

数据分析-以AMS宇宙线通量测量为例



骆首栋

2026/06/29

什么是数据分析

数据分析就像一位“科学侦探”：面对海量、嘈杂、甚至存在错误的**数据**，通过**统计学、数学和计算机工具**，把隐藏其中的物理规律一步步挖掘出来，最终让数据“开口说话”。

淘出数据深处的“物理真金”

实验结束后，我们往往得到的不是答案，而是一大堆数据。这些数据就像混杂着泥沙的河水。而数据分析的工作，就是：

- 去掉泥沙（剔除噪声）
- 挑出金粒（提取有效信息）
- 称重测量（定量计算）
- 判断真假（金子还是黄铜）

最终找到隐藏在数据里的物理规律。

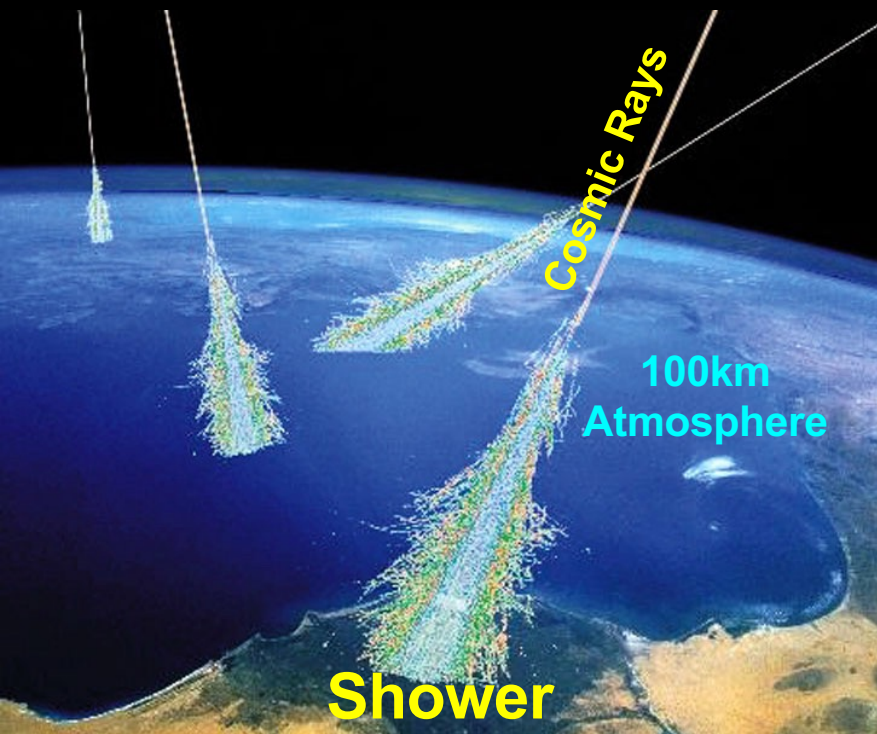


实验产生数据，物理数据分析赋予数据意义。
没有数据分析，再精密的实验也只是一堆数字。

AMS 国际空间站实验

阿尔法磁谱仪(AMS)是国际空间站(ISS)上唯一的大型磁谱仪，也是人类送入太空的第一个大型磁谱仪。AMS实验是丁肇中教授领导的大型国际合作项目，其科学目标是寻找宇宙中的反物质、暗物质及精确测量宇宙线的成分和能谱。目前AMS实验已经累计运行超过14年，收集到约2600亿个宇宙线事例

带电宇宙线会被100km厚的地球大气吸收(~ 10m of water)。



因此，带电宇宙线组成的精确测量在地面无法完成。为了测量宇宙线的质量和动量，我们需要在太空放一个磁谱仪。

AMS: a unique TeV precision spectrometer in space

穿越辐射探测器 (TRD):
Identify e^+ , e^- , Z



AMS 主要测量:

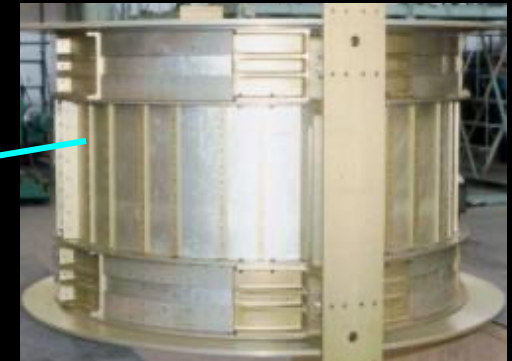
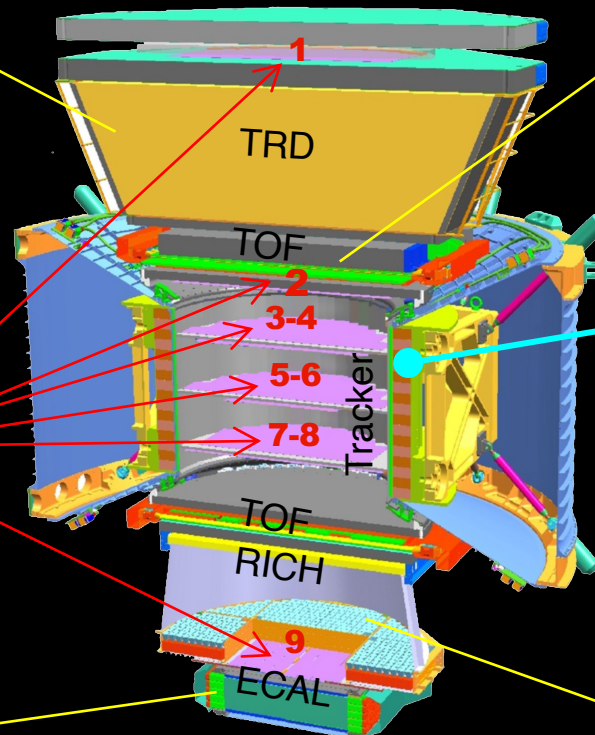
- 电荷 (Z)
- 能量 (E, GeV/A)
- 刚度 $R = P/Z$ (动量/核子电荷) (GV)

飞行时间探测器 (TOF): Z, E



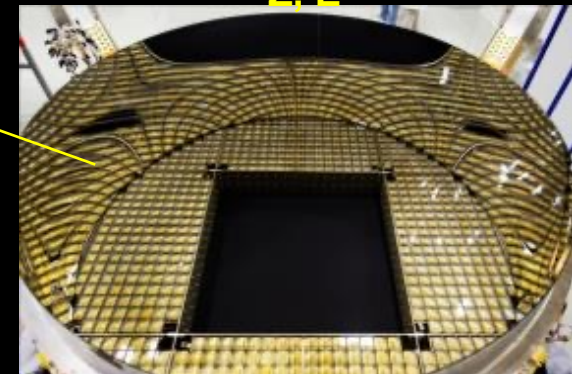
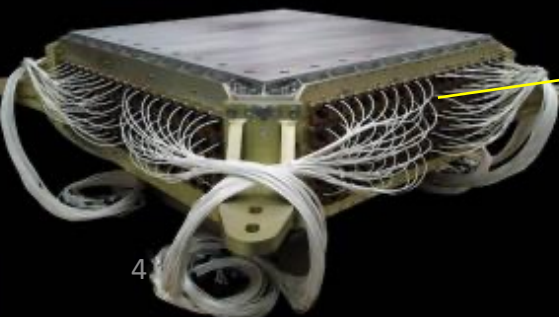
永磁体: $\pm Z$

硅微条径迹探测器 (Tracker):
Z, P



环形切伦科夫探测器 (RICH):
Z, E

电磁量能器 (ECAL): E of e^+ , e^-



Nuclei are measured independently by Tracker, RICH, TOF, and ECAL

AMS: a unique TeV precision spectrometer in space

穿越辐射探测器: Identify e^+ , e^- , Z



1. 穿越辐射探测器安装在整个探测器的顶部，它的主要作用是通过穿越辐射（带电粒子穿过不均匀的介质时发射的电磁辐射）判别 e^\pm/p ，带电粒子产生的穿越辐射光子数取决于洛伦兹 $\gamma = E/m$ ，同能量的质子和 e^\pm 的 γ 相差 1800 多倍，所以 e^\pm 相较质子更易产生穿越辐射光。
2. 穿越辐射探测器可以通过粒子的电离能损对粒子的电荷进行独立测量。

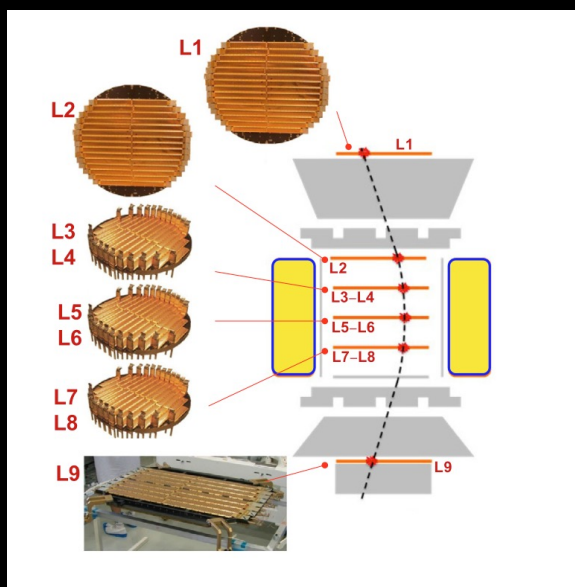
飞行时间探测器: Z, E



1. 4 层 TOF 分别由 8, 8, 10, 8 个塑料闪烁体条 (paddle) 组成，每个 paddle 两端通过光导连接在 2-3 个光电倍增管 (PMT) 上
2. TOF 的信号大小由带电粒子在 TOF 中的电离能损 $dE/dx \propto Z^2$ 决定，TOF 对 $Z=6$ 原子核的电荷分辨为 $\Delta Z/Z = 2.67\%$ ；对 $Z=26$ 原子核的电荷分辨为 $\Delta Z/Z = 1.46\%$

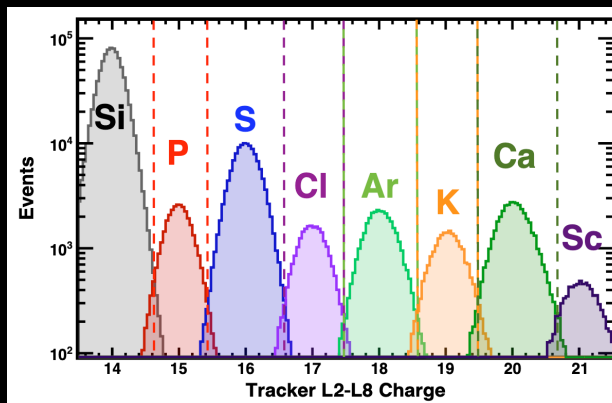
AMS: a unique TeV precision spectrometer in space

硅微条径迹探测器: Z, P

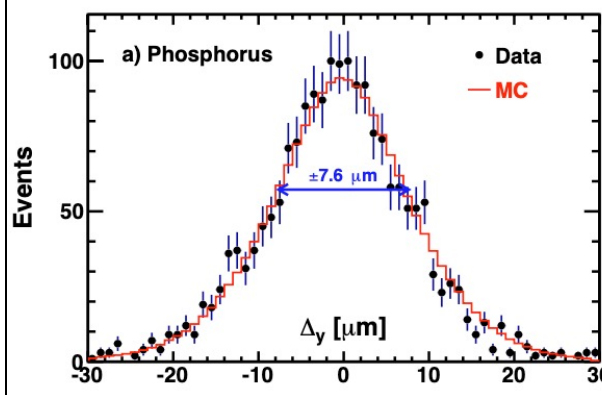


1. 径迹探测器是 AMS 的核心探测器之一，它可以通过多层的位置信息 (x, y 方向坐标) 对带电粒子在磁场中的运动轨迹进行测量，还可以通过电离能损对粒子电荷进行测量。通过 Tracker 重建径迹和磁场强度信息，可以得到精确的带电粒子刚度信息。
2. 当带电的宇宙线粒子穿过一层 tracker，会在击中探测器 sensor 上位置相邻的铝制微条 (stripes) 产生信号，击中位置的坐标由相邻的 strip 信号大小的比值决定。
3. Tracker 硅微条得信号大小正比于 Z^2 可以通过信号大小对粒子电荷的绝对值 $|Z|$ 进行测量。同时 Tracker 可以联合多层探测器信号 (Tracker L2-L8, Inner Tracker), 得极其优益的电荷分辨

Tracker L2-L8 电荷分辨



Tracker 位置分辨



硅径迹探测器是AMS中最重要、最复杂的子探测器，详细介绍参考严老师之前发的ppt

以磷核子宇宙线通量的测量为例

Flux Definition:
$$\phi_i = \frac{N_i}{A_i T_i \epsilon_i \Delta R_i}$$

N_i : 事例数 (经过背景修正以及刚度迁移影响修正)

A_i : 有效接受度

T_i : 曝光时间

ϵ_i : 效率修正

ΔR : Bin 宽

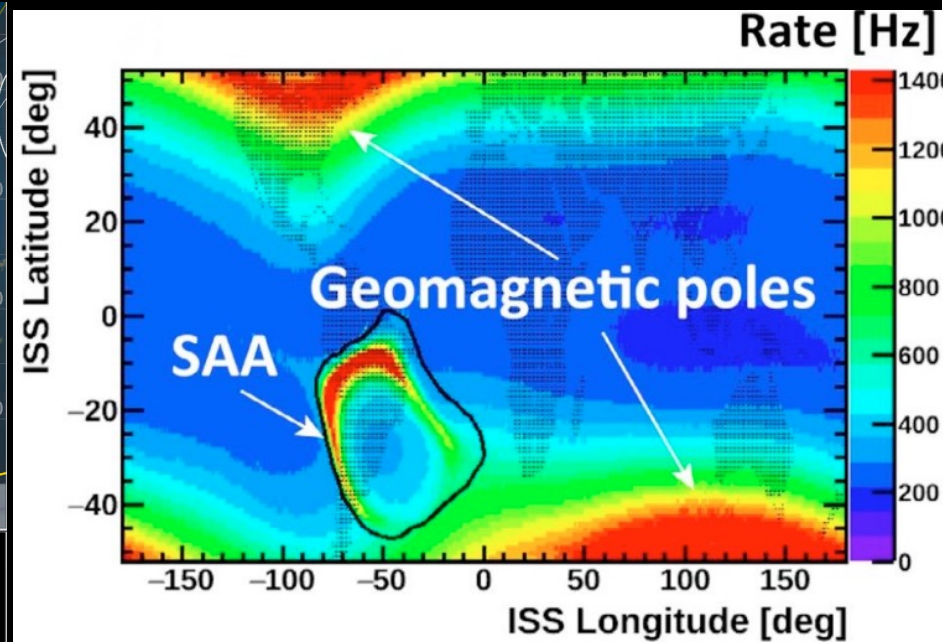
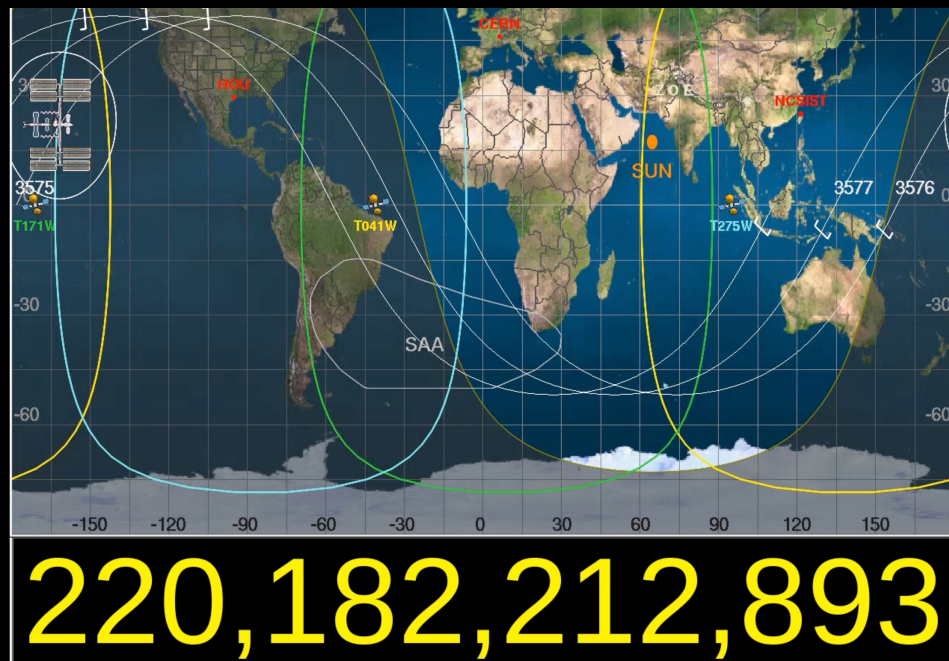
物理分析里面，所有人为引入的或者探测器响应导致的对最终物理结果的影响都需要被仔细考虑

物理数据分析是一个严谨科学的研究方法，需要权衡不同的筛选条件对结果的影响，考虑探测器自身的响应，最后追本溯源还原真实的物理规律。

淘金我们希望最后得到的东西里面泥沙越少，那意味着可能漏掉的金子也越多，反之亦然。此外可能由于工具限制，那些看起来都是金子，里面可能也有一些杂质。

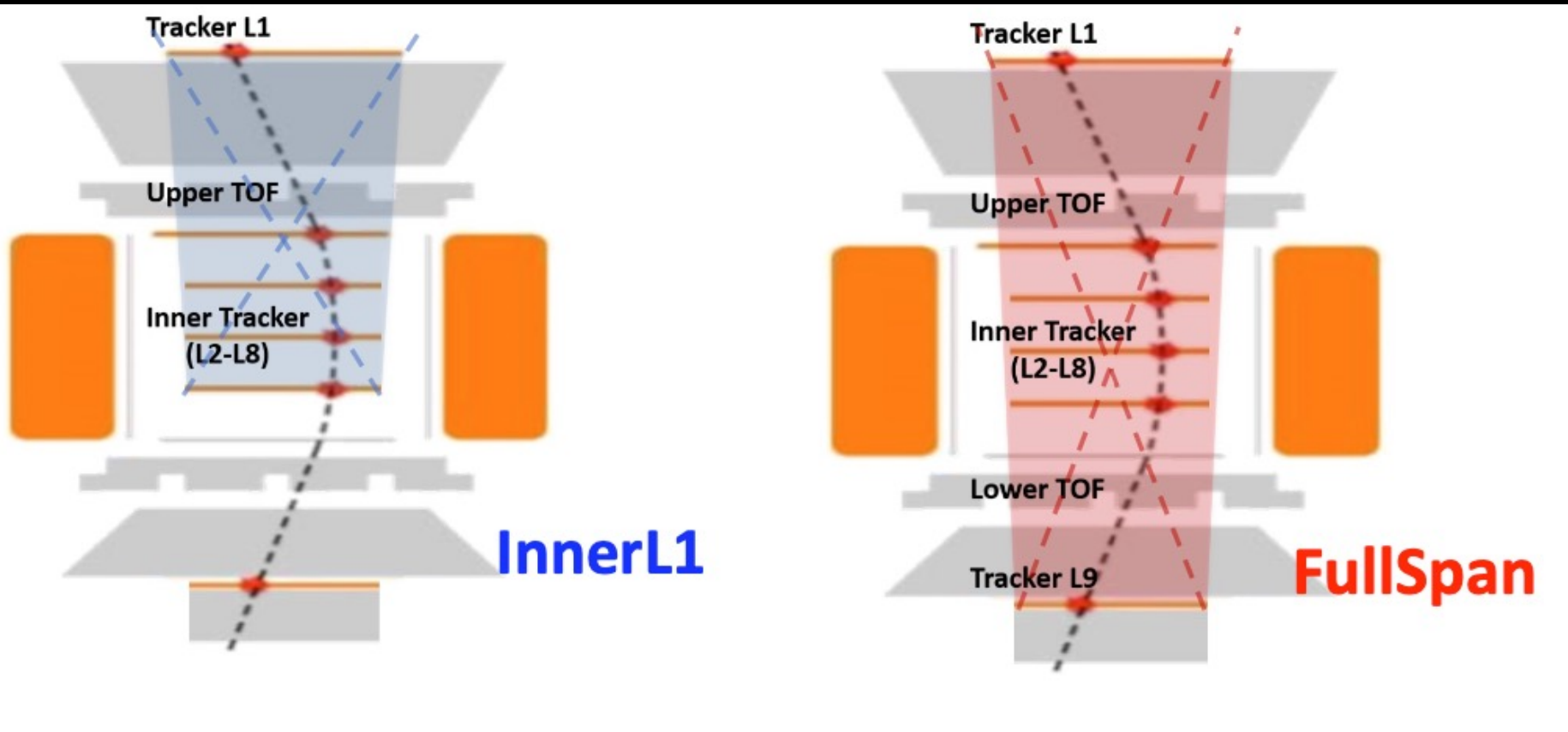
AMS 曝光时间

- 当宇宙线穿过探测器后，探测器通过一系列条件判断是否要对其进行探测。当宇宙线粒子通过触发后，探测器会对信号进行处理并将数据通过一系列操作后传输到地面，信号处理和数据传输过程需要时间，所以在一个粒子通过触发后的一段时间内不能再触发到另一个粒子，这段时间为探测器死时间。同时，地磁场的存在使得低于一定刚度（地磁截止刚度）的宇宙线受到地磁场调制，所以这个地磁截止刚度以下能区的事例已经丢失了部分原始的能量成分信息，不能用于宇宙线能谱的测量。排除地磁场影响以及修正探测器死时间后的有效探测器工作时间称为曝光时间。
- 计算曝光时间时，需要选取各个子探测器均正常工作状态的时间；而且要选取活时间的比例大于一定值，因为探测器触发过于频繁可能会导致部分子探测器信号异常或不完整



Geometry for AMS Nuclei Flux Analysis

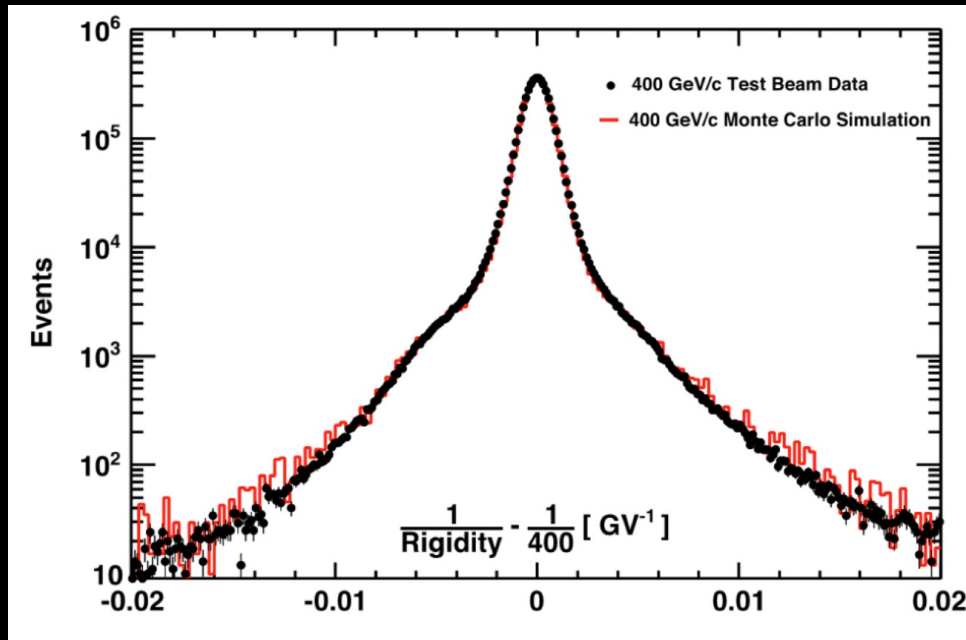
- 探测器几何：AMS选用了两种探测器几何，L1Inner(Tracker L1-L8)和 FullSpan(Tracker L1-L9)，进行分析，其中L1Inner 几何利用Tracker L1、L2-L8、UpperTOF，TRD进行分析。
- 不同的探测器几何有着不同的优缺点：InnerL1几何有着更高的几何接受度（能接收到更多的粒子），但是刚度测量上限为1.2TV;FullSpan 能测量更高的刚度，但是几何接受度更低，接收到的事例数少。



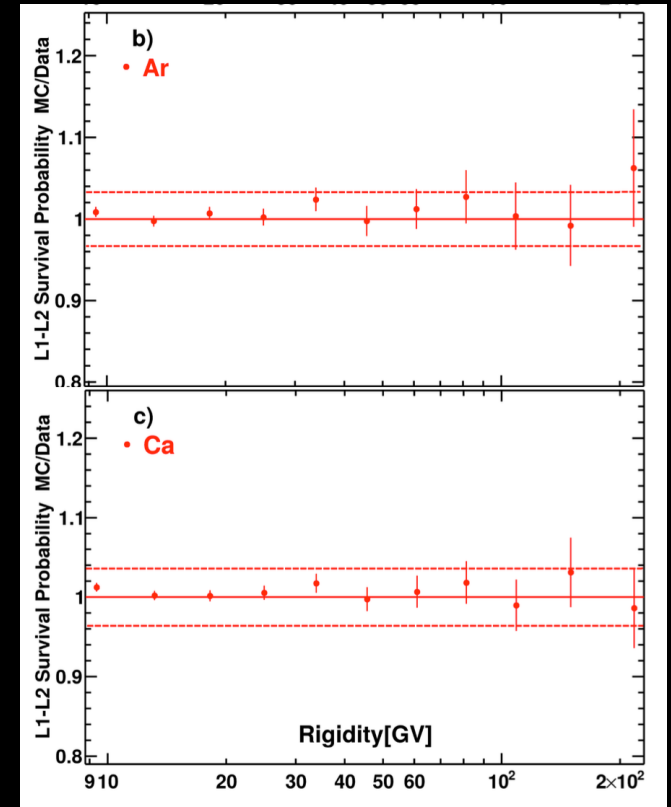
AMS Acceptance

- AMS 的蒙特卡洛模拟了原子核与探测器的各种相互作用，进行真实的数字化。在数据分析过程中，MC 是飞行数据的重要参考对象。AMS MC 基于成熟的相互作用模型和对探测器细致全面的了解，可以模拟粒子在探测器中的真实情况。在流强谱的测量过程中，以 MC 为参照可以有效减少由飞行数据统计量不足造成的效率计算偏差；同时 MC 的产生信息可以弥补分析过程中无法从飞行数据获取的信息，例如探测器的有效接收度，能量分辨对能谱形状的影响。

AMS 探测器束流测试和MC模拟的刚度分辨对比



AMS 探测器真实数据和MC模拟的L1-L2之间Survival probabilities



AMS Acceptance

- 接受度：表征探测器“接收”并“探测”到宇宙线粒子的能力，决定了单位时间被粒子探测到的粒子数目
- AMS 的探测器结构和内部磁场较为复杂，我们可以根据 MC 对探测器的接收度进行准确估算：先在 AMS 探测器上部定义了一个足够大的平面，要求从上到下通过 AMS 的粒子一定会穿过该平面AMS 的 MC 事例即产生于从该平面上的随机投点，用来模拟穿过该平面的各向同性的宇宙线事例。

经过事例筛选后得到的事例数

$$A_{eff} = A_{plane} \cdot \frac{N_{i,sel}}{N_{i,gen}}$$

MC投点平面产生的总事例数

- A_{plane} 为 MC 产生平面对从上方入射粒子的接收度，考虑宇宙线为各向同性，可以通过 (球坐标) 公式得到

$$\begin{aligned} A_{plane} &= \int_{\Omega} d\Omega \int_S d\vec{S} \cdot \hat{r} = \int_0^{\frac{\pi}{2}} \sin\theta d\theta \int_0^{2\pi} d\varphi \int_S \cos\theta dS \\ &= \frac{1}{2} \cdot \int_S \int_0^{2\pi} \int_0^1 d(\sin^2\theta) d\varphi dS = \pi \cdot S \end{aligned}$$

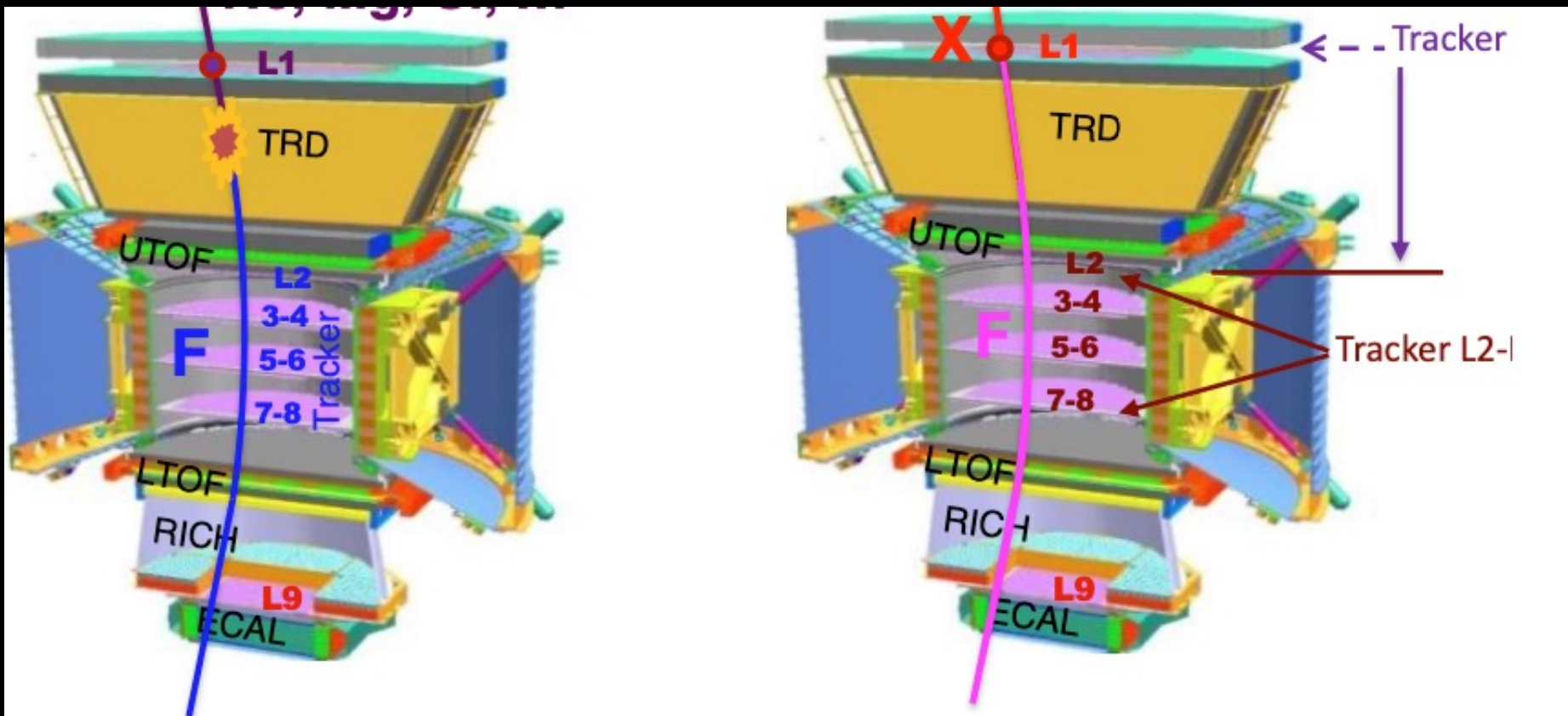
- MC 产生平面是一个边长 L 为 3.9m 的正方形平面，在 z 方向的高度为 +1.95m 计算为 $47.784m^2sr$ ，也称为 AMS 的几何因子。

本底估计

- 通过筛选的事例中依然混杂了部分本底。AMS 的电荷分辨能力可以排除绝大部份其它原子核的本底，通过各子探测器的电荷筛选基本可以确定穿过inner tracker 的粒子即为信号。但是在 inner tracker 之上发生碎裂并通过了事例筛选的本底无法被完全排除，需要根据探测器响应进行本底估计，这部分本底按发生碎裂的位置可以分成两类进行研究。

TRD 和 UTOF 上碎裂 (Below L1)

Tracker L1 及 L1 之上支撑结构上碎裂 (TOI)

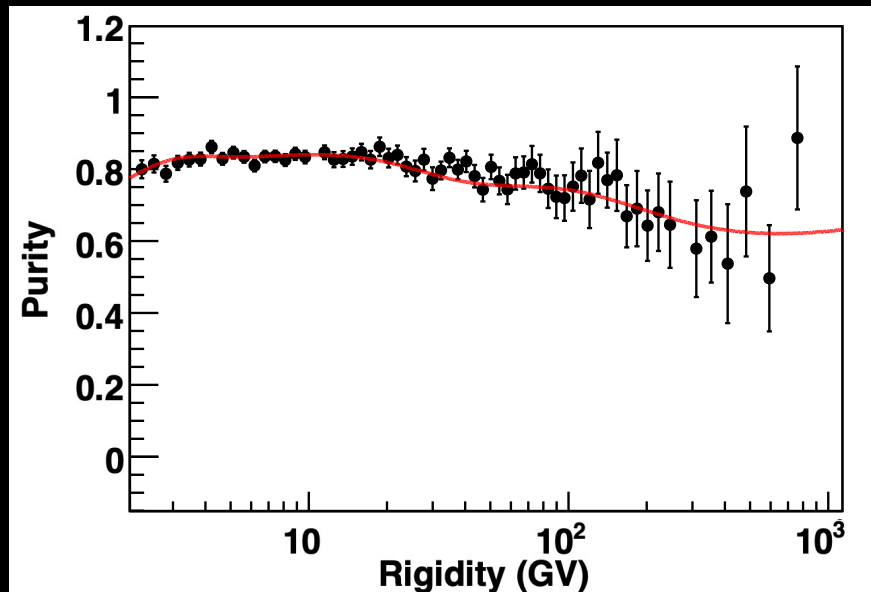
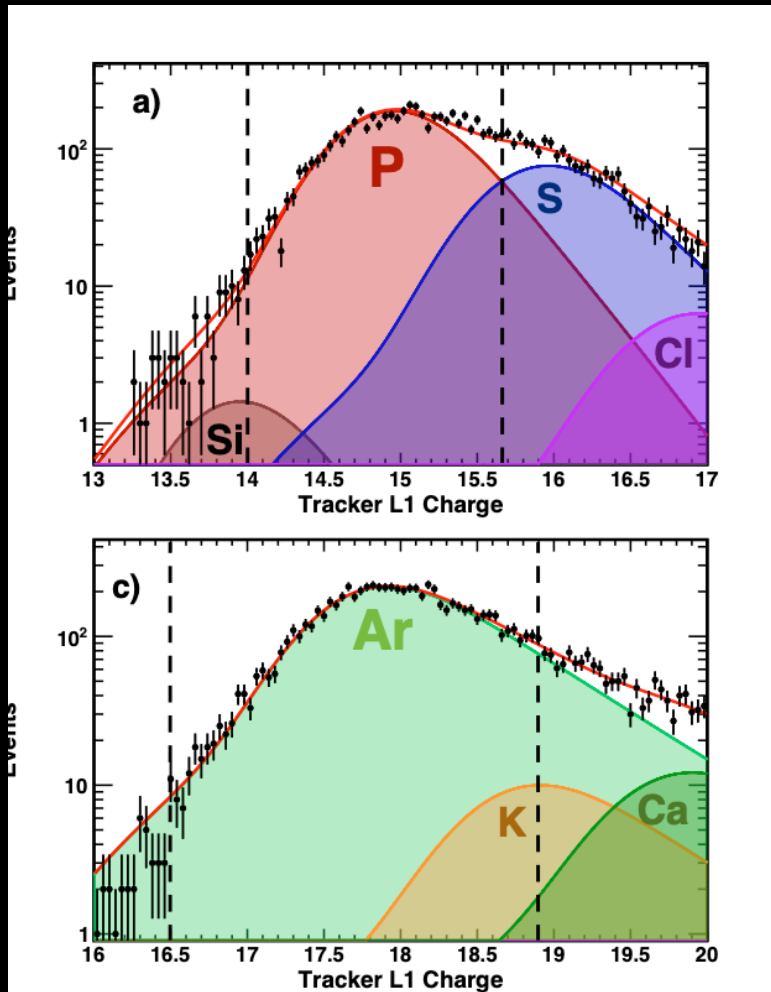


Below L1

- 由于Tracker自身探测器分辨的限制以及TRD物质质量相对较多, 有部分更高Z 原子核通过了 L1 的事例判选且在 TRD 上发生碎裂, 并且通过了 TRD 之下各子探测的事例筛选条件, 这部分主要通过得到不同核子在Tracker L1上的电荷响应模版, 然后对数据进行拟合得到.

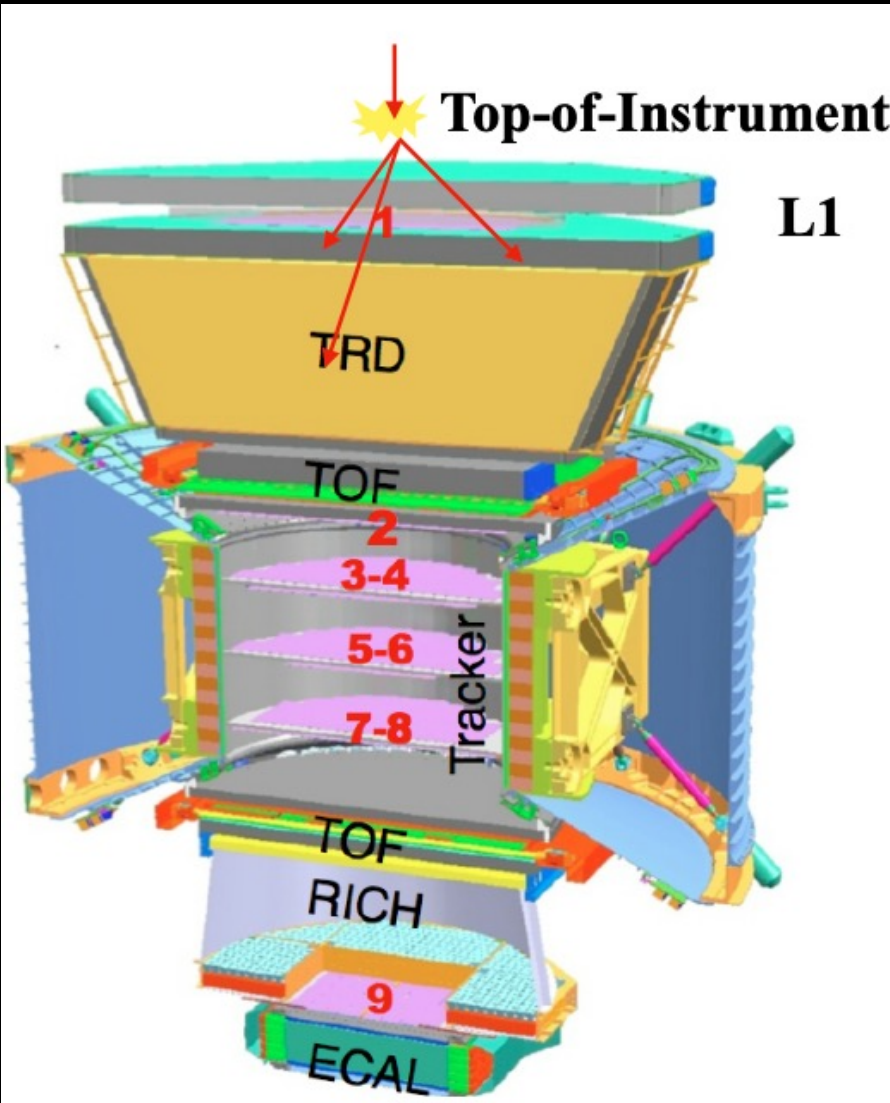
在实际测量中很难从数据中得到 Tracker L1 上干净的信号电荷测量分布形状, 然而 Tracker L1 和 L2 的硬件构造基本相同, 在理想状况下对相同电荷的测量值分布形状相同, 而且 L2-L8 之间物质质量很少, 几乎不会造成粒子碎裂, 可以筛选得到纯净的信号

通过模版拟合得到的P Below L1本底占比



TOI

- 在 tracker L1 及 L1 之上支撑结构上碎裂的本底：更高 Z 的原子核宇宙线在 L1 和 L1 之上 AMS 支撑结构上发生碎裂，产生信号原子核并通过了事例筛选，由于 L1 之上没有电荷测量信息，这部分本底只能用 MC 进行估计



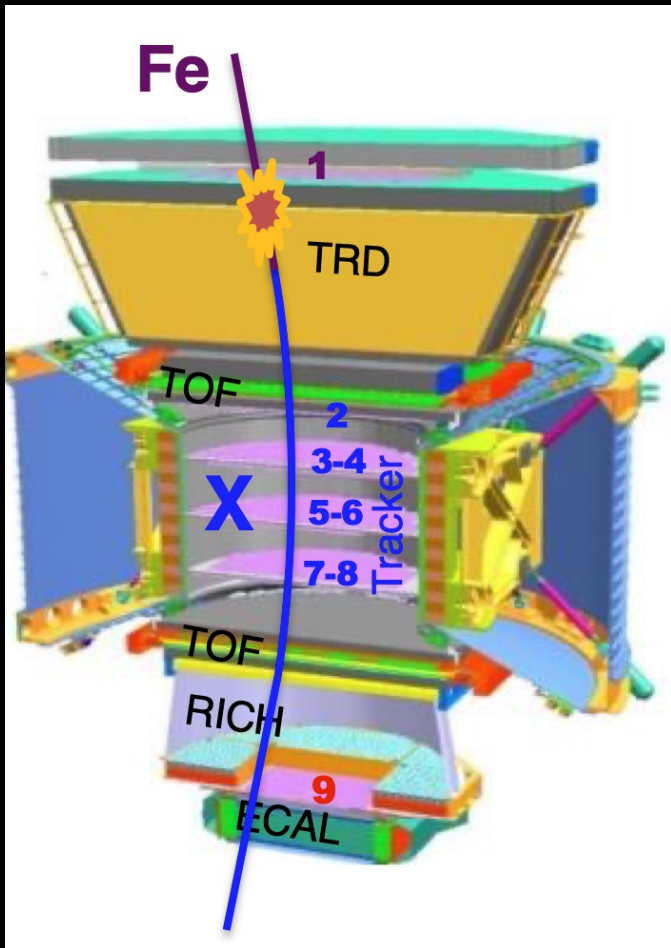
$$\begin{aligned}\frac{\sum_X N_{X \rightarrow Z}^{\text{Rec}}}{N_Z^{\text{Raw, Rec}}} &= \frac{\sum_X A_{X \rightarrow Z}^{\text{Rec}} \times \Phi_X^{\text{Gen}} \times \Delta R \times T}{N_Z^{\text{Raw, Rec}}} \\ &= \frac{\sum_X A_{X \rightarrow Z}^{\text{Rec}} \times \Phi_X^{\text{Gen}} \times \Delta R \times T}{A_Z^{\text{Rec}} \times \Phi_Z^{\text{Raw, Gen}} \times \Delta R \times T} \\ &= \frac{\sum_X A_{X \rightarrow Z}^{\text{Rec}} \times \Phi_X^{\text{Gen}}}{A_Z^{\text{Rec}} \times \Phi_Z^{\text{Raw, Gen}}}\end{aligned}$$

考虑到粒子碎裂会对刚度产生一定的影响，在above L1本底的计算中，我们需要在测量刚度（探测器真实测得的刚度）下，对这一部分本底进行估计

例如对于原子核序数Z=16的核子，在宇宙中是按10GV的流强谱产生的，但是由于碎裂以及探测器自身刚度分辨影响，最终探测器可能测到的8GV的事例，我们在计算时考虑这一部分影响

TOI

- 考虑到AMS的MC模拟的核子碎裂的分支比和真实数据可能存在一定的差异，在利用MC估计TOI碎裂本底估计的时候，我们需要对这个进行修正。
- 我们以Tracker L1作为束流，具有高物质的量的TRD和TOF作为靶，计算特定核子碎裂分支比的Data/MC差异。例如在TrackerL1上选择铁核,在Inner Tracker 上选择磷原子核，来研究Fe->P 的碎裂分支比差异



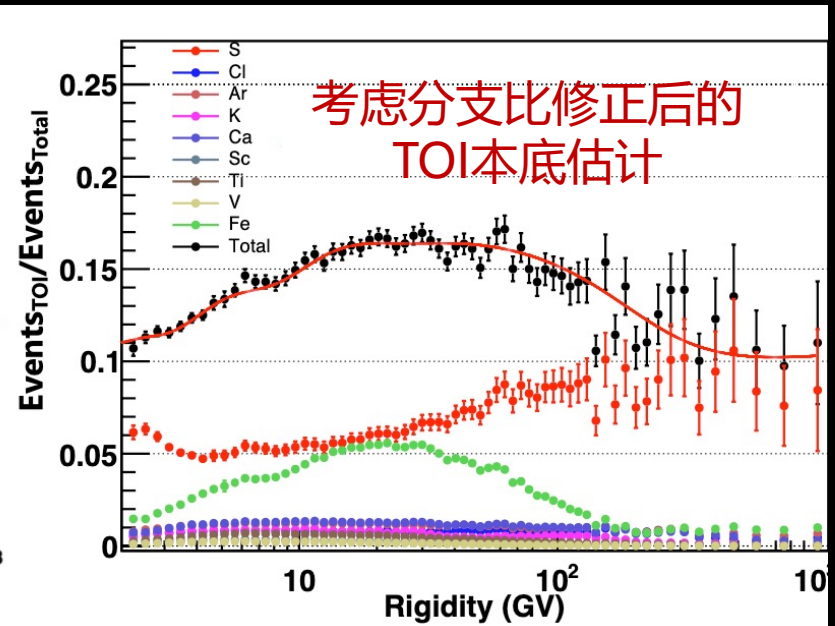
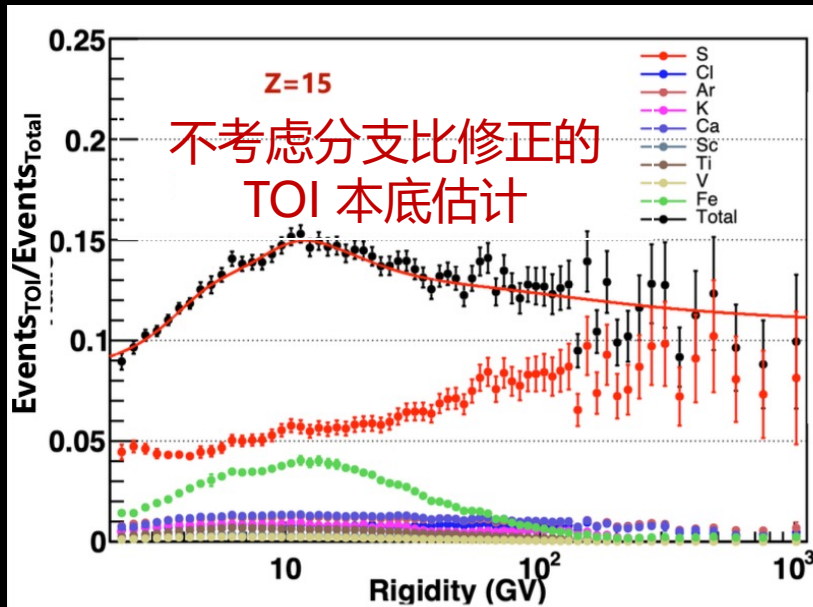
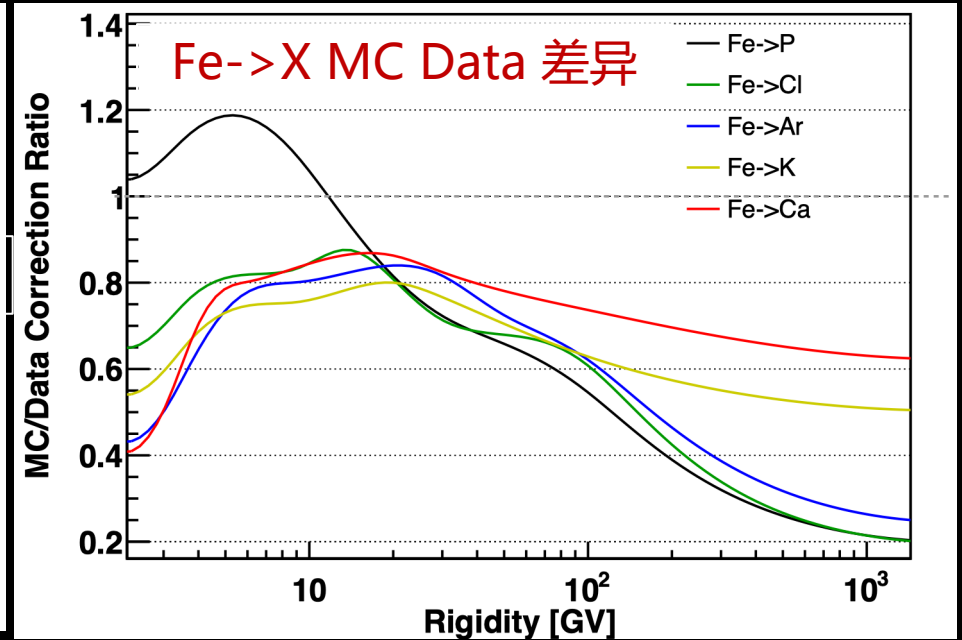
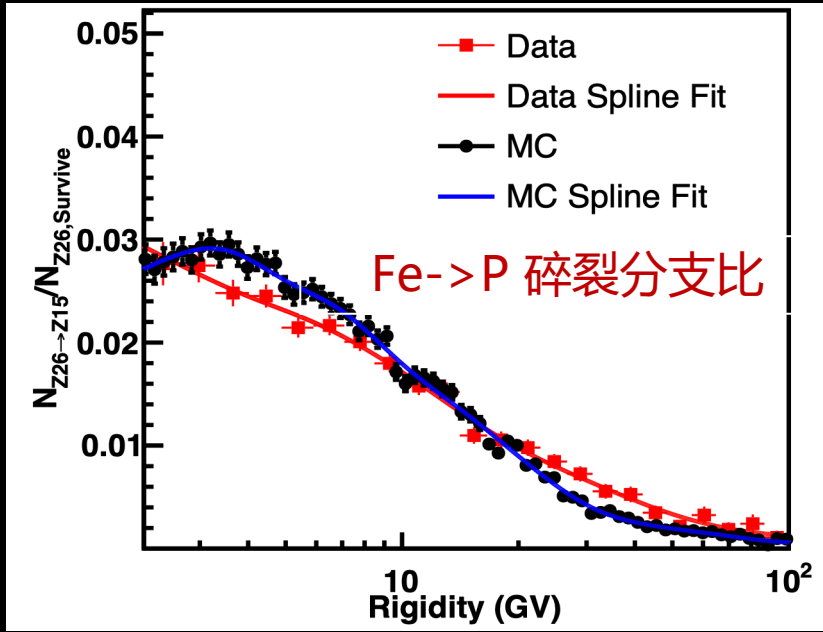
相互作用概率定义为碎裂事件数与未发生相互作用事件数之比：

$$P_{int} = \frac{N_{Z \rightarrow X}}{N_{Z, survive}}$$

其中， $N_{Z \rightarrow X}$ 为 L1 以下发生 $Z \rightarrow X$ 碎裂反应的事件数， $N_{Z, survive}$ 为 L1 以下未发生核相互作用而保持存活的事件数。

Tracker L1 的电荷分辨率有限，因此在 L1 的电荷选择中会引入一定的污染。例如对于铁 (Fe) 而言，由于Inner Tracker 具有很高的电荷分辨率，因此对 Fe→Fe (即未发生碎裂的铁核) 事件的影响很小；但是，对于 Fe→X (铁核碎裂成其他核) 的事件，污染效应仍然不可忽略。我们可以通过前面所述的电荷模版拟合等方法，剔除L1上引入的本底

Above L1



作业

- 什么蒙特卡洛模拟，为什么我们需要这个？探测器的接受度的物理意义是什么，这个是如何计算出来的？
- 我们可以通过一些基础的电荷筛选等，可以得到一个核子的事例数，参考下列筛选条件，Event Selection for data analysis部分，思考我们为什么要加这些筛选条件（不一定每个条件都要写，自己任选）？背景修正在这里主要有哪些背景？
- 由于探测器分辨率有限，测到的刚度可能偏大或偏小，这种测错导致的刚度迁移需要用数学方法修回来（即 Unfolding），这种刚度迁移的修正（unfolding）是什么？（可选）

Event Selection for data analysis

- [TrTrack](#) selection ([V6 alignment](#)+[GBL track fitting algorithm](#))
 - [InnerNHitY](#) >= 5 && L2 & (L3|L4) & (L5|L6) & (L7|L8)
 - [L1XY](#)
 - [L9XY](#) (for FS analysis)
 - [InnerNormChisqY](#) < 10
 - [L1InnerNormChisqY](#) < 10 && [L1InnerChisqY](#) - InnerChisqY < 10 (for [L1Inner](#) analysis)
 - [L1InnerL9NormChisqY](#) < 10 && ([L1InnerL9ChisqY](#) - InnerChisqY) / 2 < 10 (for FS analysis)
- Charge selection $Z \geq 9$ with tracker charge based on YJ's calibration:
 - $\text{fabs}(q_{\text{inn}} - Z) < (Z \geq 14) ? 0.5 : 0.0075 * \text{pow}(Z, 1.414) + 0.198;$
 - $Z - 0.0585 * \text{pow}(Z, 1.15) - 0.35 < q_{\text{l1}} < Z + 0.0334 * \text{pow}(Z, 1.15) + 0.20$ && $\text{fabs}(q_{\text{l1x}} - q_{\text{l1y}}) / (q_{\text{l1x}} + q_{\text{l1y}}) < 0.2$ && $(q_{\text{status_l1}} \& 0x10013D) == 0$
 - $Z - 0.0284 * \text{pow}(Z, 1.15) - 0.17 < q_{\text{l9}} < Z + 0.0585 * \text{pow}(Z, 1.15) + 0.35$ && $\text{fabs}(q_{\text{l9x}} - q_{\text{l9y}}) / (q_{\text{l9x}} + q_{\text{l9y}}) < 0.2$ (for FS analysis)
 - $Z - 0.625 - 0.0225 * (Z - 9) < q_{\text{utof}} < Z + 1.5$
 - $q_{\text{ltof}} > Z - 0.625 - 0.0225 * (Z - 9)$ (for FS analysis)

AMS Acceptance

- 接收度涉及探测器对宇宙线事例的选择效率，真实情况下探测器对飞行数据的探测效率和 MC 模拟存在稍许差别，需要根据各个子探测器的筛选条件对两者的差异进行修正。我们主要进行以下效率修正：
 - L1BZ: Tracker L1 对原子核击中位置和电荷的重建效率
 - L1PickUp: Tracker L1 对 inner tracker 径迹外推到 L1 的信号的重建效率和电荷筛选效率
 - UTOF 效率: UTOF 的电荷选择效率。
 - Inner tracker 效率: 分为 inner tracker 径迹重建效率和电荷筛选效率