



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

粒子物理实验 数据处理

孙 胜 森

sunss@ihep.ac.cn

第三届高能物理计算暑期学校

2022年8月17日





STANDARD MODEL OF ELEMENTARY PARTICLES



BES III



• 标准模型

✓ 电弱统一理论

实验对电弱理论有很多精确的测量和检验；2012年 Higgs粒子的发现标志着电弱理论的巨大成功

✓ 量子色动力学

高能下的QCD得到实验的大量检验，QCD的渐进自由

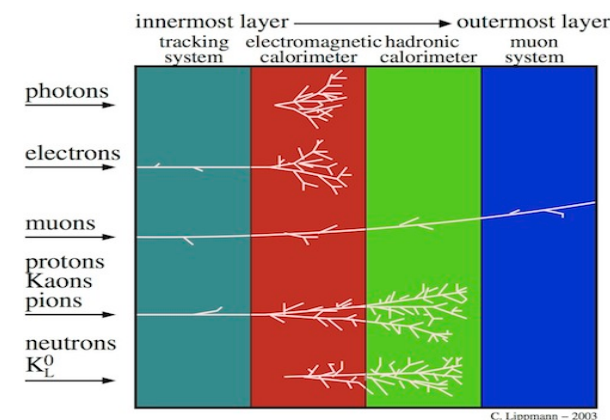
• 标准模型不能回答的问题

- 电弱对称性破坏和电弱相变
- 味对称和味结构
- 质量等级
- 暗物质
- 暗能量
- 反物质不对称
- 宇宙膨胀
- 引力的量子化和量子属性

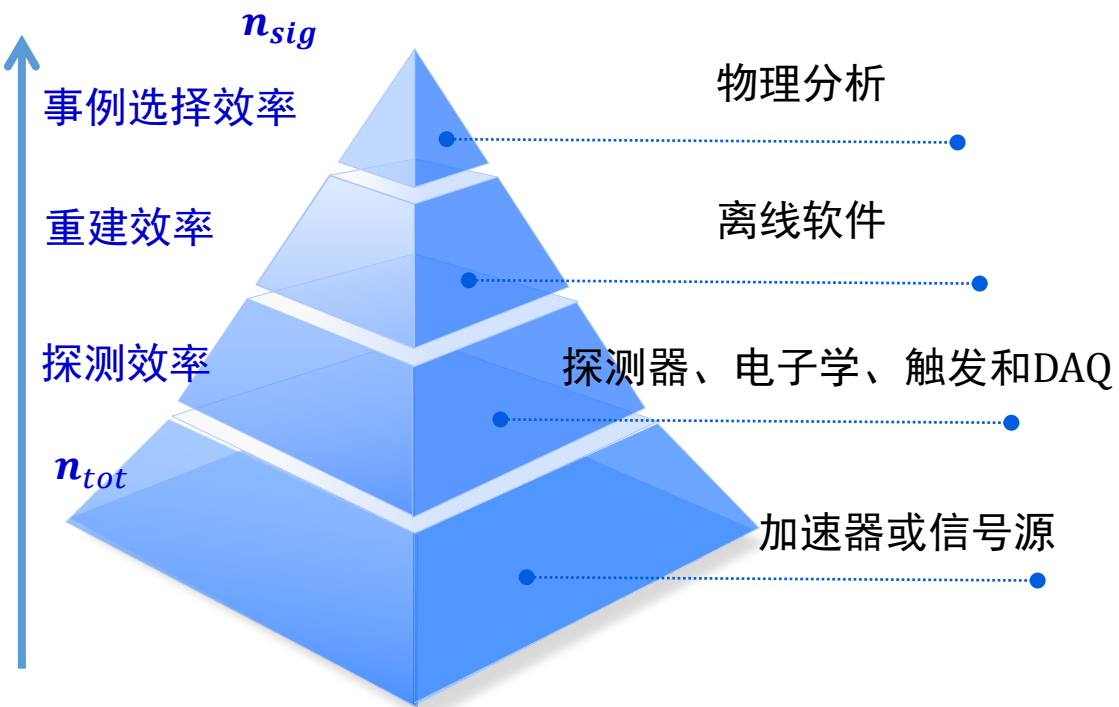
标准模型不是终极理论

粒子物理实验

- 非常简单粗暴地划分：**理论** — 动力学；**实验** — 运动学
- 粒子物理实验的研究内容：
 - 粒子的产生：截面、产生机制
 - 粒子的性质：质量、寿命、共振宽度、自旋、宇称，……
 - 相互作用的物理规律：衰减常数、衰变分支比、能谱、形状因子、物理量的守恒与破坏，……
 - 寻找新粒子，新物理现象
- 测量对象：带电粒子 e, μ, π, K, p 中性粒子 γ, K_L, ν, n 的四动量 (p_x, p_y, p_z, E)
- 微观粒子的探测是通过**粒子与宏观物质的相互作用**实现探测的
 - 带电粒子与介子的分子，原子及核的碰撞
 - ✓ 电离能损
 - ✓ 多次散射
 - ✓ 切伦科夫辐射
 - ✓ 穿越辐射
 - ✓ 韧致辐射
 - ✓ 正负电子湮灭
 - 光子与探测器介质的作用
 - ✓ 光电效应
 - ✓ 康普顿散射
 - ✓ 电子对产生
- 粒子与探测器相互作用输出的电信号，经过电子学提取电荷、时间等信息，形成原始数据



数据产生、获取、处理和分析



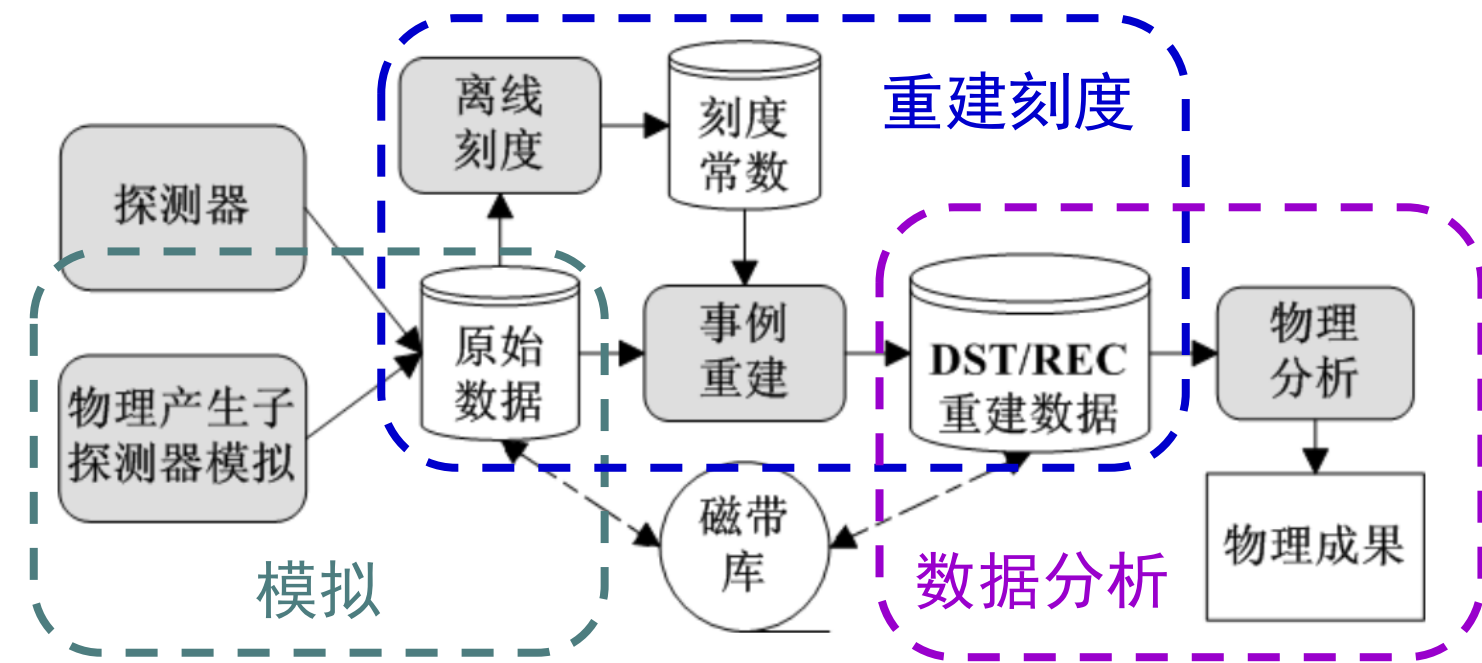
物理事例的产生、获取、处理和分析

- 高亮度的加速器或信号源 → 高统计量的实验数据 n_{tot}
- 高精度的探测器
 - 探测效率 / 探测器分辨
 - 物理过程：电离能损、多次散射、康普顿效应等等
 - 束流本底与电子学噪声
- 数据处理
 - 重建效率 / 重建算法性能
- 物理分析
 - 事例选择效率： n_{sig} 统计误差
 - （质量）分辨：信号显著性

$$BR = \frac{n_{sig}}{n_{tot} \cdot BR_{inter} \cdot \epsilon}$$

原始数据记录的电子学信号不能直接用于物理分析，需要在计算平台通过数据处理形成重建数据
计算与软件是粒子物理实验的重要环节！

数据处理



数据处理和物理分析流程

- 模拟: signal MC / Inclusive MC
探测器设计 / 事例选择效率 / 本底分析
- 重建: 电子学信号 → 物理量
- 数据分析: Real Data / MC
研究物理过程与规律

- 软件框架提供有效的数据管理工具，不同软件模块的组合和动态库的链接机制
- 通过刻度与重建，压低各种条件的影响和排除噪声本底，最大限度挖掘加速器和探测器性能
- 实现精确的探测器模拟，为事例选择效率计算、选择条件优化和本底估计等提供可靠依据
- 利用优秀数理方法联合各种实验信息开发物理分析工具软件，进一步提高实验精度

离线软件的构成

• 基础软件库和工具

- 软件开发通用库和工具
- 高能物理领域通用的流行软件：Geant4, ROOT等

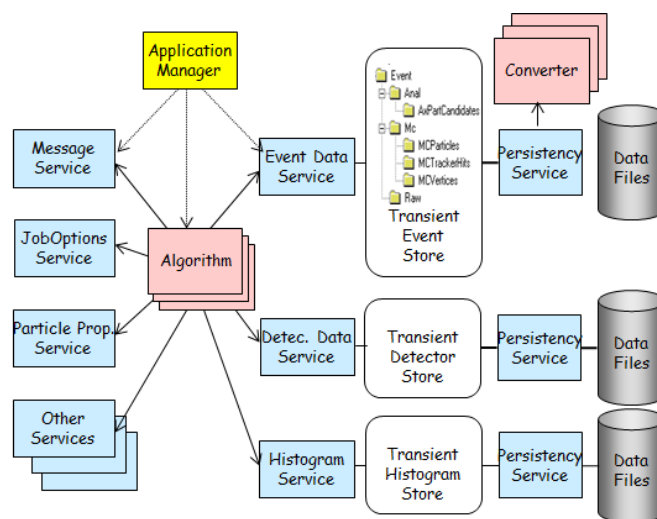
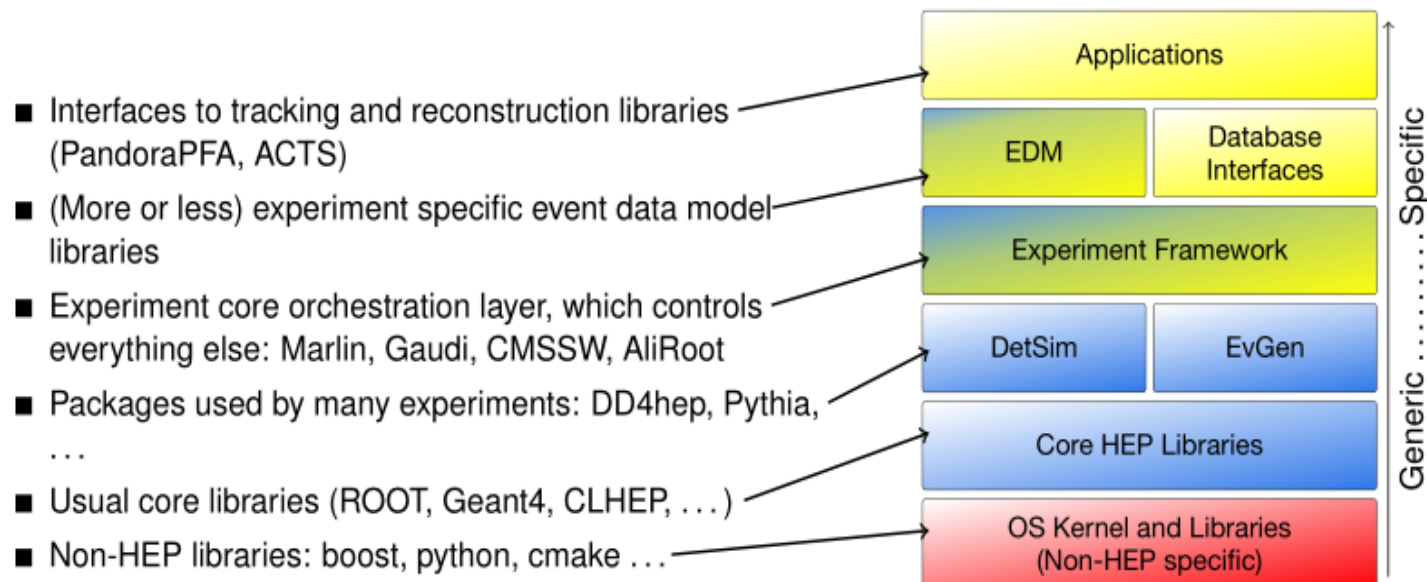
• 框架软件

- 狭义上指通用的离线软件框架，与实验无关，如Gaudi、SNiPER等
- 通常也包括与实验相关的通用功能模块，如数据模型及其I/O、数据库接口等

• 数据处理和分析软件

- 与实验紧密相关，需要有物理背景人员开发
- 专注于对数据的计算（算法），不用关心底层软件技术细节

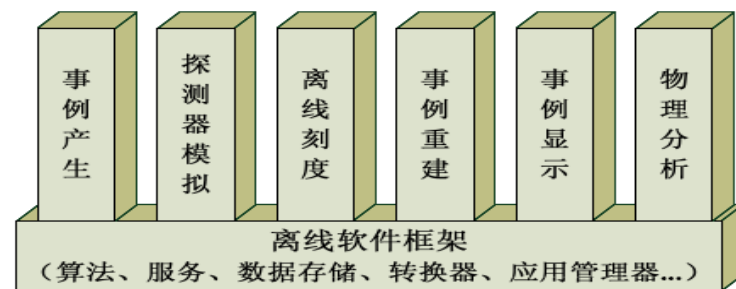
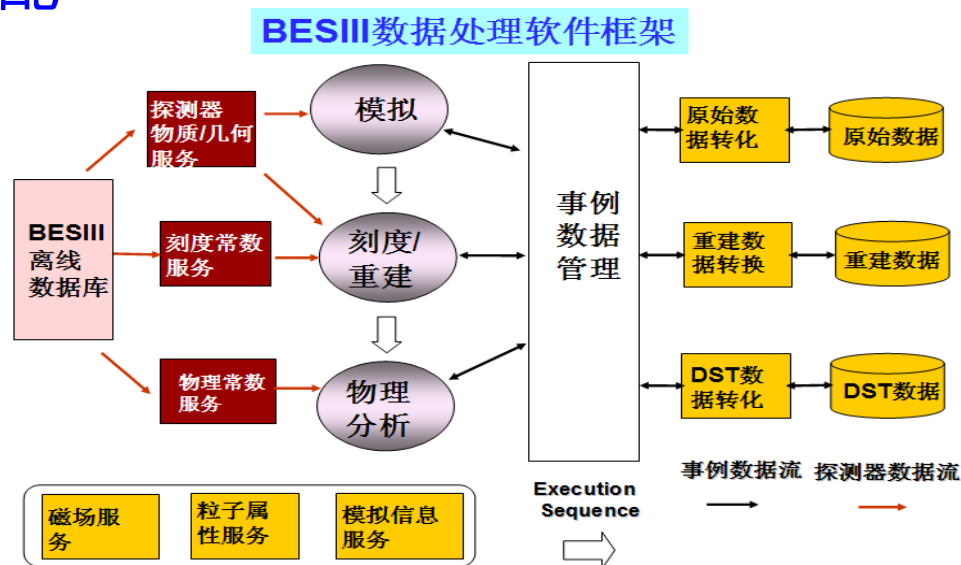
Applications usually rely on large number of libraries, where some depend on others



- GAUDI为应用程序提供了软件框架，开发人员将代码植入框架中，同时提供通用功能
- 算法（多变）和数据（稳定）分离，算法以插件形式动态嵌入框架执行
- 瞬态数据（内存）与持久数据（磁盘）分离

BESIII实验离线软件

- BOSS采用通用框架软件GAUDI为基础的离线数据处理软件系统
- 根据BESIII实验的需求，设计了 Event Data Model, Event Management, File I/O System, Services,
- 粒子物理实验国际通用软件库和工具软件的支持：Geant4, ROOT, MySQL, CERNLIB, CLHEP,
- 提供简单易用、安全可靠、服务齐全的数据处理环境，高效的数据管理和灵活的软件调配



探测器刻度

- 探测器固有性能和刻度算法水平是决定物理测量结果精度/信号显著性的最主要贡献之一
- 通过刻度算法把探测器的优良的空间分辨，时间分辨等转化为物理分析中的动量/能量分辨，粒子鉴别能力等性能指标
- BESIII各个子探测器性能达到或超过设计指标；达到同类型探测器国际先进水平，部分指标处于国际领先
- 探索利用先进数理方法，如深度学习等进一步优化探测器性能

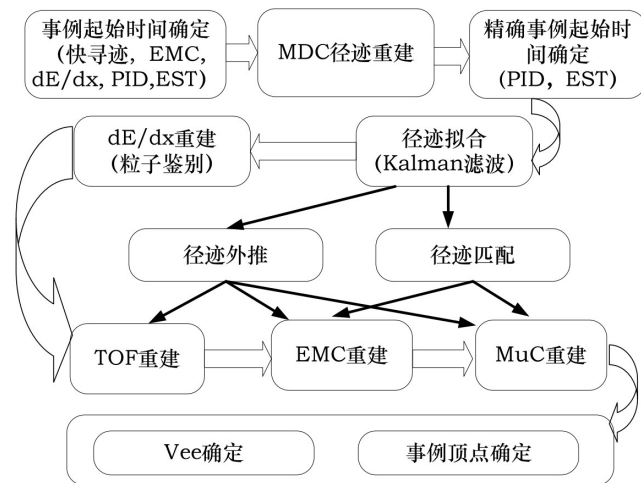
Exps.	MDC Spatial resolution	MDC dE/dx resolution	EMC Energy resolution
CLEOc	110 μm	5%	2.2-2.4 %
Babar	125 μm	7%	2.67 %
Belle	130 μm	5.6%	2.2 %
BESIII	115 μm	<5% (Bhabha)	2.4%

Exps.	TOF Time resolution
CDFII	100 ps
Belle	90 ps
BESIII	68 ps (BTOF) 60 ps (ETOF)

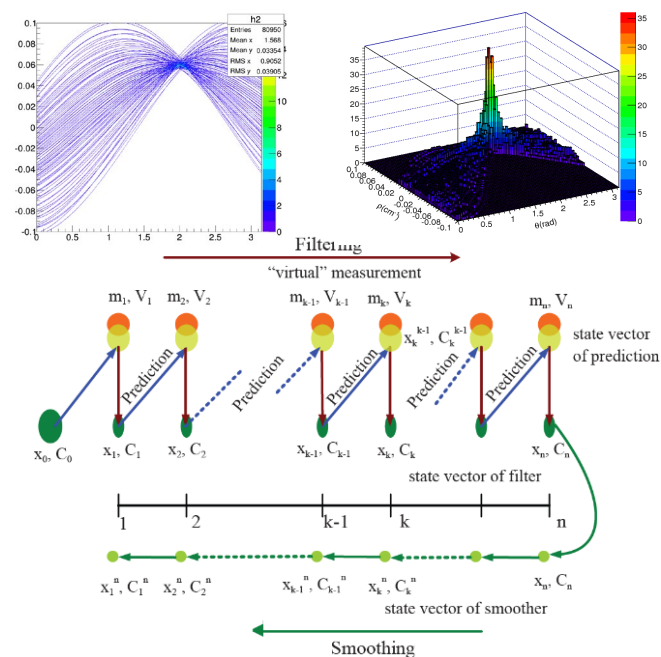
MUC: Efficiency $\sim 96\%$
BG level:
< 0.04 Hz/cm²(B-MUC),
< 0.1 Hz/cm²(E-MUC)

事例重建

- 事例重建软件将原始数据中记录的电子学信号转化为粒子的动量、能量等物理量，生成重建数据，供物理课题研究使用
- BESIII事例重建：
 - 快寻迹和事例起始时间
 - 漂移室带电径迹重建
 - Kalman滤波径迹拟合
 - dE/dx径迹重建
 - EMC、TOF、MUC 重建和径迹外推与匹配
- 开发和应用基于模式匹配、共形变换、霍夫变换的径迹寻找算法，采用卡尔曼（Kalman）滤波的径迹拟合算法，实现精确计算带电径迹参数和误差矩阵的目标
- 加速器和探测器的噪声本底是限制重建效率和精度的最关键因素，探索利用探测器自身特点和信号噪声特征，调试与优化噪声本底排除机制是未来主要研究方向之一



BESIII事例重建流程



带电径迹重建

• 寻迹（模式识别）

测量点与径迹的对应关系

- 局域方法：如径迹跟踪、径迹路径、径迹元等
- 全局方法：如模式匹配，霍夫变换方法等

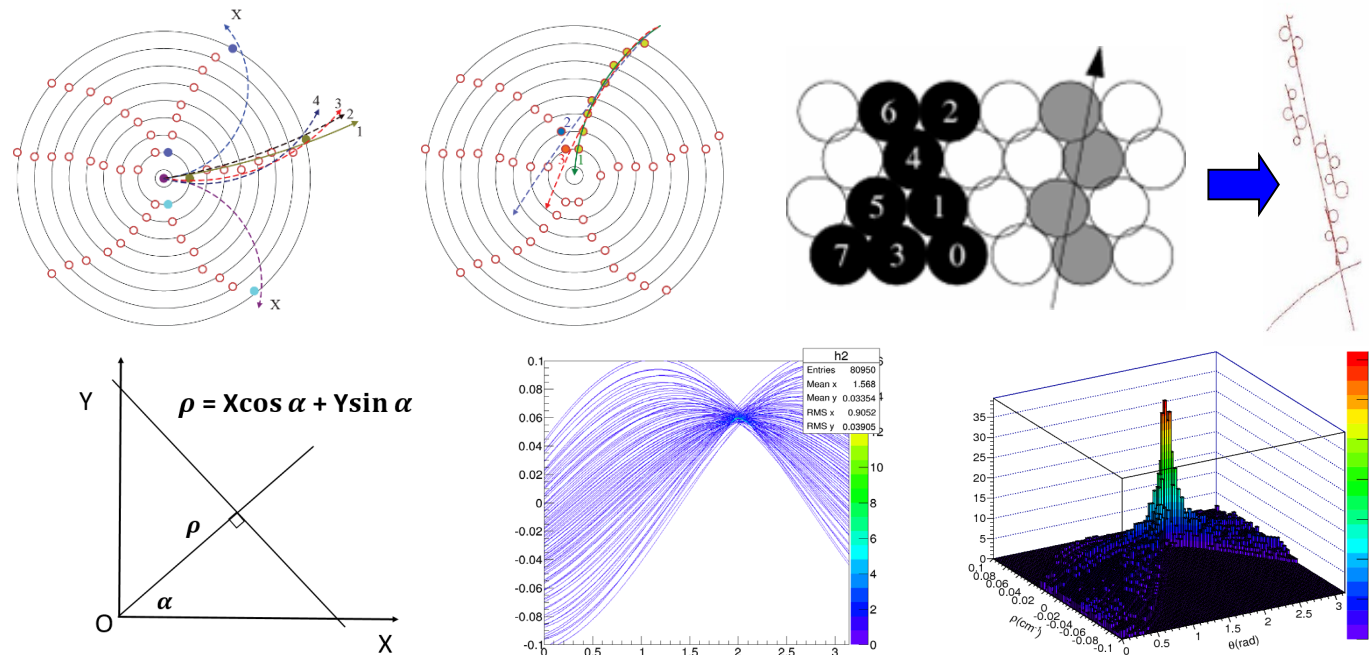
• 径迹拟合（参数估计）

径迹参数 + 误差矩阵

- 最小二乘法拟合
- Kalman 滤波方法

• 统计检验

$$\chi^2 = \sum_{i=1}^n \frac{(d_{mea} - d_{track})^2}{\sigma^2}$$



径迹参数（螺旋线）： d_ρ , d_z , κ , ϕ_0 , λ

$$\begin{aligned} x &= x_0 + d_\rho \cos \phi_0 + \frac{\alpha}{\kappa} (\cos \phi_0 - \cos(\phi_0 + \phi)) \\ y &= y_0 + d_\rho \sin \phi_0 + \frac{\alpha}{\kappa} (\sin \phi_0 - \sin(\phi_0 + \phi)) \\ z &= z_0 + d_z - \frac{\alpha}{\kappa} \lambda \cdot \phi \end{aligned}$$

$$\begin{aligned} p_x &= -\frac{1}{\kappa} \sin(\phi_0 + \phi) \\ p_y &= \frac{1}{\kappa} \cos(\phi_0 + \phi) \\ p_z &= -\frac{1}{|\kappa|} \lambda \end{aligned}$$

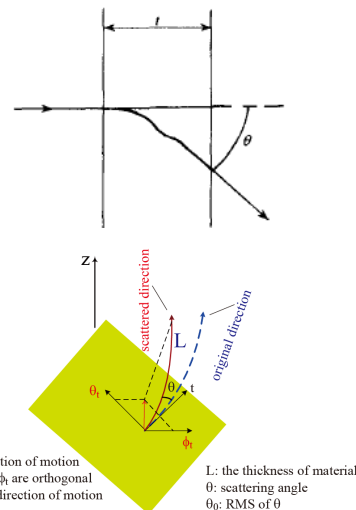
带电径迹重建

- Kalman Filter 方法由一系列递归数学公式描述
 - 高效的可计算的方法来估计过程的状态，可以估计信号的过去和当前的状态，甚至能估计将来的状态，及时不知道模型的确切性质
 - 应用广泛，功能强大。可用于粒子物理实验中参数的优化估计，如径迹重建，顶点重建和运动学拟合等领域
- 理想的空间螺旋线拟合，没有考虑到：
 - 多次散射
 - 能量损失（束流管，漂移室内壁等物质效应）
 - 磁场的不均匀性
- Kalman Filter 方法包括径迹参数和误差矩阵的改进

多次库伦散射

带电粒子通过介质时与原子的电子发生相互作用，当入射带电粒子与介质原子最接近距离比原子半径小时，入射粒子将受到原子核电场的相互作用，入射粒子的轨迹将受到核电场的偏转，这种过程称为介质原子核的多次库伦散射。

$$\theta_{RMS}^{proj} = \sqrt{\langle \theta^2 \rangle} = \frac{13.6 \text{ MeV}}{\beta c p} \cdot z \cdot \sqrt{x/X_0} \cdot [1 + 0.038 \cdot \ln(x/X_0)]$$



能量损失

带电粒子通过介质时与介质原子发生相互作用，在相互作用中入射粒子将部分能量传递给核外电子，使介质原子产生电离或激发而损失能量。

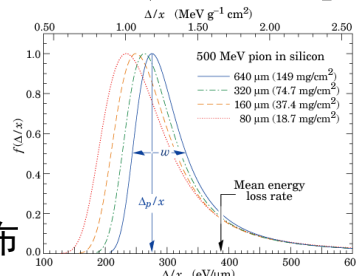
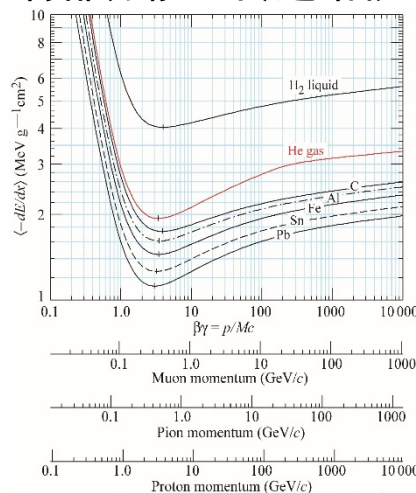
$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\ln \left(\frac{2m_e c^2 \beta^2 \gamma^2}{I} \right) - \beta^2 - \frac{\delta}{2} \right]$$

低能区：~1/β²

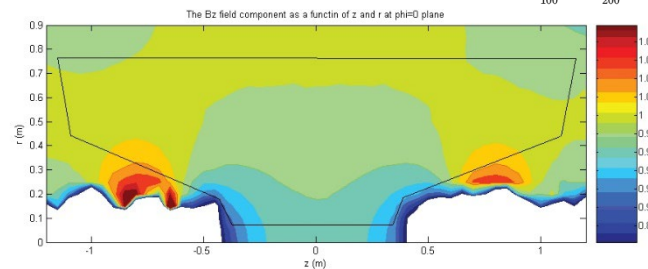
最小电离粒子（MIP）：βγ~3.2

相对论上升和密度效应

电离能损围绕平均值有统计涨落，穿过薄介质层时，Landau(朗道)分布

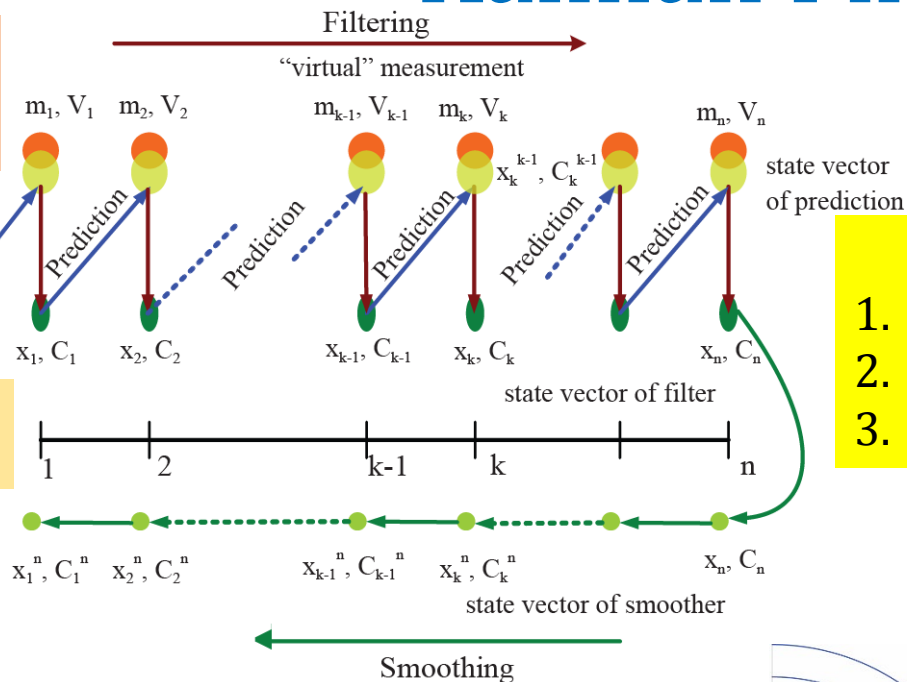


磁场的不均匀性



Kalman Filter径迹拟合

被估计的信号:
State Vector
状态矢量



Measured Vector
测量矢量

- 三个步骤
1. 预言 Prediction
 2. 过滤 Filtering
 3. 平滑 Smoothing

系统传输方程:

把上一时刻的状态量影射到当前时刻的状态量

$$x_k = F_{k-1}x_{k-1} + \eta_{k-1} \quad cov(\eta_{k-1}) = Q_{k-1}$$

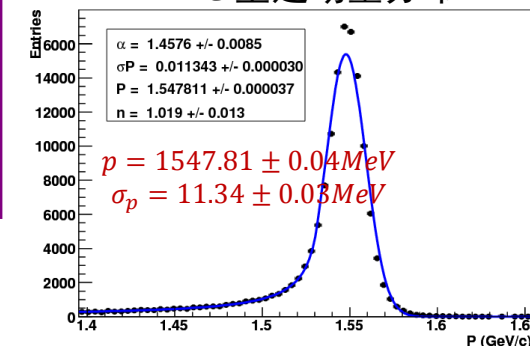
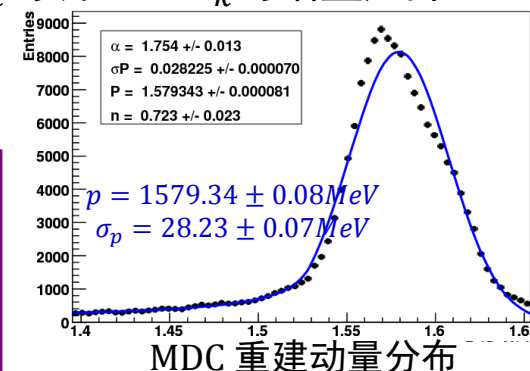
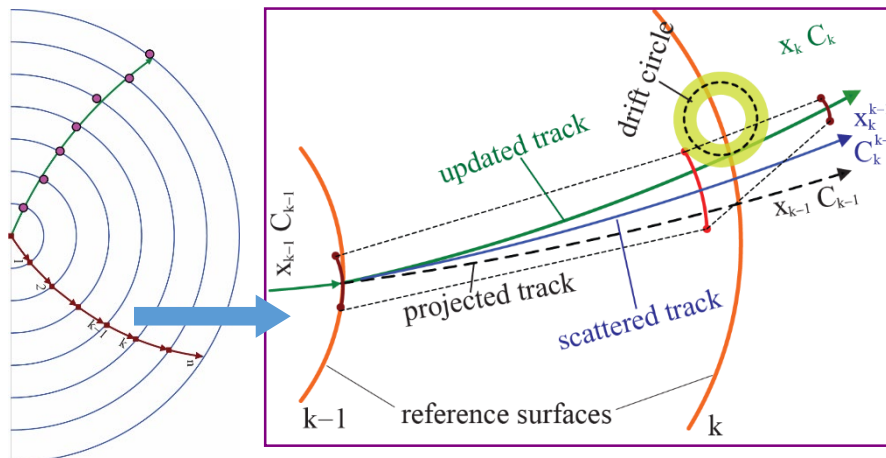
F_{k-1} : 传输矩阵, η_{k-1} : 随机的传输误差

测量响应方程: 描述观测量与状态矢量的关系

$$m_k = H_k x_k + \varepsilon_k \quad cov(\varepsilon_k) = V_k$$

H_k : 状态矢量 x_k 对测量量 m_k 的增益矩阵

ε_k : 测量误差



	State Vector	Error Matrix	State Vector Example
Measured Vector	m_k	V_k	Drift distance d_{drift}
True Vector	x	C	5 helix parameters $d_\rho, d_z, \kappa, \phi_0, \lambda$
Predicted Vector	x_k^{k-1}	C_k^{k-1}	5 predicted helix parameters $d_\rho, d_z, \kappa, \phi_0, \lambda$
Updated Vector	x_k	C_k	5 updated helix parameters $d_\rho, d_z, \kappa, \phi_0, \lambda$
Smoothed Vector	x_k^n	C_k^n	5 smoothed helix parameters $d_\rho, d_z, \kappa, \phi_0, \lambda$

电磁量能器的刻度与重建

- 电磁量能器用于精确测量光子能量和位置

- 刻度

- 探测单元（晶体）的能量刻度

Monte caro中
簇射团的沉积能量
预期沉积能量

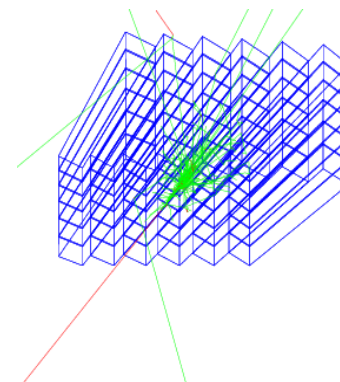
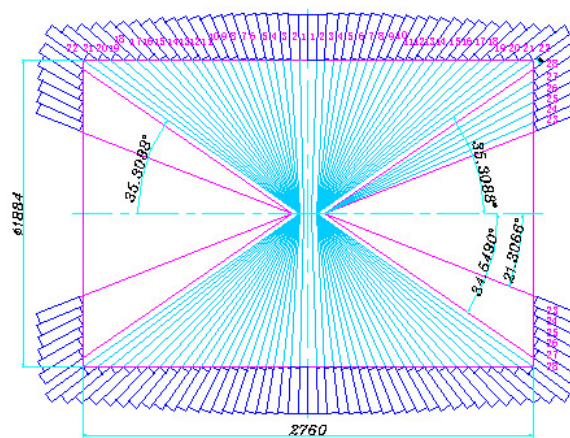
$$\chi^2 = \sum_{k=1}^N \frac{E_{exp}^k - \sum_i^{5 \times 5} g_i \cdot E_i^{elec}}{\sigma(\theta, \phi)}$$

- 簇射绝对能量刻度

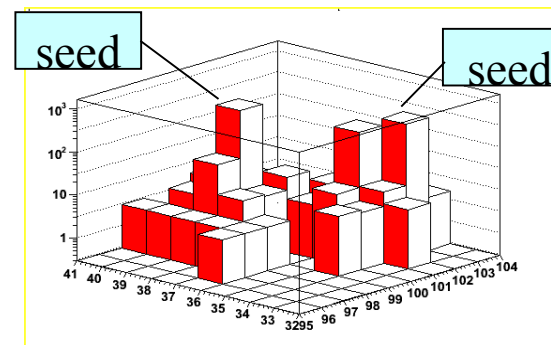
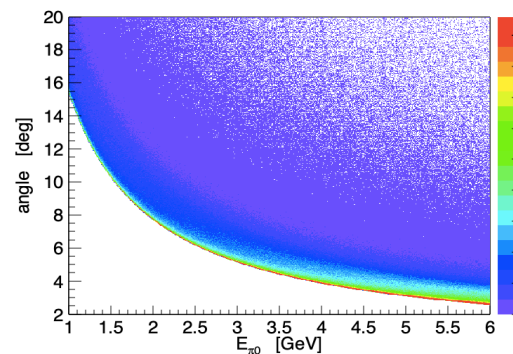
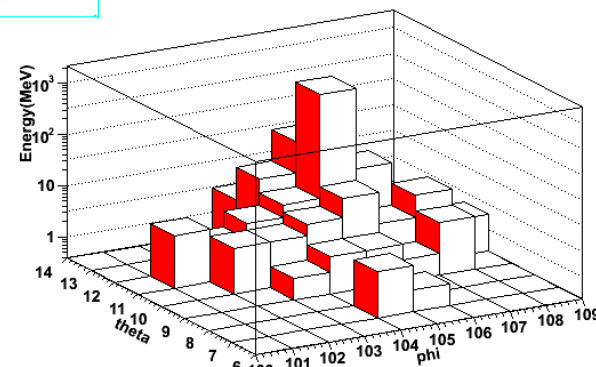
- 位置修正

- 重建

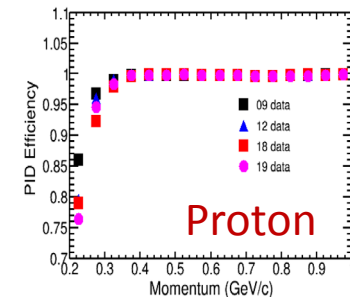
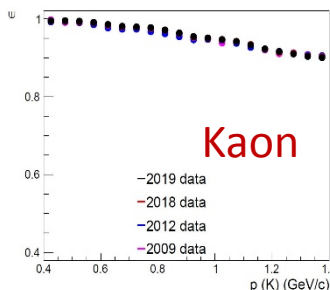
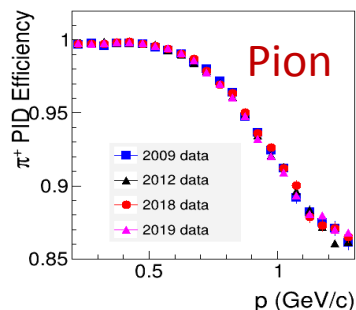
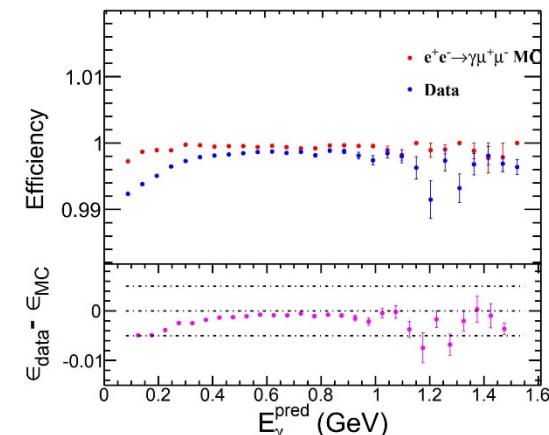
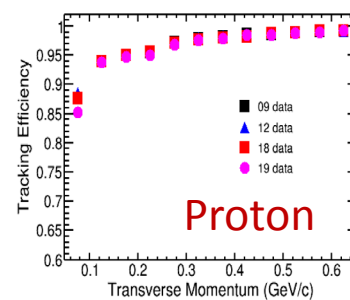
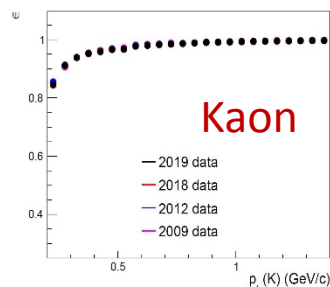
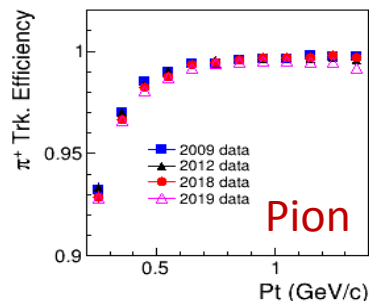
- 簇射团的寻找：一系列沉积能量大于某个阈值的晶体集合
- 簇射团的劈裂 ($\pi^0 \rightarrow \gamma\gamma$)：每块晶体测量能量包括来自多个簇射的贡献



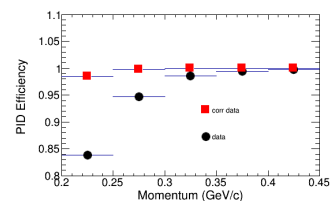
数据中簇射团的
沉积能量



重建效率



光子重建效率



低动量质子修正

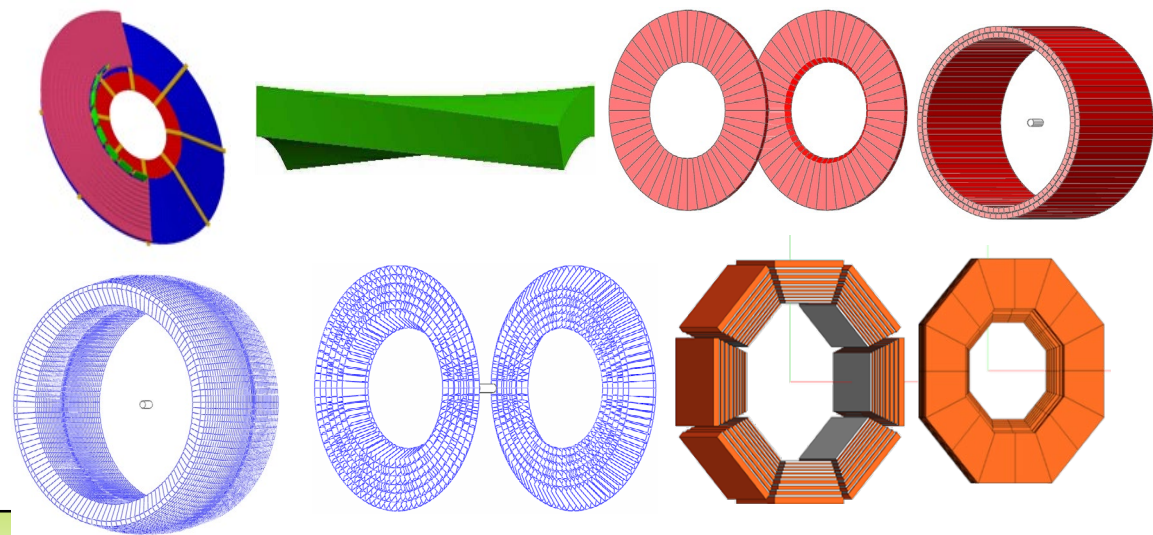
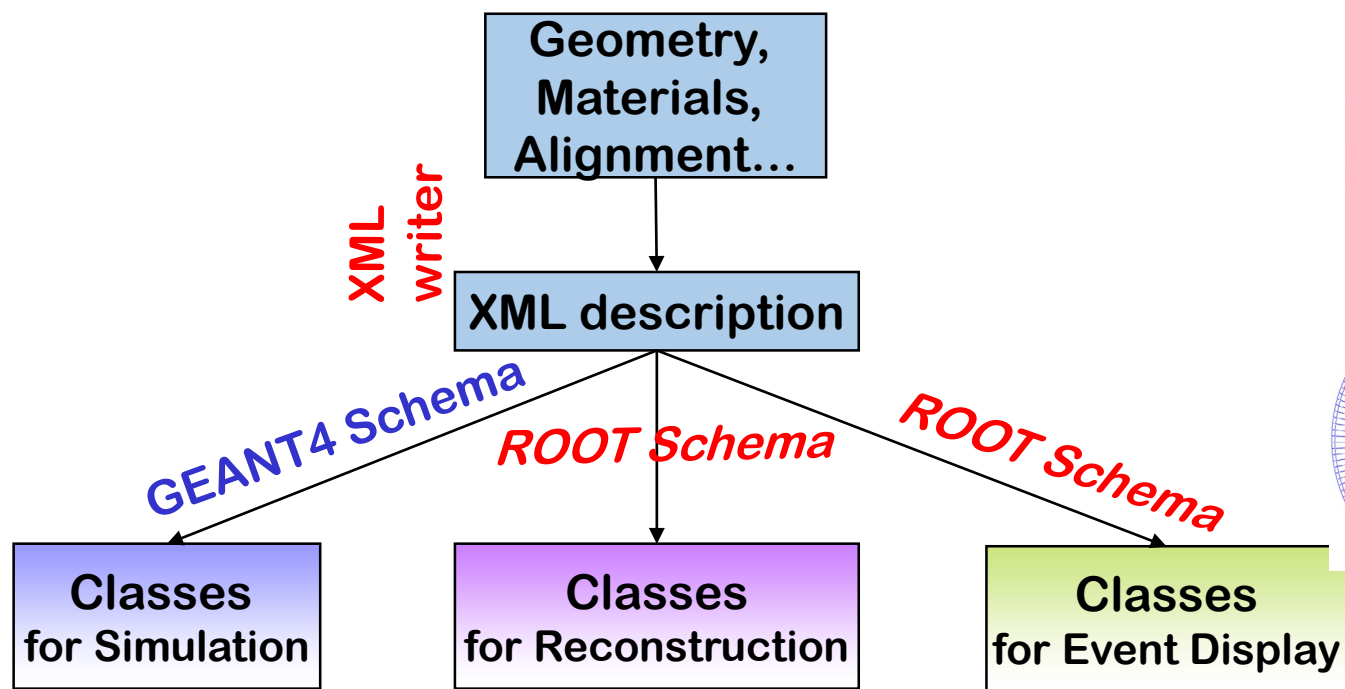
带电径迹重建效率

粒子鉴别效率

- 高横动量带电径迹重建效率接近100%；提高低横动量（<200MeV）径迹重建效率，压低假径迹比例是提高事例选择效率，对粲重子和粲介子研究十分重要
- 光子重建效率接近100%
- 粒子鉴别效率实现探测器设计要求；通过低动量质子修正获得正确粒子鉴别效率；尝试利用先进的机器学习/深度学习技术提高粒子鉴别效率

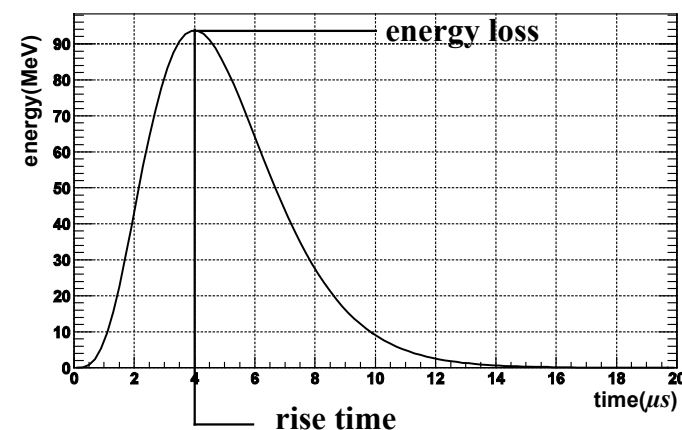
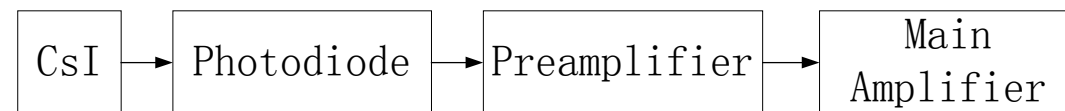
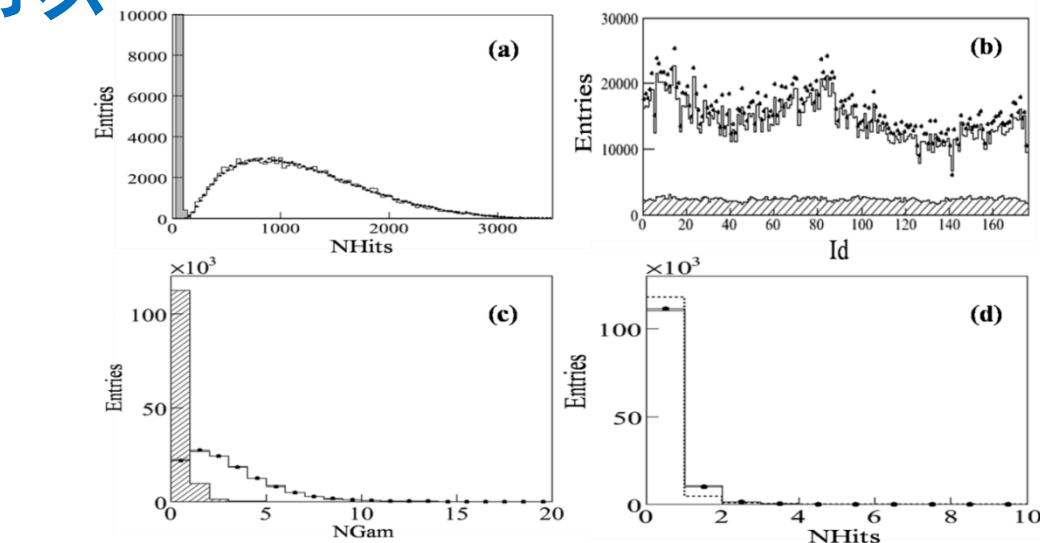
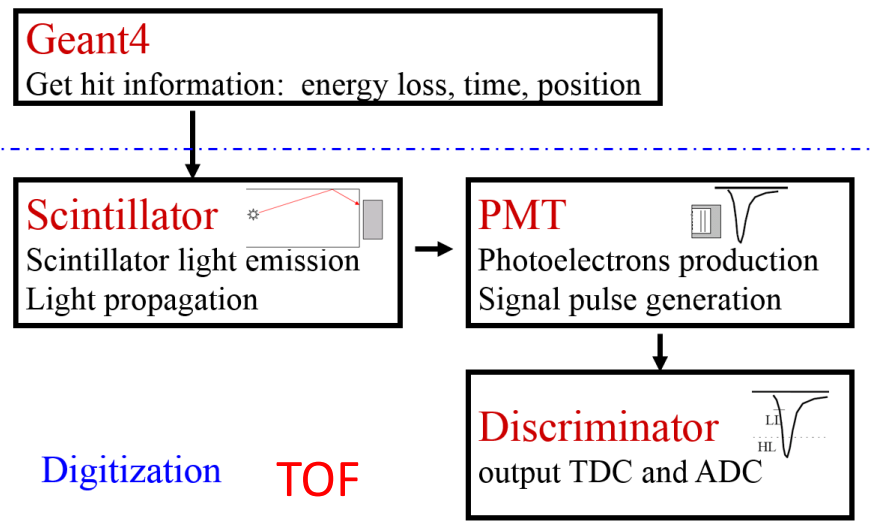
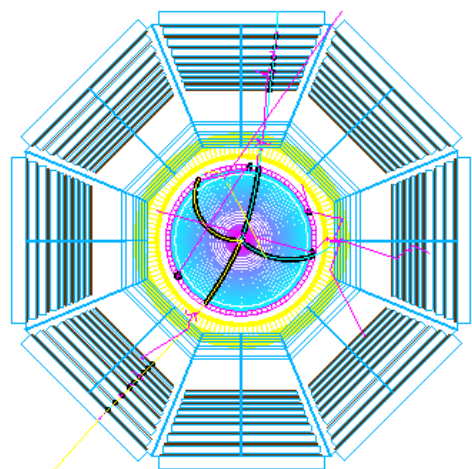
探测器几何构建

- BESIII探测器几何基于 GDML (Geometry Description Markup Language) 方式构造
- 实现了一系列不规则和复杂几何结构的精确探测器描述
- 基于GDML格式的统一的几何管理服务



探测器模拟

- 探测器模拟软件BOOST
- 粒子在探测器中的传输和相互作用
- 探测器的响应（数字化）
- 真实化模拟
- 调试和测试

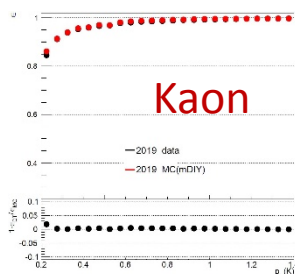
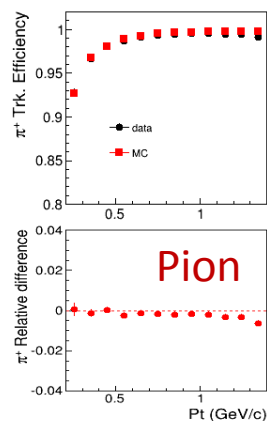


EMC

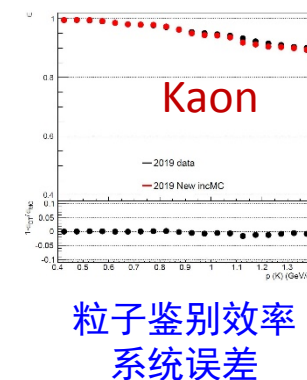
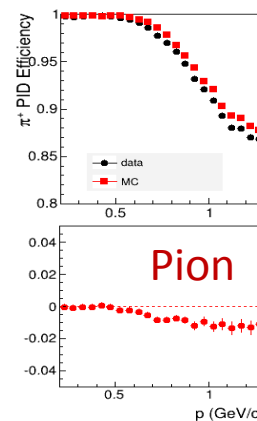
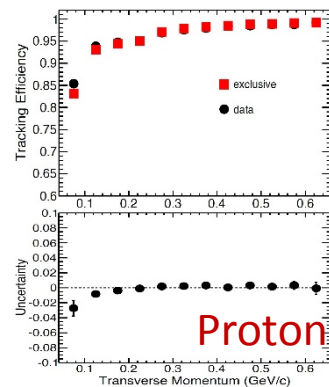
Readout electronics simulation for EMC

$\psi' \rightarrow \pi^+ \pi^- J / \psi \rightarrow \pi^+ \pi^- \mu^+ \mu^-$ Developed parameterized fast simulation model

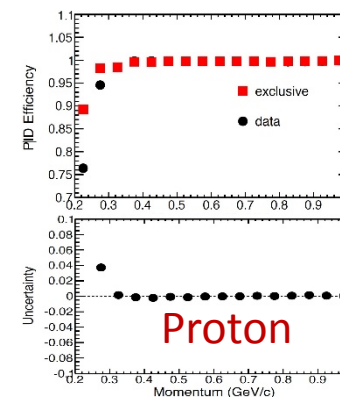
模拟 — 系统误差



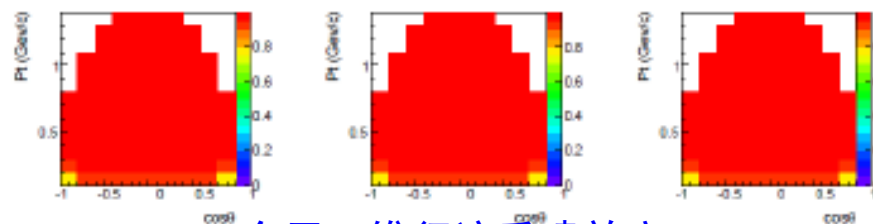
带电径迹重建效率
系统误差



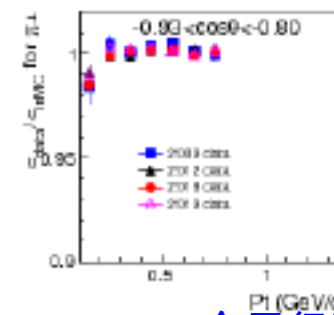
粒子鉴别效率
系统误差



Data exclusive MC inclusive MC



π 介子二维径迹重建效率

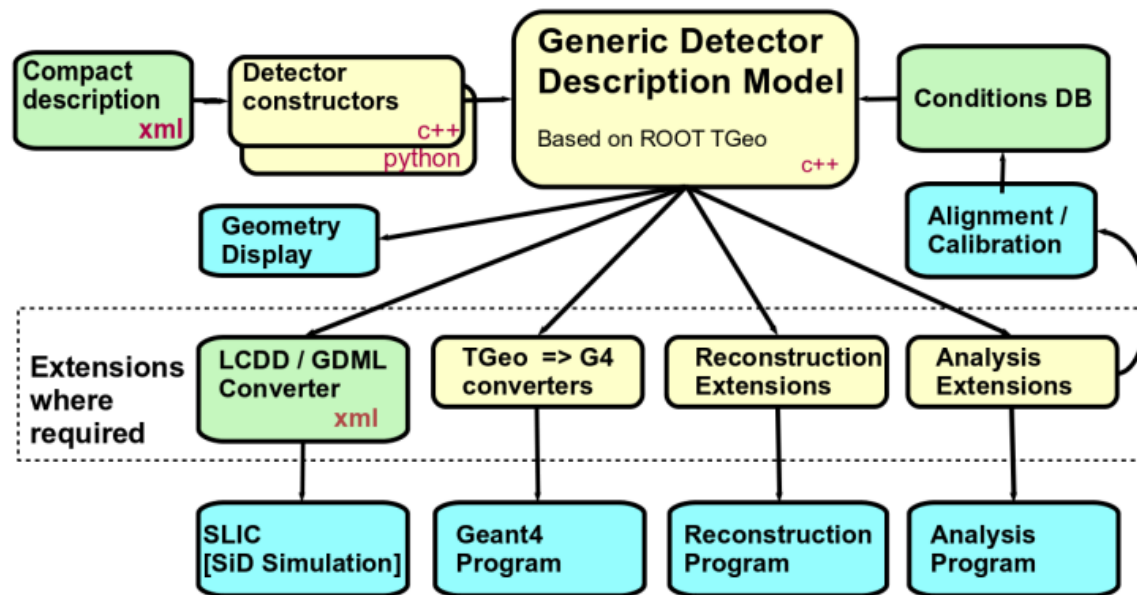


π 介子径迹重建效率修正因子

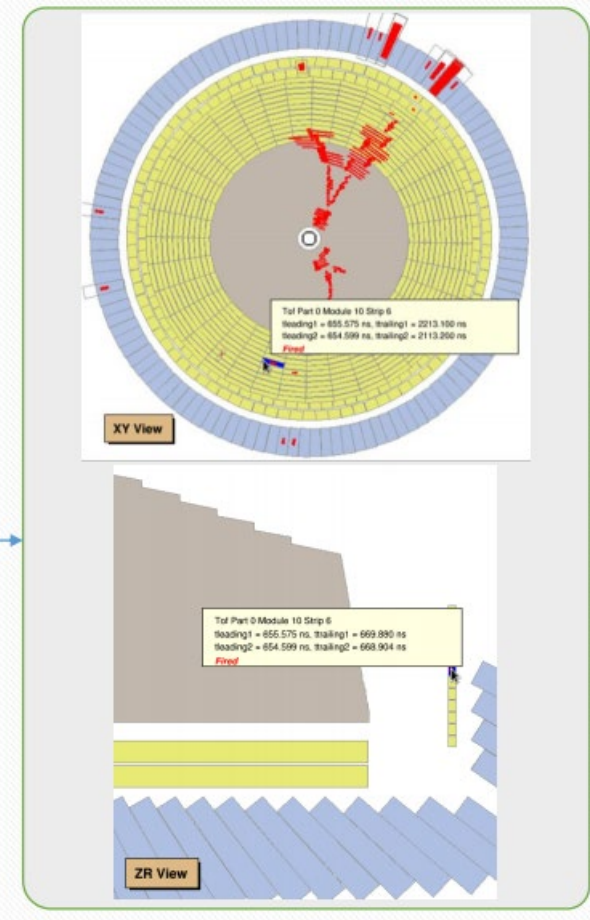
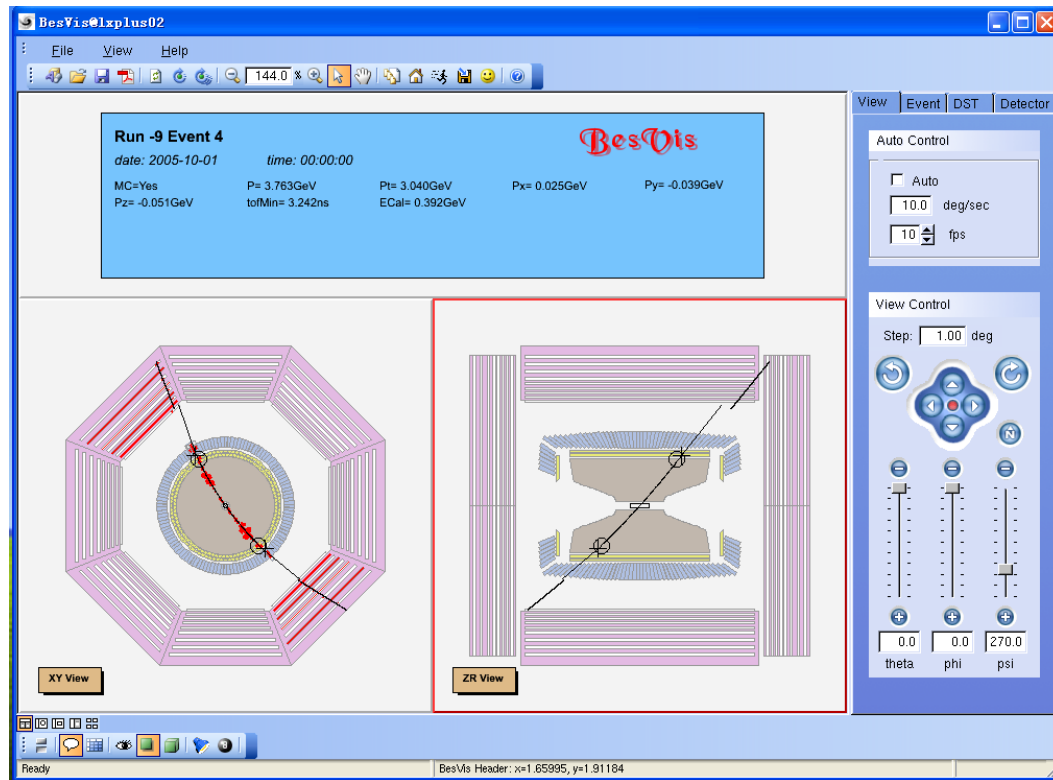
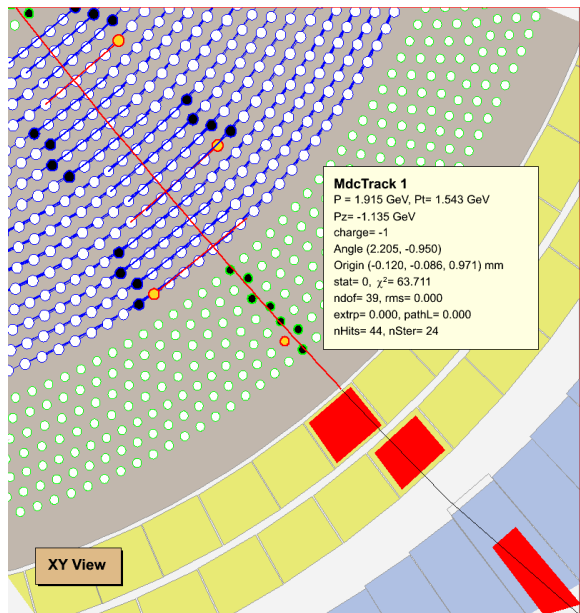
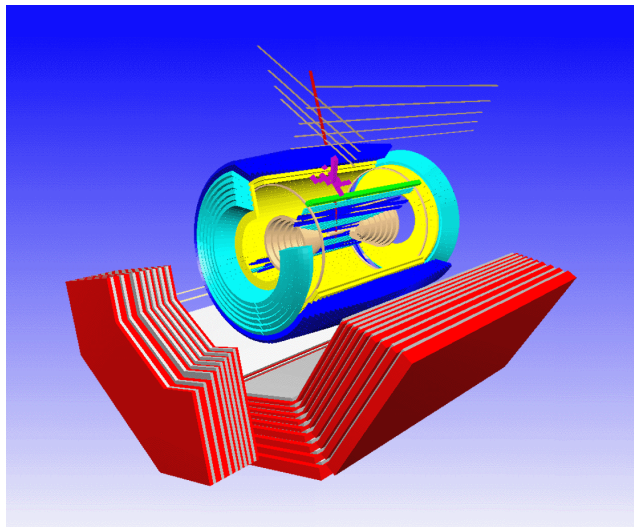
- 系统误差与统计误差一样，是衡量物理测量精度的重要参数
- 数据与模拟的不一致被认为是物理结果**系统误差的最主要来源**
- 高亮度（或高截面）的实验使得统计误差不断降低，要求进一步压低系统误差，使之达到与统计误差一致的水平
- 利用数据驱动的新思路解决系统误差问题

探测器描述工具：DD4hep

- 一套完整的探测器描述工具，提供唯一的信息来源
 - 几何、物质描述、可视化、读出、校准、刻度和重建算法
- 支持实验研究的整个生命周期
 - 探测器概念设计、优化、建造和运行
- 已经被 ILC、CLIC、FCC、CEPC 和 STCF 等实验采用



事例显示



- 事例显示软件是在事例级别对实验数据的可视化分析
- 广泛应用于探测器设计、监控、事例产生、重建、探测器模拟、数据分析等等

总结

- 数据处理不仅仅是粒子物理实验的“工具”，是实验的重要环节
- 原始数据主要包括电子学测量的时间幅度等信息，同时包括噪声和本底，不能直接用于物理研究；需要经过数据处理，得到各个粒子的测量的物理量
- 刻度和重建算法的性能直接影响物理事例的探测性能，如探测效率和能量分辨率等
- 模拟与数据的不一致是系统误差的主要来源，利用各种方法提高模拟精度是改善物理测量精度的重要手段
- 新技术新思维将为物理研究提供更加强有力地支撑和推动

谢谢大家！

