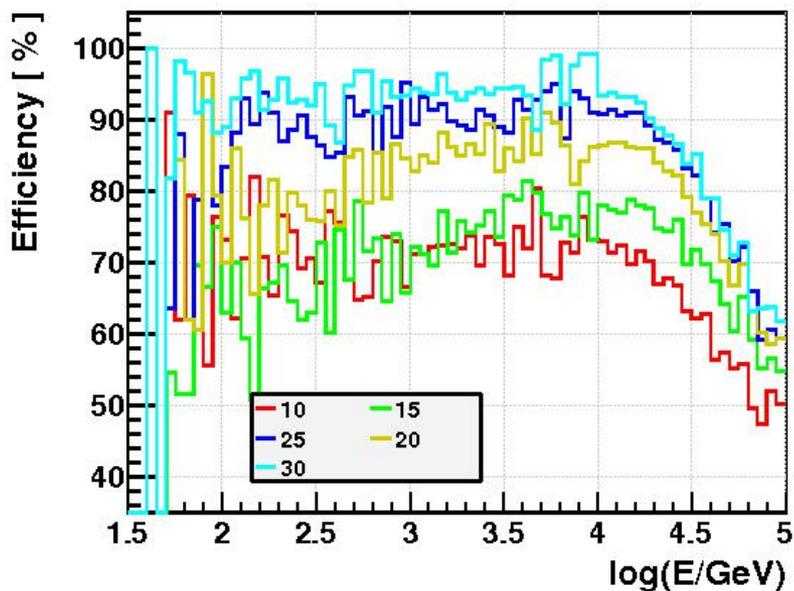
- 速度相关
- 窗口窄则sort函数中需要排序的数据个数n越小

时间窗口优化---keep窗口



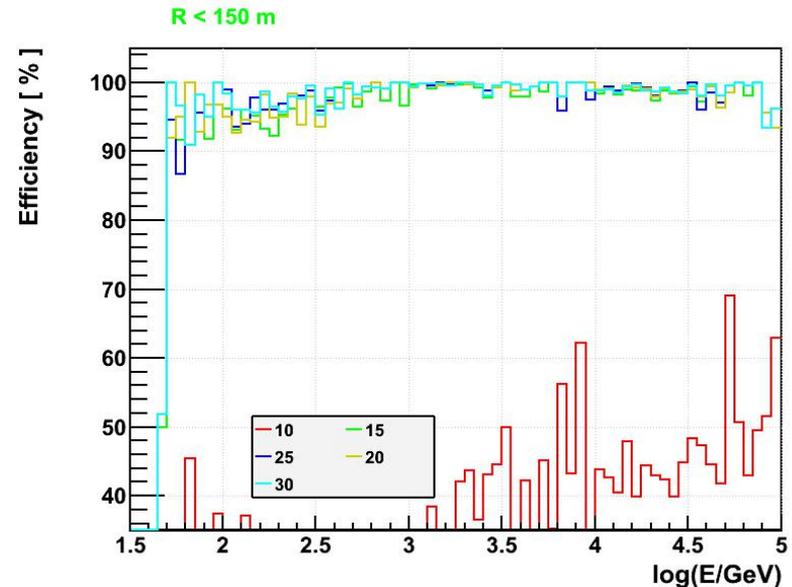
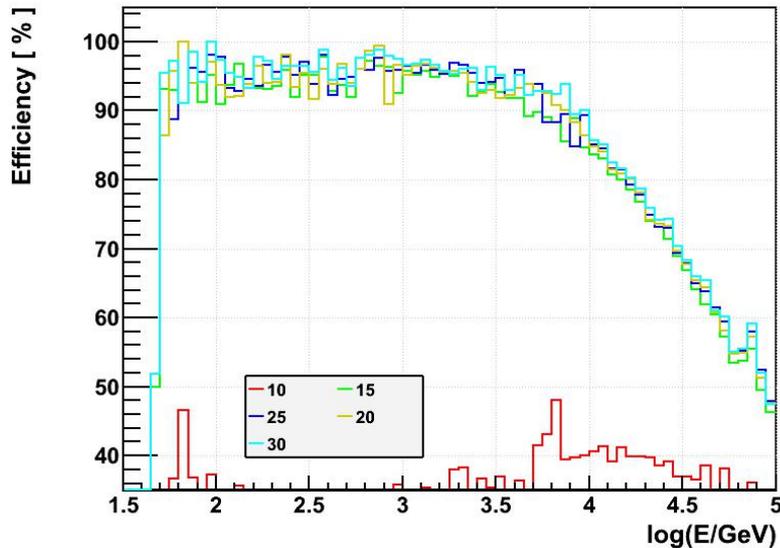
- ◆ 要求窗口窄、hit保留率高
 - 与效率及数据存储量相关
 - 效率计算公式中“真实信号保留比”与窗口有关
 - 只有keep窗口内的数据会被保留做离线分析

优化前效率图



$$\eta = \frac{\text{真实信号保留比} > 90\% \text{的触发事例数}}{\text{触发事例总数}}$$

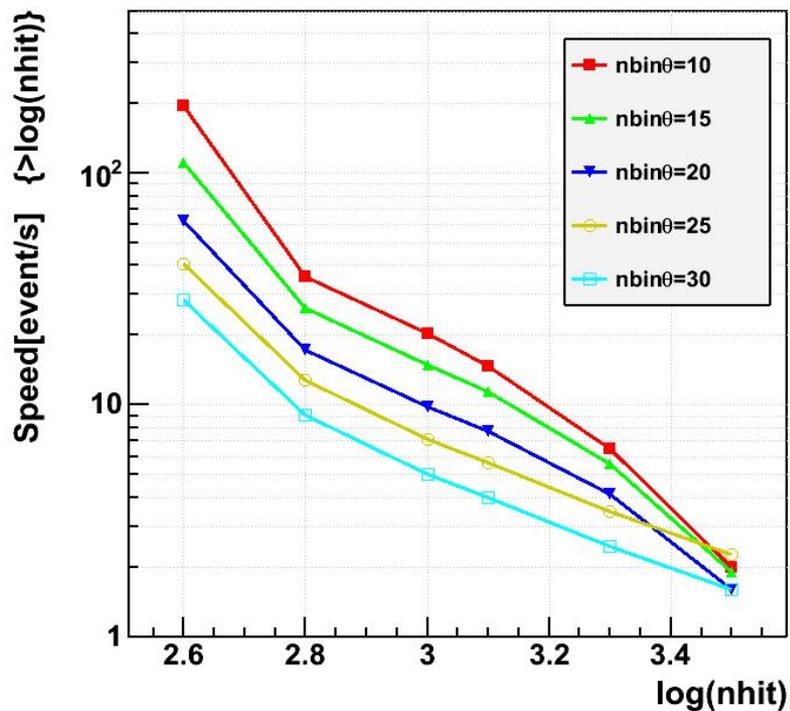
优化后效率图



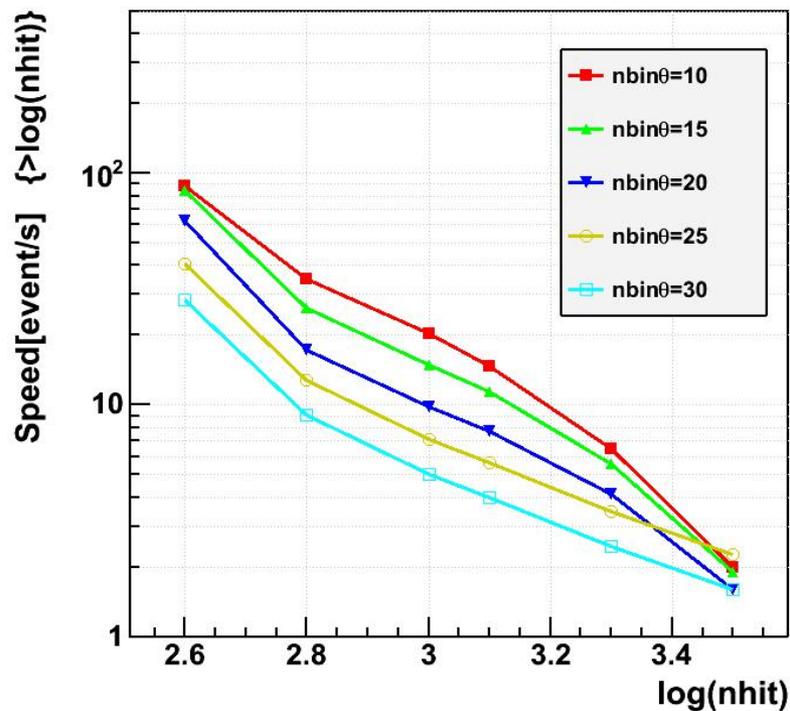
- ◆ 100GeV~30TeV能区，优化后效率达到80%以上，芯内事例几乎都在90%以上

优化后速度

速度上限

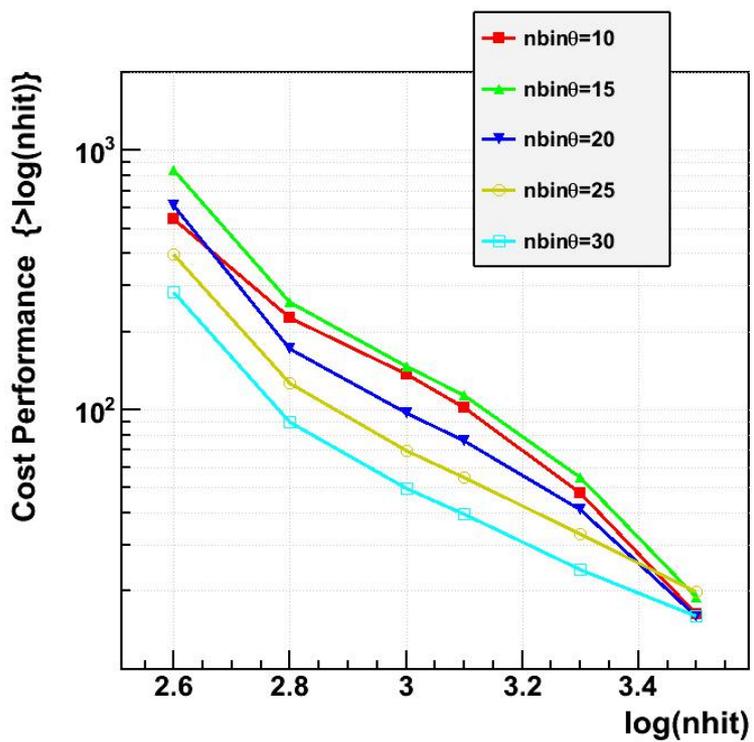


速度下限



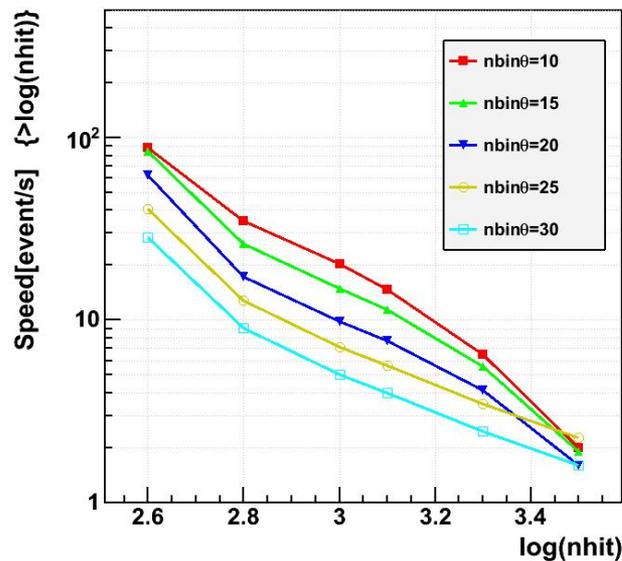
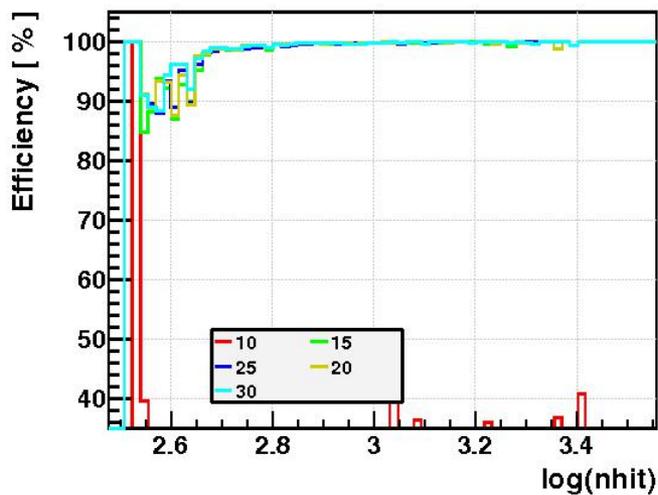
参数选择合适，速度可达80events/s

性价比



$$CP = \sqrt{\eta} * V$$

- η 平均效率
- V 平均速度
- $N_{\theta}=15$ 最优



芯位重建

◆重心法

- 简单直接
- 精度不够
- 不超出阵列范围

◆NKG函数拟合

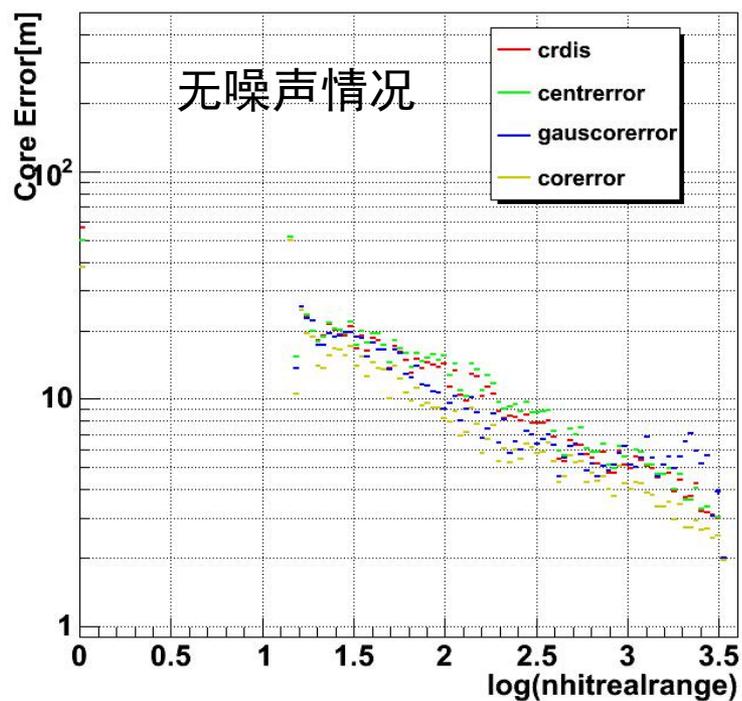
- 参数多
- 对簇射涨落敏感

◆2D-Gaussian

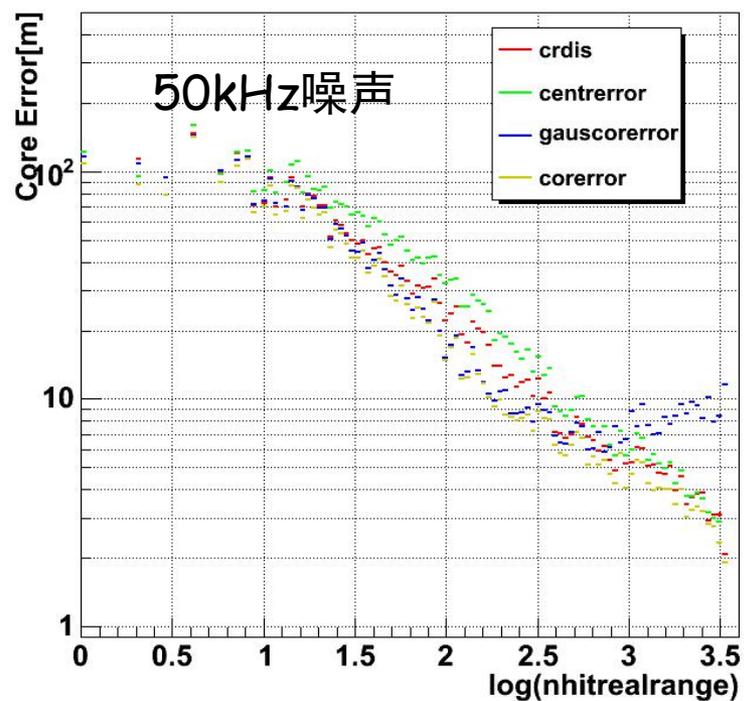
- 信息利用不充分（位置+电荷）
- 精度不够

初步芯位重建精度

R < 150 m



R < 150 m



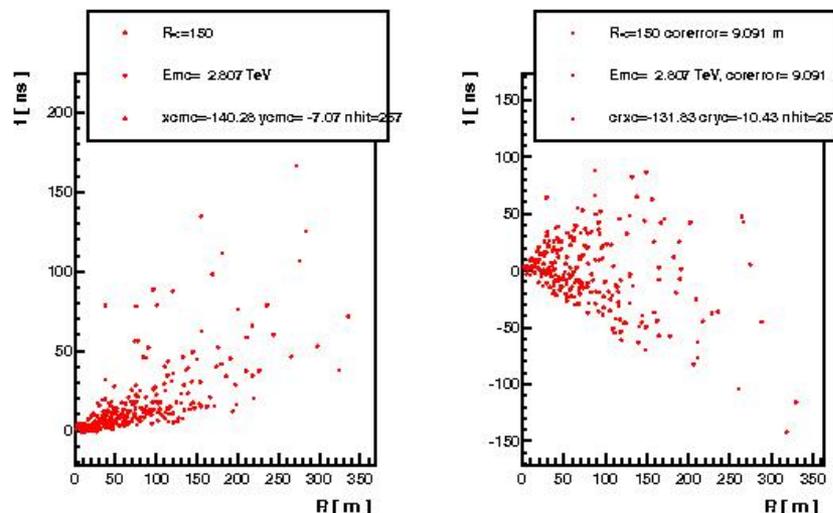
遍举法寻找芯位

◆ 步骤

- 地面上特定区域按照特定大小分为若干单元
- 循环，按照一定规则挑出最可几的单元坐标做为芯位
- 利用每个距离bin里面最早到达的hit信息

◆ 特点

- 不局限于阵列面积
- 充分利用 $q, t, \text{位置}$ 等信息
- 减弱噪声影响



距簇芯距离和时间关系图。分别为，真实芯位+真实方向，重建芯位+重建方向

总结

◆在线快速算法优化

- 参数优化
- 算法优化
- 过滤噪声、压缩数据量，保证真实信号保留比和速度

◆优化结果

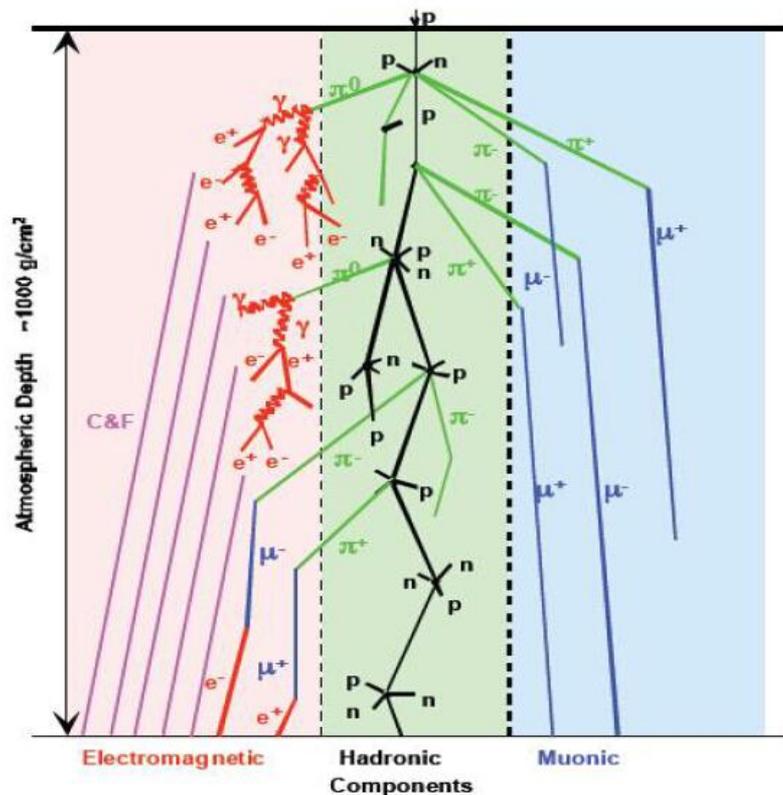
- $N_{\theta}=15$, 80event/s
- 真实信号保留比>80%情况下, 100GeV~30TeV的效率达到80%以上

◆芯位重建新方法尝试

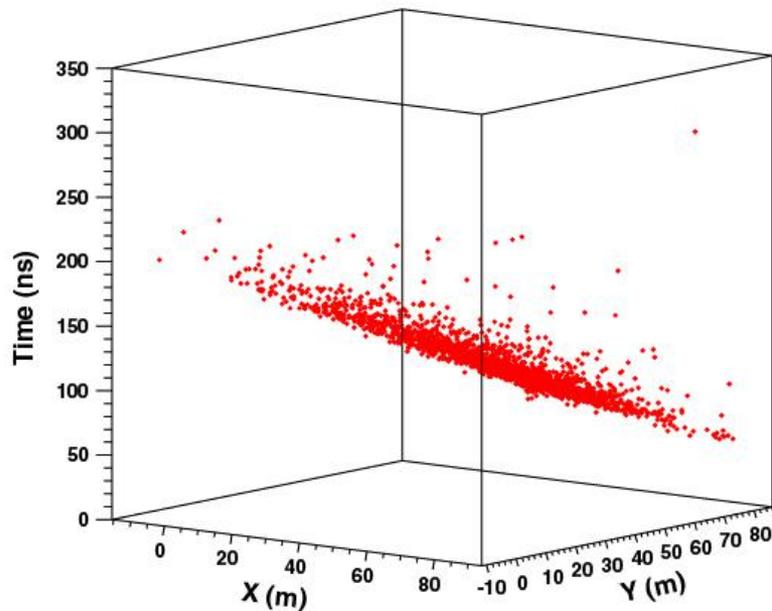
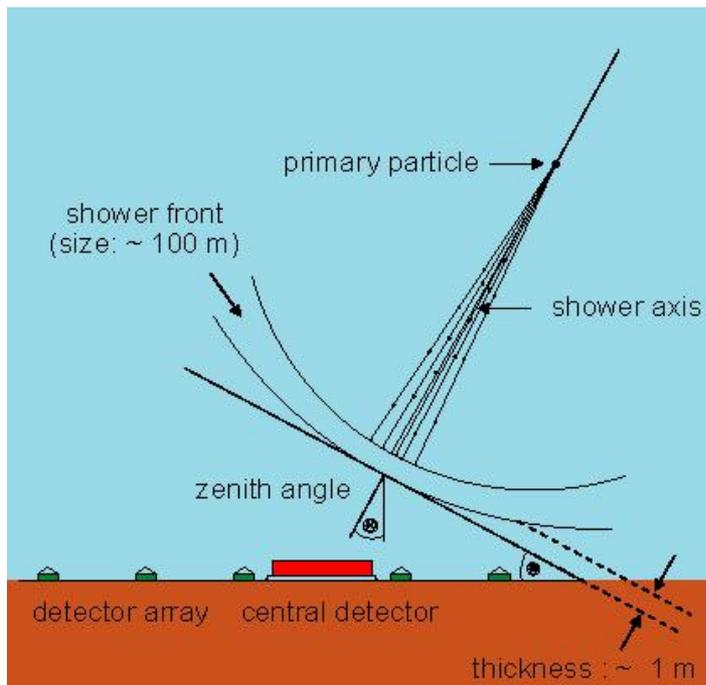
backup

高能 γ 天文

- 高能 γ 射线产生机制：
 - ✓ π 介子衰变；
 - ✓ 逆康普顿散射；
 - ✓ 轫致辐射；
- γ 天文意义：
 - ✓ 与宇宙线三大问题直接相关；
 - ✓ 作为探针研究天体的内部物理过程；
 - ✓ 新物理；
- 问题：
 - ✓ 甚高能 γ 射线源样本统计不足；
 - ✓ 对已探测到的源的能谱和源区精细观测不足；



方向重建----平面拟合



$$\chi^2 = \sum_i n_i (c \cdot (t_i - T_0) - x_i \cdot L_1 - y_i \cdot L_2)^2$$

T0: 参考时间，可以是第一个着火的时间

L1: x方向的方向矢量

L2: y方向的方向矢量

簇射轴垂直与此平面，平面的法线即为簇射轴的方向

方向重建----锥面拟合

- ◆ 因为距离芯位约源的粒子的到达时间相对与前峰面有所延迟，使得到达粒子的前封面不再是简单的平面，而呈现锥面
- ◆ 需要在平面拟合的基础上进行锥面拟合
- ◆ α 为锥面因子。在拟合中可以放开，作为一个拟合参数，也可以固定，根据模拟0.03时能够使伽马的角分辨最好

$$\chi^2 = \sum_i n_i (c \cdot (t_i - T_0) - x_i \cdot L_1 - y_i \cdot L_2 - c \cdot (\alpha R_i))^2$$

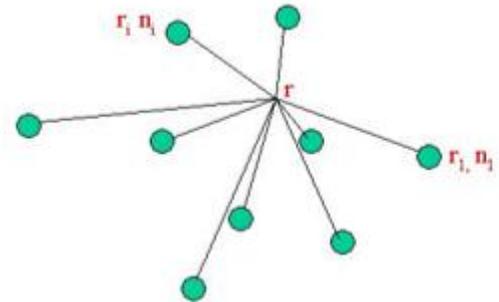
- ◆ 堆芯以 i 的相反任一处任反心以积一心位的是廷精度

芯位重建

- **N** = 给定事例的PAD击中数
- 对于某一击中 PAD i ($i=1, \dots, N$):

- ◆ $x_i, y_i \equiv$ PAD 的中心坐标
- ◆ $t_i \equiv$ 粒子到达时间
- ◆ $r_i^{(2)} \equiv (x_i, y_i)$
- ◆ $r_i^{(3)} \equiv (x_i, y_i, z_i)$ (with $z_i = t_i \cdot c_{\text{light}}$)
- ◆ $n_i \equiv$ 击中Strip数

✚ 一个二维r-tree 示例



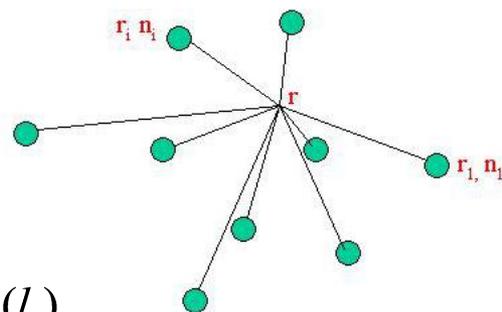
✚ r-tree 长度定义如下:

$$l(\mathbf{r}) = \sum_{i=1}^N d(\mathbf{r}, r_i) n_i$$

Tree&Dtree 簇芯重建法

◆重心法→Tree →Dtree

$$\frac{\sum_{i=1}^N r_i^{(2)} n_i}{\sum_{i=1}^N n_i} \longrightarrow \frac{\sum_{i=1}^N r_i^{(2)} w_i(l_i)}{\sum_{i=1}^N w_i(l_i)} \longrightarrow \frac{\sum_{i=1}^N r_i^{(2)} w_i(l_i)}{\sum_{i=1}^N w_i(l_i)}$$



◆ Tree : $l_i = l(r_i) , w_i = \frac{n_i}{l_i}$

◆ Dtree: $l_i = l(r_i) , P_M = \max\{n_i \cdot l(r_i)\}_{i=1,2,\dots,N}$

$$w_i = \left(1 - \frac{n_i \cdot l_i}{P_M}\right)$$

◆芯位：考虑到Tree长度时击中PAD位置的加权平均

芯位重建方法----极大似然法

- ◆ 簇射中次级粒子数目的横向分布可以用NGK函数来描述

$$\rho(N_e, r) = \frac{N_e}{R_M^2} \frac{\Gamma(4.5 - s)}{2\pi\Gamma(s)\Gamma(4.5 - 2s)} \left(\frac{r}{R_M}\right)^{s-2} \left(\frac{r}{R_M} + 1\right)^{s-4.5}$$

- ◆ 根据NGK函数可以估计出距离芯位R处的探测器接受到的粒子数，假设为 μ ，
- ◆ 而实际观测到的粒子数m是符合泊松涨落的，实际观测到m个粒子的概率为

$$p(m) = \frac{\mu^m}{m!} e^{-\mu}$$

- ◆ 假设探测观测到>1个粒子着火，0个粒子不着火，则探测器着火和不着火的概率为：

$$\bullet p_k(OFF) \equiv p_k(0) = e^{-\mu_k}$$

- ◆ 似然函数为：

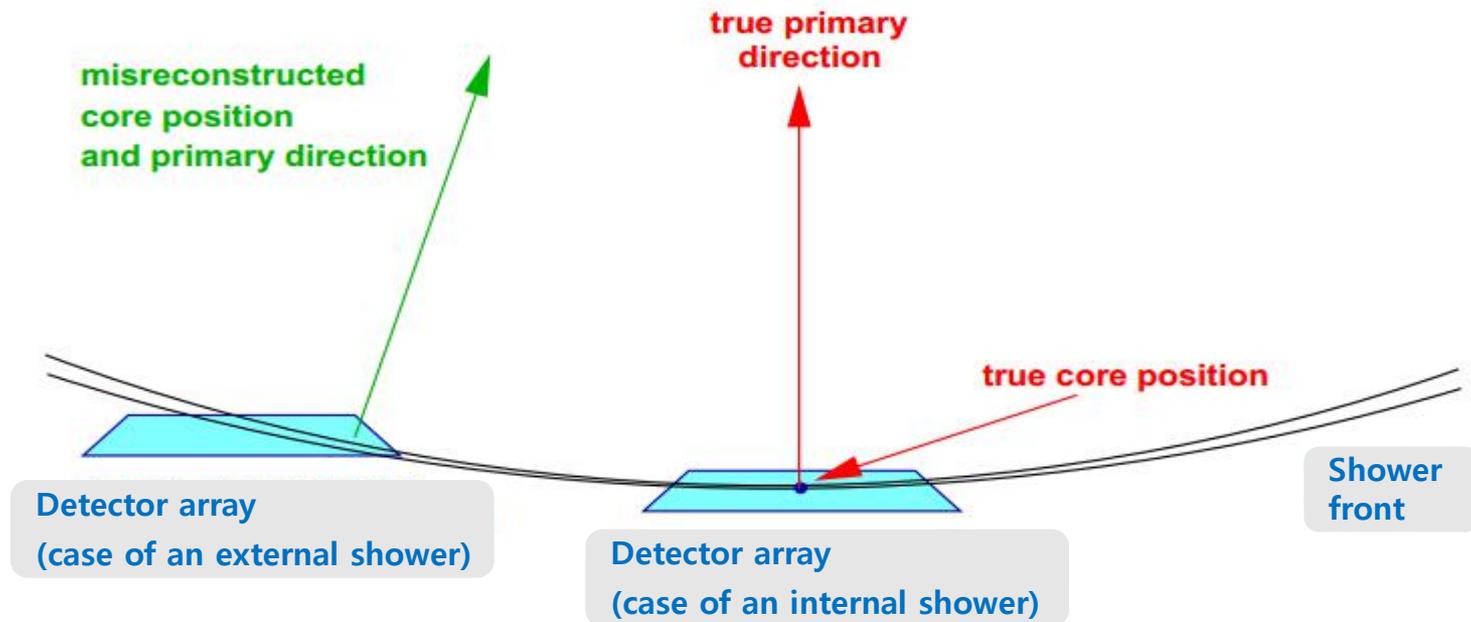
$$\bullet p_k(ON) \equiv p_k(> 0) = 1 - e^{-\mu_k}$$

- ◆ 重建精度取决与探测单元的大小

- 以ARGO为例，只有以strip条为最小探测单元时重建精度才最好，但是速度交慢

$$LF1 = \prod_{i=1}^{N_S^{OFF}} p_i(OFF) \prod_{j=1}^{N_S^{ON}} p_j(ON)$$

簇射芯位重建



- ◆ 重心法；
- ◆ 树形分析法：MTree法、Tree法、Dtree法；
- ◆ NKG函数拟合（利用其对簇射横向发展的分布描述间接得到簇芯位置）；
- ◆ 2D-Gaussian 拟合

2D-Gaussian fitting

- ◆ 1.setup map of outrigger grid to quadrant(8 quads,2*4 overlaping)
- ◆ 2.zero out covariance matrix prior to doing the fit; zero out hit map
- ◆ 3.check which layer to use
 - ON POND: caldata->nOUT<5
 - OFF POND: caldata->Nout>20||ratioOutToTop>0.1
&&(fabs(xOUT)>3000||fabs(yOUT)>3000)
- ◆ 4.zero out pe & pe weighted sums
- ◆ 5.collect up hits in used layers & sum up pes and pe weights posions
- ◆ 6.looping to find out the best parameter with the least chisquare

looping

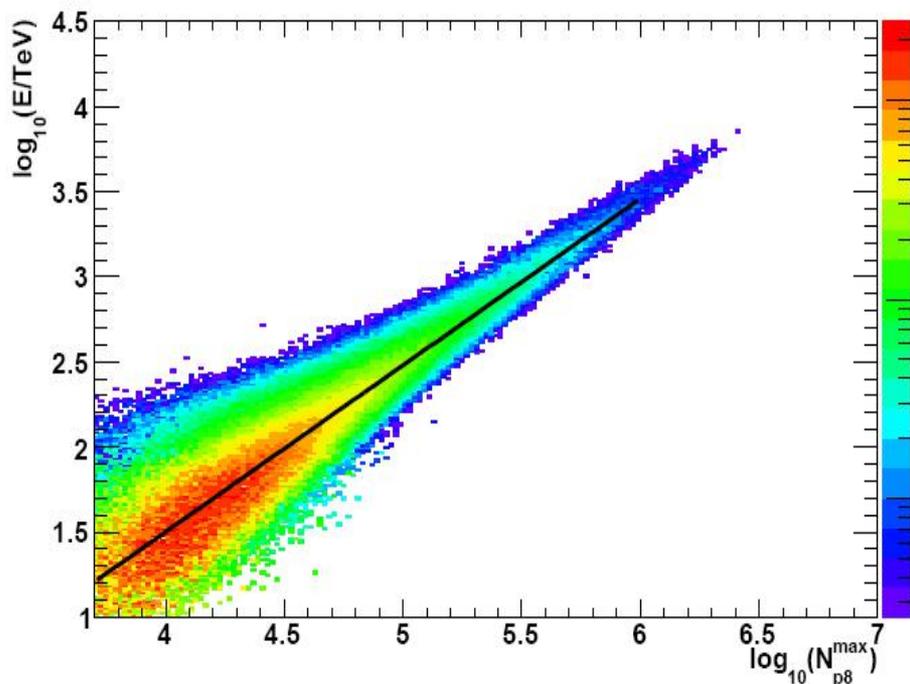
- ◆ `alambda = -1.0;`
- ◆ `Success = mrqmin(x,y,pe,sig,nhit,parms,covar,alpha,&chisq,&alambda);`
- ◆ `if (! Success) {一系列赋值, xfit, yfit, etc. return 1;}`
- ◆ `coreInfo::xCoreFPass[coreInfo::nPass] = parms[0];`
- ◆ `coreInfo::yCoreFPass[coreInfo::nPass] = parms[1];`
- ◆ `// Minimize chi-squared, or stop when no improvement found`
- ◆ `do { ochisq = chisq;`
- ◆ `Success = mrqmin(x,y,pe,sig,nhit,parms,covar,alpha,&chisq,&alambda);`
- ◆ `if (! Success) {一系列赋值。。。 return 1; }`
- ◆ `if (coreInfo::nPass < MAX_GAUSS_CORE_PASS) {`
- ◆ `coreInfo::xCoreFPass[coreInfo::nPass] = parms[0];`
`coreInfo::yCoreFPass[coreInfo::nPass] = parms[1];`
- ◆ `coreInfo::CoreErrPass[coreInfo::nPass] = 0; coreInfo::nPass++; }`
- ◆ `else { coreInfo::nPass++; // keep track of nPass even if don't fill further info. }`
- ◆ `} while (((chisq >= ochisq) || (fabs(ochisq - chisq) > 1.0)) && (fabs(log10(alambda)) < 30.0));`

looping

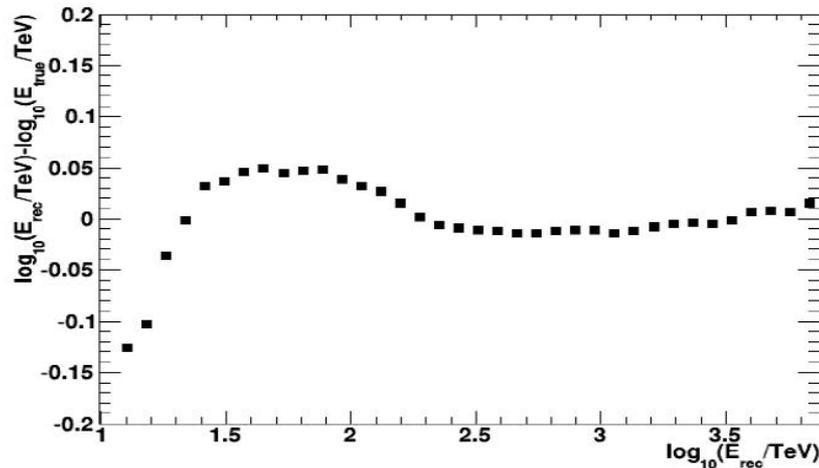
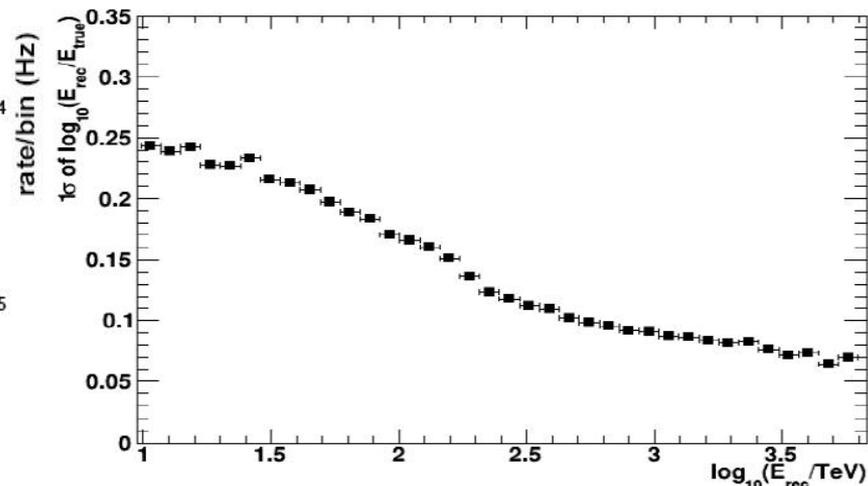
- ◆ `mrqmin(x,y,pe,sig,nhit,parms,covar,alpha,&chisq,&alambda);`
- ◆ `int mrqmin(float x[], float y[], float z[], float sig[], int ndata, double a[4], double covar[4][4], double alpha[4][4], double *chisq, double *alambda)`
- ◆ Levenberg-Marquardt Method: 给定一系列初始值 $x[1..ndata]$, $y[1..ndata]$, 有标准偏差 $sig[1..ndata]$ 的 $z[1..ndata]$ 和有 ma 个系数 $a[1..ma]$ 的一个非线性函数, 求这一系列初始值和非线性函数拟合值之间的最小 $chi-sq$ 值。初始 $ia[1..ma]$ 由应该被拟合的非零事例和一些应该被修正的零事例得到。改程序会返回当前能得到对于 $a[1..ma]$ 和 $chisq$ 最好的拟合值。Covar 和 $alpha$ 在迭代时充当存储空间。程序 `funcs(x,y,a,zfit,dzda,ma)` 用来计算拟合函数 $yfit$ 和它的导数 $dyda[1..ma]$ 关于 (x,y) 处的拟合参数 a , 初始化时设置 $alambda < 0$ (which then sets $alambda = 0.001$). 如果成功, $chisq$ 变小, 则 $alambda$ 除以 10, 如果失败, $alambda$ 乘以 10, 重复调用此函数直至收敛。
- ◆ Then make one final call with $alambda=0$, so that `covar[1..ma][1..ma]` returns the covariance matrix, and `alpha` the curvature matrix. (Parameters held fixed will return zero covariances)

地面粒子阵列能量重建

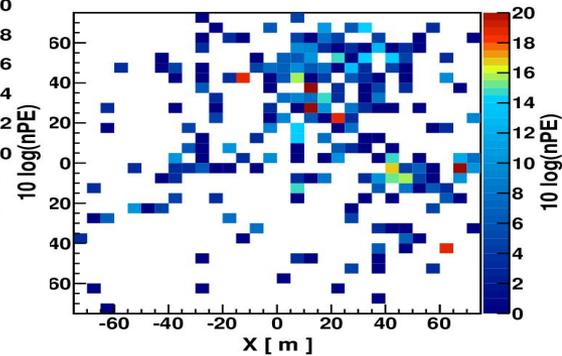
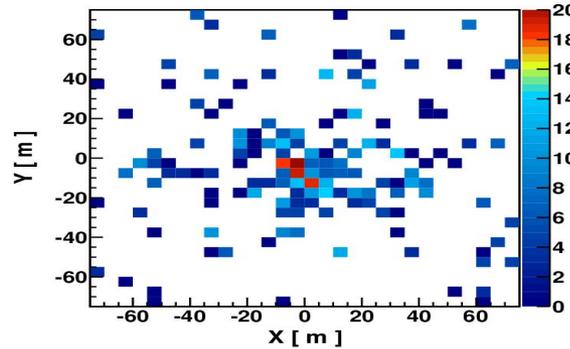
- ◆ 利用所观测到粒子数和能量之间的关系



$$\text{Log}(E/\text{TeV}) = a \cdot \text{Log}(N_{p8}^{\max}) + b$$



伽马质子区分

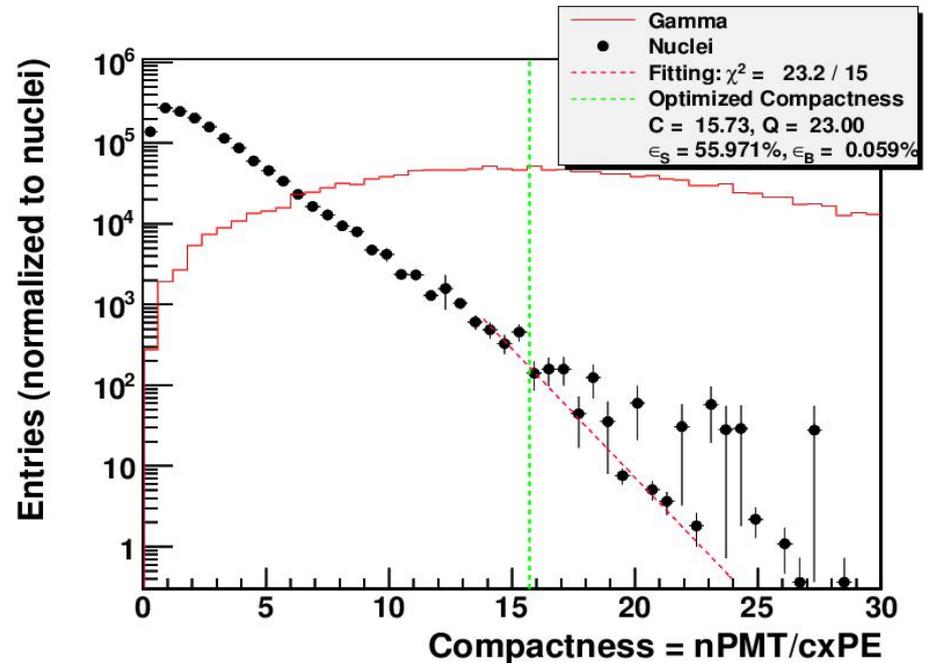


- ◆ 为提高阵列对 γ 射线的灵敏度，应尽量排除强子（质子）背景。
- ◆ 我们定义

$$\text{compactness} = \text{nPMT} / \text{cxPE}$$

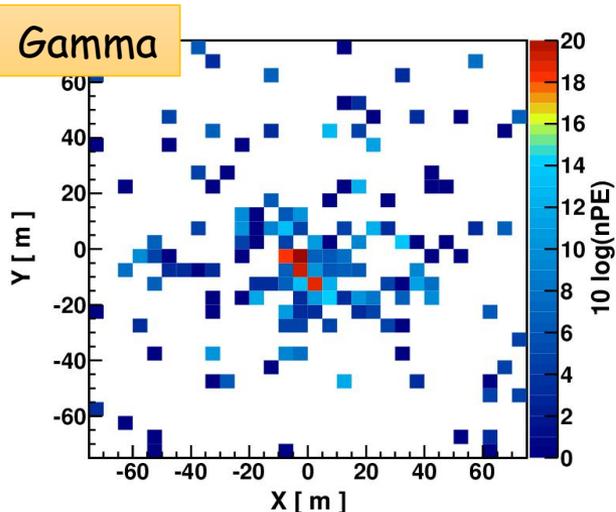
其中cxPE是指重建芯位45m以外，单个PMT上探测到的最多的PE数，nPMT指簇射触发的总PMT数目。

- ◆ 对于给定的nPMT值，compactness越大，越趋向于gamma-like簇射。
- ◆ 缺点：适用于低能段，高能 γ 簇射在45m外也会沉积大量能量

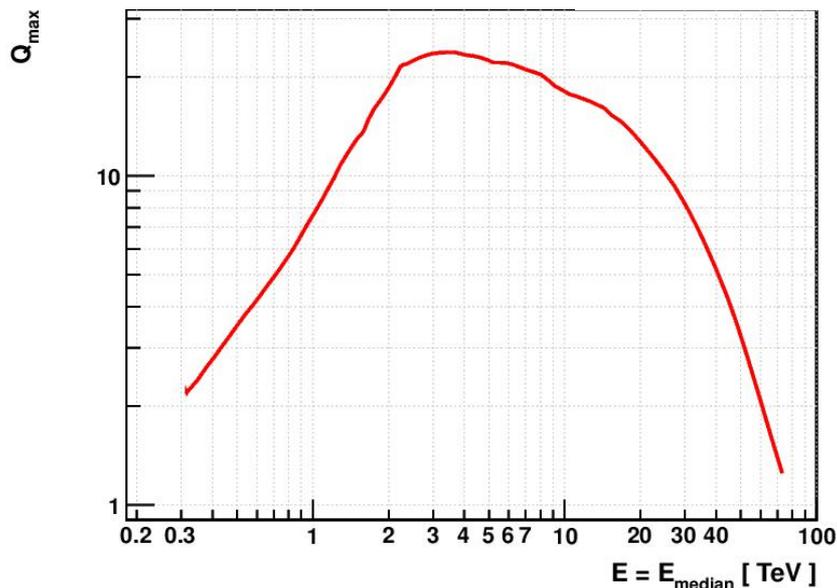
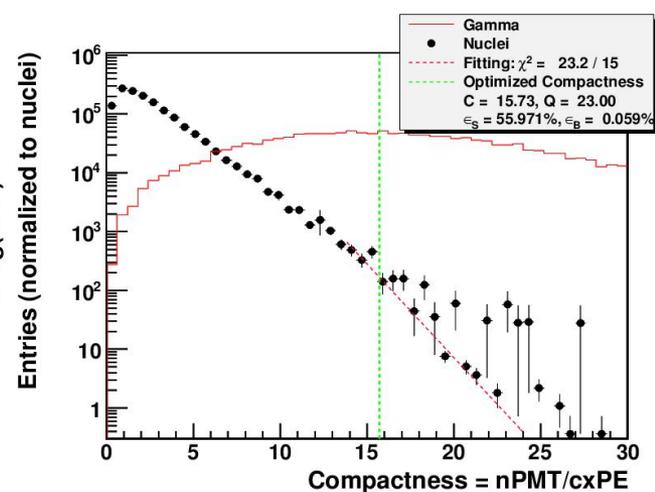
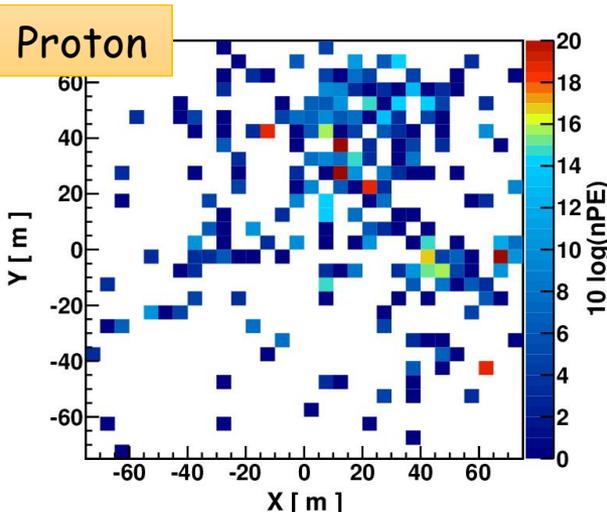


质子伽马鉴别- WCDA

WCDA 150×150 m² | Gamma, E = 1 TeV | nPMT = 142



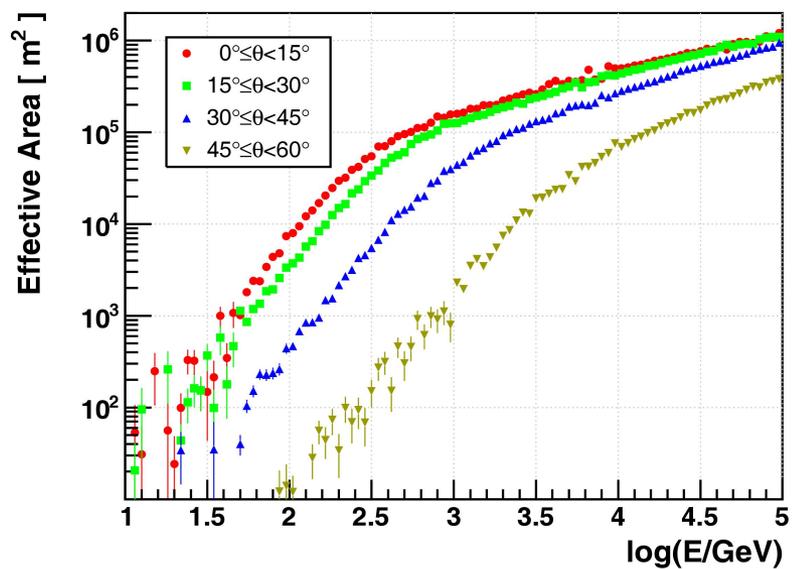
WCDA 150×150 m² | Proton, E = 2 TeV | nPMT = 212



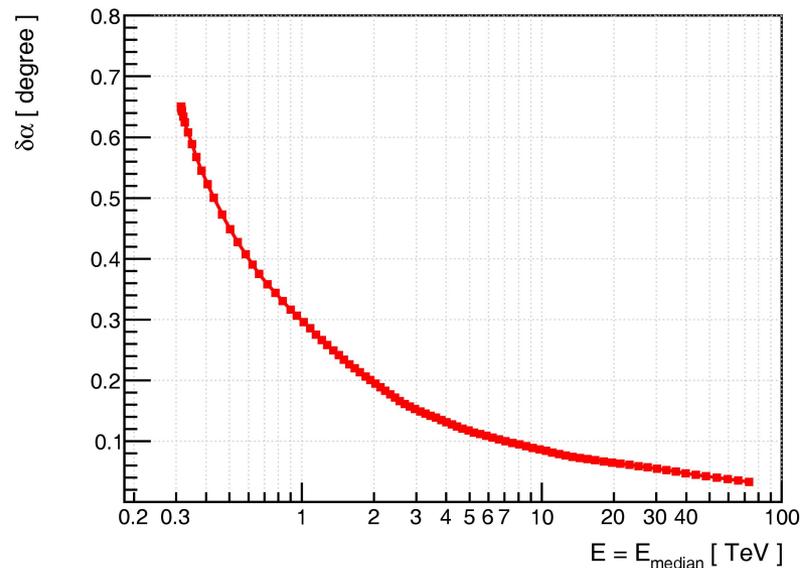
◆ 利用质子shower中的次芯:

- Signal of the brightest PMT outside the shower core region (e.g., 45 m);
- 着火PMT数目与45米外最亮PMT中PE数的比值

WCDA性能参数



Gamma的有效面积



Angle-resolution

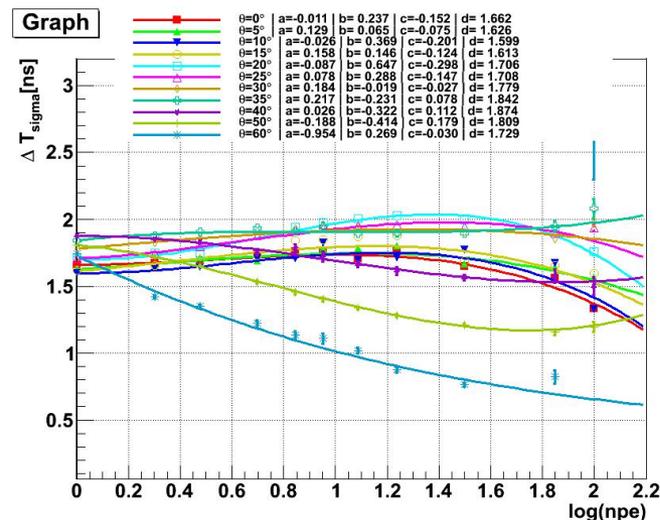
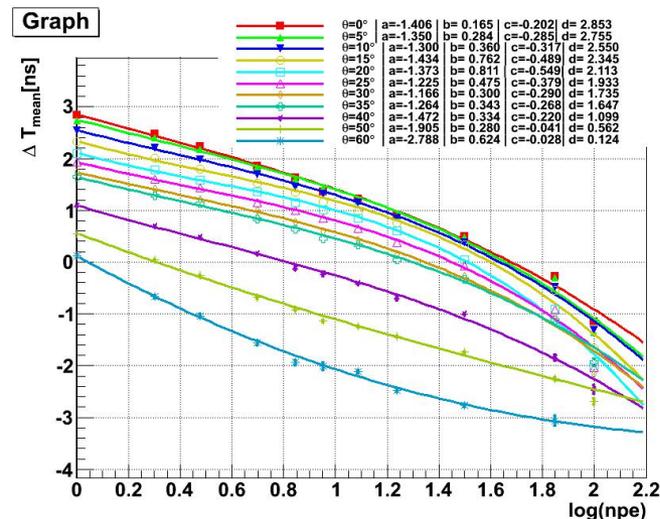
时间误差simulation

◆影响探测器单元接受信号时间信息的多种因素:

- 簇射前锋面的厚度及其涨落
- 粒子的击中位置（重建时我们认为是单元中心）
- 入射角度
- 粒子数的多少

◆全阵列时间模拟

- t_0 =探测器中心处粒子的到达时间
- $\Delta t = t_i - t_0$ 的均值和方差分布如右图



小结

◆ 优化总结

■ 参数优化选择

- X_σ : 三次多项式
- N_θ : 15
- Sort window size : 800ns (-570,230)
- Keep window size : 200ns (-65,135)

■ 算法优化选择

- 小角度事例再重建与分天区次序重建相结合

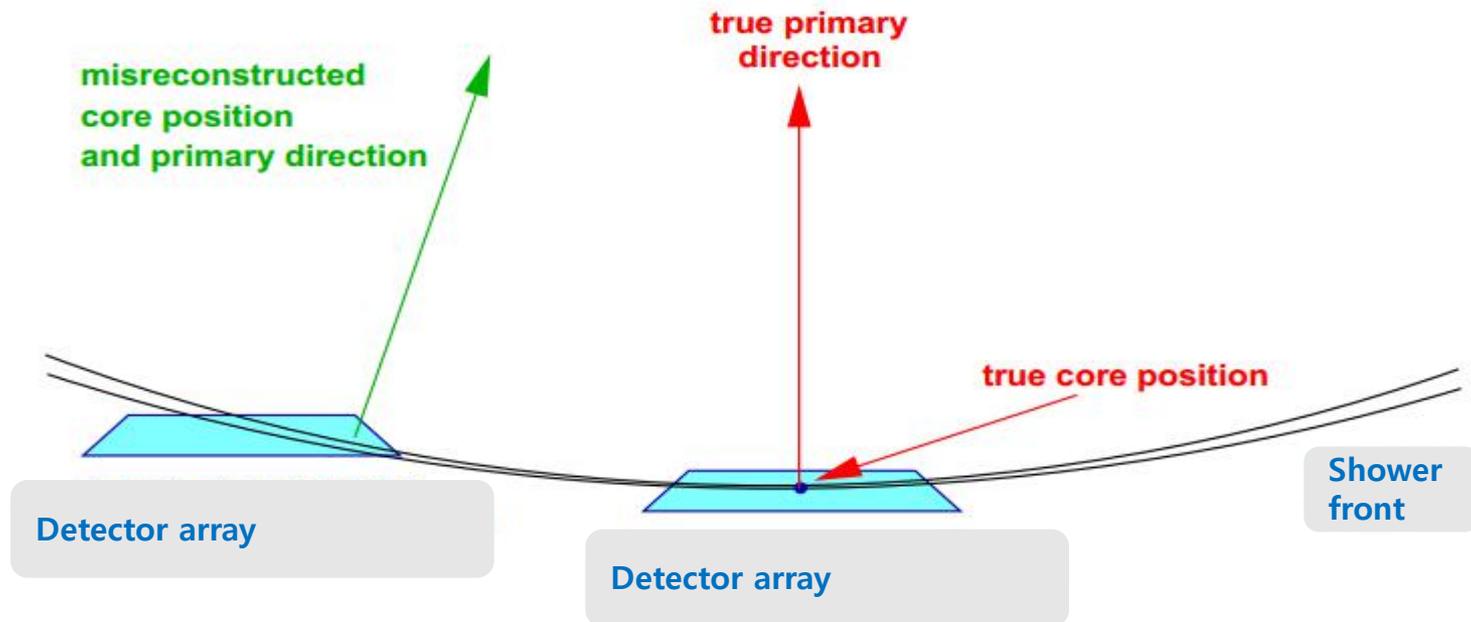
■ 初步优化结果

- $N_\theta=15$,积分速度可以达到80event/s
- 以真实信号保留比>80%作为计算重建效率的cut条件, 100GeV~30TeV的效率达到80%以上

◆ 预重建算法优化主要包括了参数优化和算法优化两个部分, 目前主要工作完成, 程序效率和速度都有较大提高。

◆ 下一步工作对经预重建效率筛选的数据重建, 检验效果。

簇射芯位重建



- ◆ 重心法；
- ◆ 树形分析法：MTree法、Tree法、Dtree法；
- ◆ NKG函数拟合（利用其对簇射横向发展的分布描述间接得到簇芯位置）；
- ◆ 2D-Gaussian 拟合