

# 数据流说明

Pipeline程序 是一个 **线性** 的数据处理 **管道**

数据依次经过各层级模块，原始数据作为第一个模块的输入，每个模块都会产生对应的各级数据输出，上一级的输出又作为下一级的输入。最后一个模块产生的数据将被作为最终的科学数据产品，直接用于天文分析。

1. **输入**：原始 .dat 文件，二进制字节流文件。
2. **管道 1 - 数据包提取**：
  - 输入：**原始数据** .dat 文件。
  - 处理：提取LVDS包、应用层包、特征量模式 / 高通量模式 / housekeeping 数据帧。
  - 输出：FITS文件，包含 EVENTS 、 SPECTRA 和 HOUSEKEEPING 的数据包，数据仍然是字节流。
3. **管道 2 - 数据包解析**：
  - 输入：**管道 1 输出**的FITS文件。
  - 处理：解析字段（事例的 UTC、ADC值、能谱与载荷硬件监测数据等）。
  - 输出：FITS文件，包含解析后的字段（UTC / ADC），浮点数格式。
4. **管道 3 - 事件重建**：
  - 输入：**管道 2 输出**的FITS文件。
  - 处理：校准时间和能量。
  - 输出：FITS文件，包含每个通道的事件数据（光子精准时间、能量）。
5. **管道 4 - 文件格式化**：
  - 输入：**管道 3 输出**的FITS文件。
  - 处理：生成标准FITS文件，包含光子的精准时间、能量区间
  - 输出：最终科学数据产品，可直接用于天文分析。

## 各级输入输出文件的组织形式

### 输入文件

- **文件类型**：原始数据文件（.dat 格式）。
- **内容**：包含GRID-11B探测器采集的低电压差分信号（LVDS）数据包，内嵌应用层数据包，包括科学数据（事件、能谱）和housekeeping数据（温度、电压、电流、星敏、GPS等）。
- **说明**：作为数据处理流水线的起点，提供未经处理的探测器原始数据。

# 各级输出文件的含义及FITS组织结构

程序通过四个处理阶段生成不同级别的输出文件，均为FITS格式。以下按阶段说明，重点介绍最后两级（事件重建和文件格式化）的输出文件，因为它们包含直接可用于天文分析的信息。

## 阶段1：数据包提取

- 输出文件：
  - 文件名： {short\_name}\_pac\_{原始文件名}.fits （如 11B\_pac\_200322671\_JL1PT02A03\_20241120001208\_tianghe.fits ）。
  - 路径：存储在 output/packets 目录。
- 含义：
  - 提取原始数据中的LVDS数据包，分类为事件（EVENTS）、能谱（SPECTRA）和 housekeeping（HOUSEKEEPING）数据。
  - 这些数据包尚未解析，仅保留原始字节流及基本时间信息。
- FITS文件组织：
  - 主HDU（PRIMARY）：空。
  - 扩展HDU（每个数据类型一个）：
    - HDU **EVENTS**：包含特征量模式的事件数据包。
      - 列： UTC （整数，单位：秒）、 TIMESTAMP （长整数，时间戳）、 PACKET （字节数组，528字节）。
    - HDU **SPECTRA**：包含能谱数据包。
      - 列： UTC （整数，单位：秒）、 TIMESTAMP （长整数，时间戳）、 PACKET （字节数组，528字节）。
    - HDU **HOUSEKEEPING**：包含housekeeping数据包。
      - 列： UTC （整数，单位：秒）、 TIMESTAMP （空）、 PACKET （字节数组，187字节）。
- 用途说明：此阶段输出为中间产物，包含原始数据包，尚未提取具体物理量，不直接用于天文分析。

## 阶段2：数据包解析

- 输出文件：
  - 文件名： {short\_name}\_unp\_{原始文件名}.fits （如 11B\_unp\_200322671\_JL1PT02A03\_20241120001208\_tianghe.fits ）。
  - 路径：存储在 output/unpacked\_data 目录。

- **含义：**
  - 解析阶段1的FITS文件中的数据包，提取具体字段（如时间、通道、ADC值、能谱等）。
  - 按数据类型（EVENTS、SPECTRA、HOUSEKEEPING）组织解析结果。
- **FITS文件组织：**
  - **主HDU (PRIMARY)：** 空。
  - **扩展HDU**（每个数据类型一个）：
    - **HDU EVENTS：**
      - 列：
        - UTC（整数，单位：秒）：事件发生时间。
        - TIMESTAMP\_REF（长整数）：参考时间戳。
        - TIMESTAMP\_OFFSET（整数）：时间偏移。
        - CHANNEL（整数）：探测器通道（0、1、2）。
        - ADC\_VALUE（整数）：未校准的ADC值。
    - **HDU SPECTRA：**
      - 列：
        - UTC、TIMESTAMP\_REF、TIMESTAMP\_OFFSET、CHANNEL（同上）。
        - LONG\_SPECTRA（82元素数组）：长能谱数据。
        - SHORT\_SPECTRA（20×8矩阵）：短能谱数据。
    - **HDU HOUSEKEEPING：**
      - 列：
        - UTC（整数，单位：秒）。
        - SIPM\_TEMP 0-3（整数，单位：0.01K）：各通道温度。
        - VMON 0-3（整数，单位：mV）：电压监测值。
        - IMON 0-3（整数，单位：uA）：电流监测值。
        - SCTR\_1\_Q1-Q4、SCTR\_2\_Q1-Q4（整数）：星敏数据（姿态四元数）。
        - LONGITUDE、LATITUDE（整数，单位：0.01度）：星下点经纬度。
- **用途说明：** 此阶段输出为解析后的字段数据，包含初步的时间和ADC值，但未进行时间和能量校准，不直接用于天文分析。

## 阶段3：事件重建

- **输出文件：**
  - **文件**
    - 名： {short\_name}\_tte\_{时间范围}\_preliminary.fits（如 11B\_tte\_2501060000\_2501062359\_preliminary.fits）。
  - **路径：** 存储在 detector.output\_path/paths.path\_events 目录。
- **含义：**

- 对阶段2的EVENTS数据进行时间和能量校准，生成精确的事件数据。
- 每个通道的事件包含精确时间（UTC）、校准后的ADC值（ADC\_CALIBRATED）和能量（ENERGY）。
- **FITS文件组织：**
  - **主HDU (PRIMARY)：**空。
  - **扩展HDU**（每个通道一个，不再存储特征量模式以外的信息）：
    - **HDU EVENTS0, EVENTS1, EVENTS2, EVENTS3**
      - 列：
        - UTC（浮点数，精确度：微秒）：精确时间（UTC时间戳，校准后）。
        - ENERGY（浮点数，单位：keV）：光子能量。
        - ADC\_VALUE（整数）：原始ADC值。
        - ADC\_CALIBRATED（浮点数）：校准后的ADC值。
- **用途说明：**
  - 提供光子事件的时间（UTC）和能量（ENERGY），可用于初步的天文分析，如光变曲线、能谱分析。

## 阶段4：文件格式化

- **输出文件：**
  - **文件名：** G11\_evt\_{时间范围}.fits（如 G11\_evt\_2501060000\_2501062359.fits）。
  - **路径：**存储在 detector.output\_path/paths.path\_evt\_files 目录。
- **含义：**
  - 将阶段3的事件数据格式化为标准化的科学数据产品，包含完整的元数据和结构化的事件信息。
  - 提供能量边界（EBOUNDS）、有效时间区间（GTI）和事件数据，符合天文分析标准。
- **FITS文件组织：**
  - **主HDU (PRIMARY)：**空。
  - **扩展HDU**（不再存储特征量模式以外的信息）：
    - **HDU EBOUNDS：**
      - 列：
        - Channel（整数）：能量通道编号（1到N）。
        - E\_MIN（浮点数，单位：keV）：通道能量下限。
        - E\_MAX（浮点数，单位：keV）：通道能量上限。
      - 用途：定义能量通道的边界。
    - **HDU GTI**（Good Time Intervals）：
      - 列：
        - START（浮点数，单位：秒）：观测开始时间（相对于2018-01-01T00:00:00）。

- STOP (浮点数, 单位: 秒): 观测结束时间。
- 用途: 观测计划指定的有效观测时间区间, 用于时间筛选。对于11B来说, 由于24小时开机, 实际上有效观测时间就是从文件开始到文件结束的全部时间。
- HDU **EVENTS{channel}** (每个通道一个, 如 EVENTS0、EVENTS1、EVENTS2):
  - 列:
    - TIME (浮点数, 单位: 秒): 事件时间 (相对于2018-01-01 T 00:00:00)。
    - PI (整数): 能量通道编号 (基于 EBOUNDS 的能量分区的编号)。
    - DEAD\_TIME (字节, 单位: 微秒): 死时间 (固定为4微秒)。
    - EVT\_TYPE (字节): 事件类型, 用于判断事件的能量是否超出了 EBOUNDS 定义的所有能量范围 (0: 能量低于最低通道; 1: 正常事件; 2: 能量高于最高通道)。
  - 用途: 提供光子事件的时间和能量通道信息, 适合能谱和光变曲线分析。
- 头信息 (Header):
  - 包含元数据: DATE (文件创建时间)、FILE\_VER (版本, 如 PRELIMINARY)、MISSION (GRID)、CUBESAT (GRID-11B)、DATE\_OBS (观测开始时间)、DATE\_END (观测结束时间)、DATE\_REF (参考时间2018-01-01 T 00:00:00)。
- 用途说明:
  - 这是最终的科学数据产品, 适合直接用于天文分析。

## 输出文件总结

- 阶段1和阶段2: 中间产物, 不直接用于天文分析。
- 阶段3: 提供初步事件数据 (时间和能量), 可用于基本分析, 但缺乏标准化元数据。
- 阶段4: 最终科学数据产品, 包含标准化的时间、能量通道和元数据。