



# 一种适用于高通道数和高计数率下的粒子物理 实验软件触发算法

彭宇

中国科学院大学  
中国高能物理研究所 实验物理中心TDAQ组  
JUNO 合作组成员

第二十届全国科学计算与信息化会议

# 主要内容

## ➤ JUNO 实验

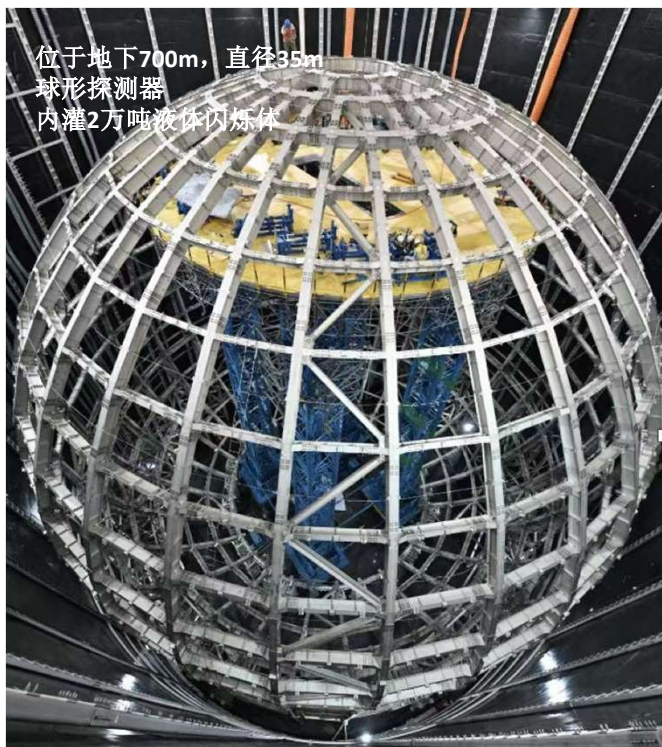
- 背景介绍
- TQ数据处理需求

## ➤ 多重数触发算法

- 算法基本原理
- 算法的实现
- 算法的改进

## ➤ 算法性能测试

# JUNO 实验——背景介绍

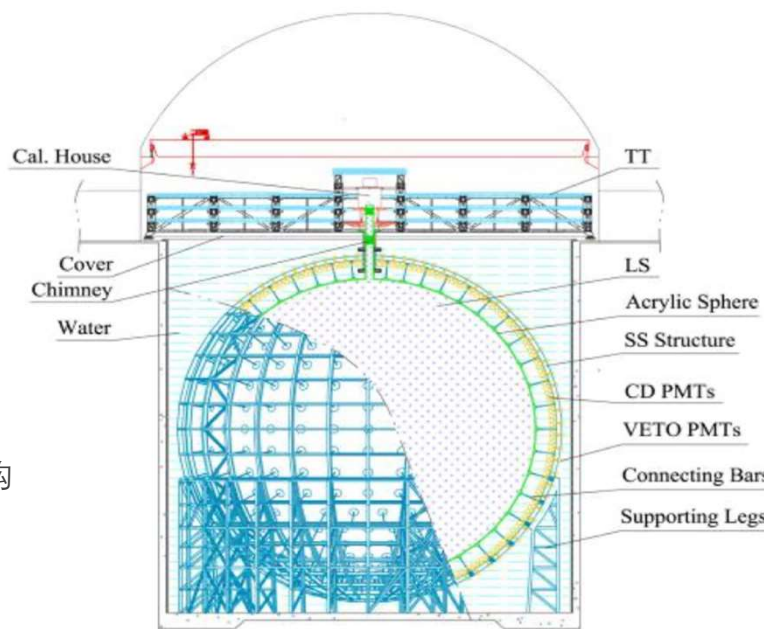


位于地下700m，直径35m  
球形探测器  
内灌2万吨液体闪烁体

中微子 物质世界的基本粒子之一

1998日本的超级神冈实验→中微子振荡  
2002加拿大SNO实验→太阳中微子振荡  
2012大亚湾中微子实验→第三种振荡模式

实验发现中微子具有微小质量，超出了标准模型的预言  
对中微子质量顺序的测量已经成为中微子研究的焦点之一



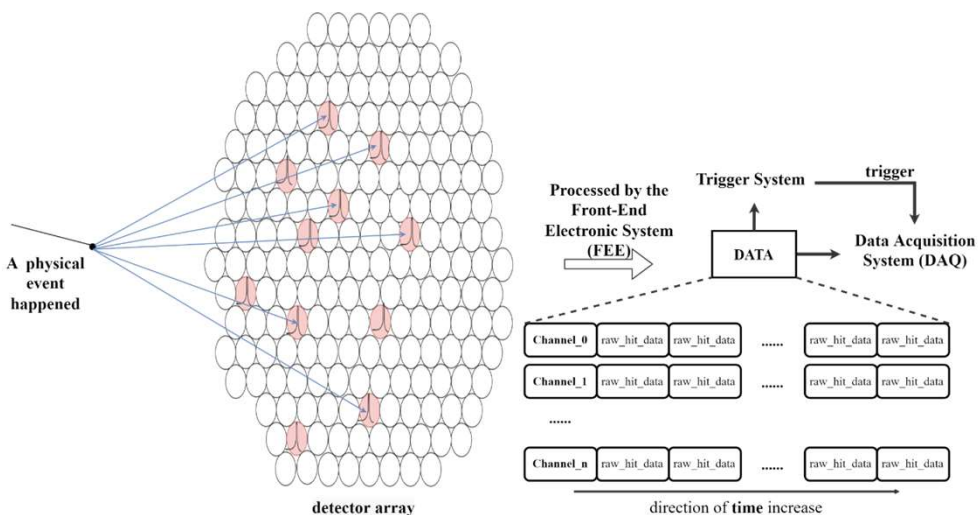
JUNO的探测器系统组成:

- 中央探测器 Central Detector, CD
  - 两万吨液体闪烁体
  - ~18000个20寸光电倍增管 LPMT (Large Photomultiplier Tubes)
  - ~25000个3寸光电倍增管 SPMT (Small Photomultiplier Tubes)
- 水切伦科夫探测器 Water Cherenkov Detector
  - ~2400个20寸光电倍增管
- 顶部径迹探测器 Top Tracker, TT
  - 塑料闪烁器阵列

高能所提出用反应堆中微子测量质量顺序的实验构想，既江门中微子实验。

江门中微子实验2015年建设，计划2023年底完成工程建设，2024年运行取数

# JUNO 实验——TQ数据处理需求



带电粒子在液体闪烁体中运动产生切伦科夫光，打到PMT上产生脉冲信号。为了精确测量，实验会保存完整的脉冲**波形数据**。同时针对超新星爆发时数据量暴增的场景设计了单独的**TQ数据流（无硬件触发）**。

- 当超新星爆发时→存TQ数据
- 正常物理取数→主要存波形数据，TQ数据需要**在线处理**以供超新星监测使用

TQ数据是粒子物理实验中一种很重要的数据格式

- T, 指时间信息
- Q, 指电荷信息

TQ数据记录了粒子与探测器相互作用产生的总电荷和作用发生的时刻

正常物理取数下，TQ数据数据量估计

detector	Channel	Rate	Data volume
CD LPMT-T/Q	~18000	~30 KHz	~5.4 GB/s
CD SPMT	~25000	~500 Hz	~375 MB/s

JUNO有大约**6GB/s**的TQ数据需要在线实时处理

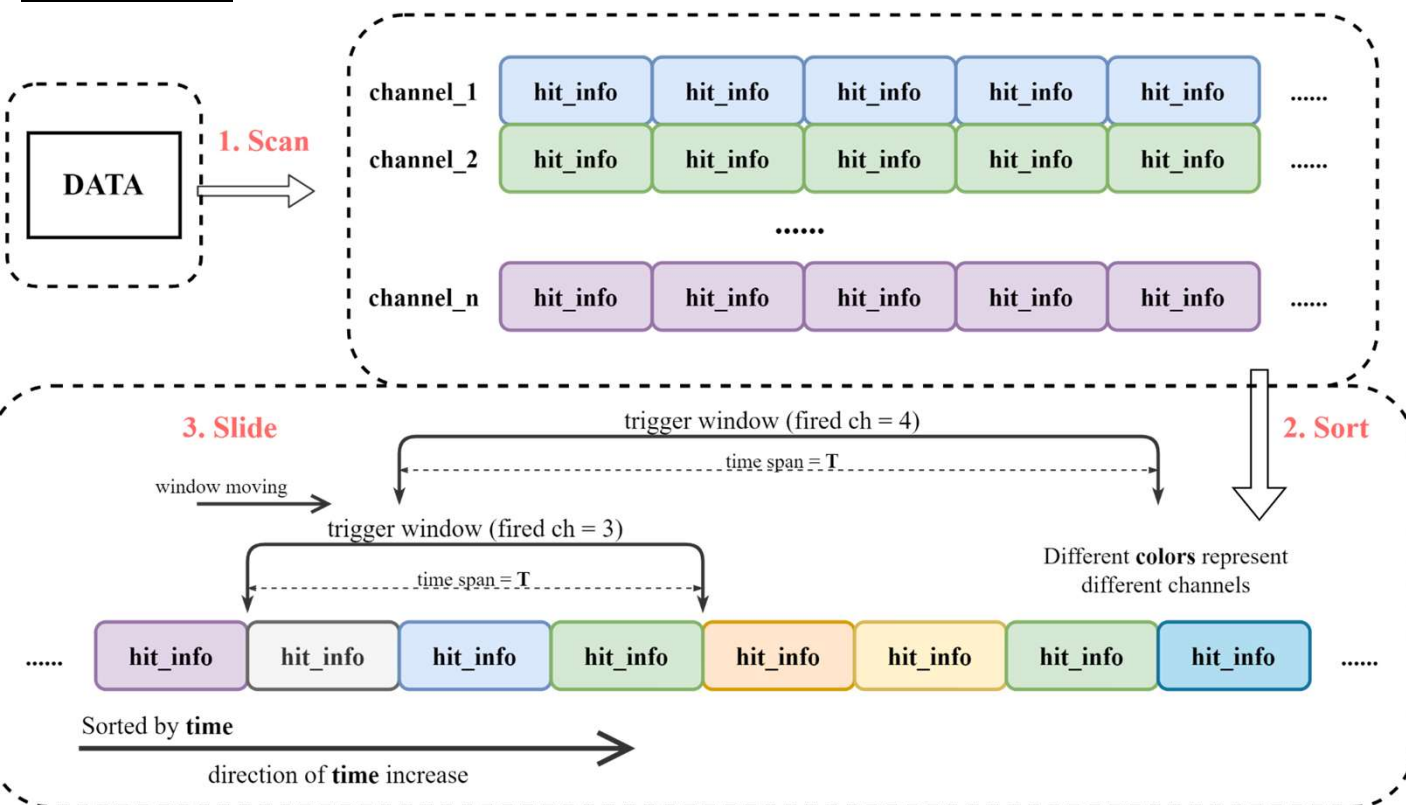
TQ在线处理包含很多步骤和处理流，但都需要**软件触发**作为最初的预处理

本报告的主要内容为介绍为TQ数据**在线处理**开发的**软件触发算法需求**：

- 能快速处理多通道（~20k）高单道计数率（~30k）的TQ数据

# 多重数触发算法——原理、实现

软件触发算法，需要将不同通道之间有关联（同一个事例）的TQ数据聚集在一起



事例：实验中将一次真实的粒子相互作用（衰变/散射）称为一个事例event

触发：后端系统从各个通道读出数据，通过一定手段分辨数据属于哪个事例，同时过滤噪声数据。

**多重数触发原理：**在固定时间窗口内，着火探测单元的数量大于某个阈值时，判定触发

**常规实现方式：**将数据按时间排序，在排序好的队列上开一个滑动窗口，统计窗口内的着火通道数。如果着火通道数高于设定的阈值，则判定触发

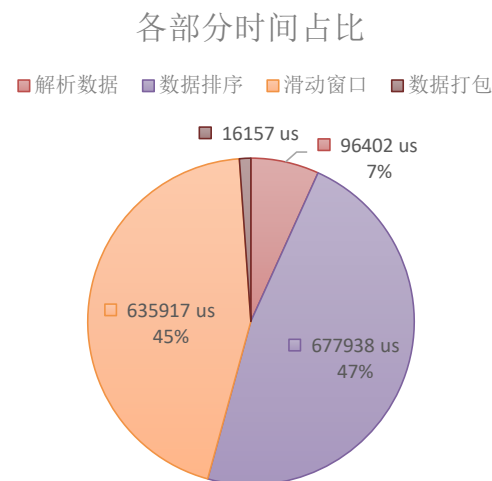
# 多重数触发算法——实现

## ■ 完整多重数触发实现（可分4部分）

- **解析数据**（从原始数据中提取时间信息）
- **数据排序**（按时间将TQ数据排序，使用std::sort）
- **滑动窗口**（在排序队列上滑窗，进行触发判定）
- **事例数据打包**（数据拷贝，将数据以事例为单位聚集）

### 测试环境：

- 64逻辑核
- CPU model: Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
- 内存: 384GB
- 操作系统: CentOS7.9
- 使用GCC8.3
- 打开编译优化-O2



江门CD-LPMT TQ数据量：

18000通道 × 30kHz = 540MTQ/s

处理10ms数据用时1426ms

**估算实时计算需要143核**

# 多重数触发算法——改进

改进方向：

- 减少排序时间
- 降低滑窗的时间

基本思路：使用**触发表**来**避免排序**（空间换时间）

**优点：**

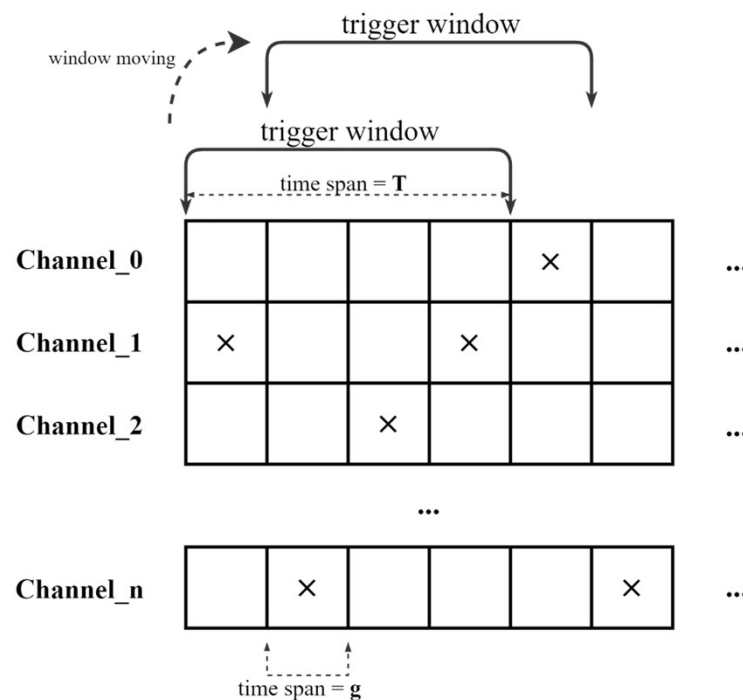
- 避免了数据排序，只要表格填完，着火通道数就可以直接数出。
- 滑窗操作数是固定的，只与表格大小相关

**缺点：**

- 内存消耗大。以JUNO实验为例，列单位为25ns时，处理10ms的TQ数据需要**7.2GB**内存
- 漏触发。滑窗时，触发窗并非在时域上连续滑动，而是以列宽度为间隔的跳动。在前后两个窗口之间满足触发阈值的事例可能会被漏触发。

**触发表：**

- 行，表示通道
- 列，表示时间段
- 表格值为布尔值，表示该通道在该时间段内有无着火



# 多重数触发算法——改进

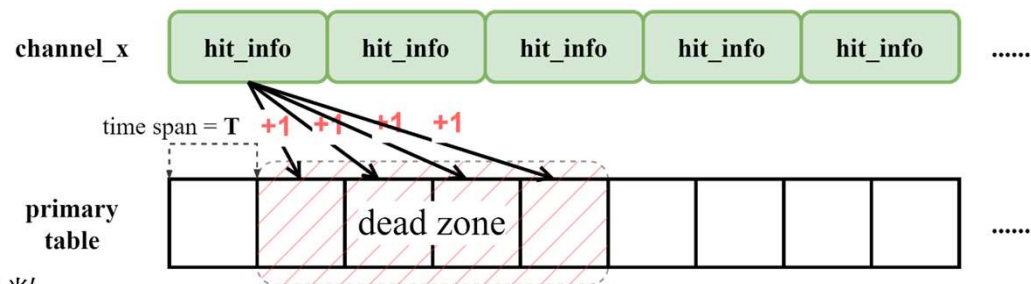
为了改善触发表的空间利用率，我们设计了新的表格，只用一行实现之前多行表格的功能

基本思路：不记录通道是否着火，而是直接记录着火通道数

优点：

- 提升了空间利用率。触发表的体积与通道数无关。以JUNO实验为例，表格值用4字节整数，触发窗口的滑动间隔为25ns时，处理10ms的TQ数据只需要**1.6MB**内存。
- 减少滑窗时间。表格体积的减小，导致滑窗需要的时间相应减少。同时表格的值即为着火通道数，滑窗只需遍历表格并进行一次比较即可完成，操作数也相应减少。

漏触发问题：漏触发可以通过减小触发窗口的滑动间隔解决。



一行触发表：

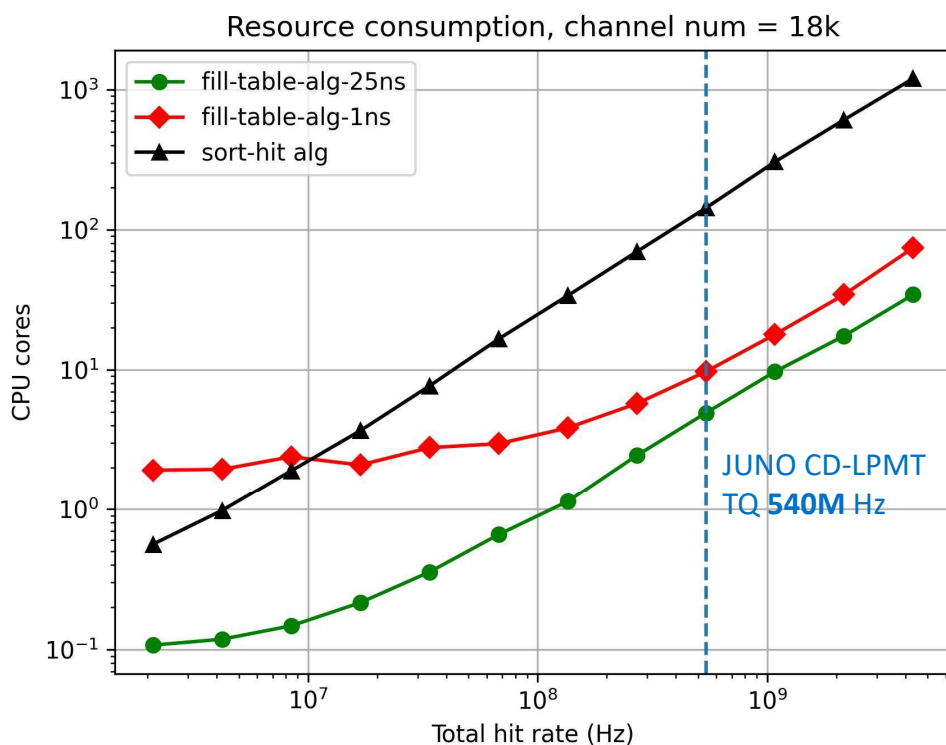
- 列，表示触发窗口
- 表格值为整数，表示在该触发窗口内着火通道数



# 算法性能测试

## 测试环境:

- CentOS7.9
- 使用GCC8.3
- 打开编译优化-O2
- CPU model: Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
- 内存: 384GB



左图是利用模拟数据得到的多重数触发算法三种不同实现在通道数固定为18K, 总计数率在[2M, 4320M] Hz 范围内的资源消耗

- Fill-table-alg-25ns 指触发窗口的滑动间隔为25ns
- Sort-hit alg 指使用排序来实现多重数触发

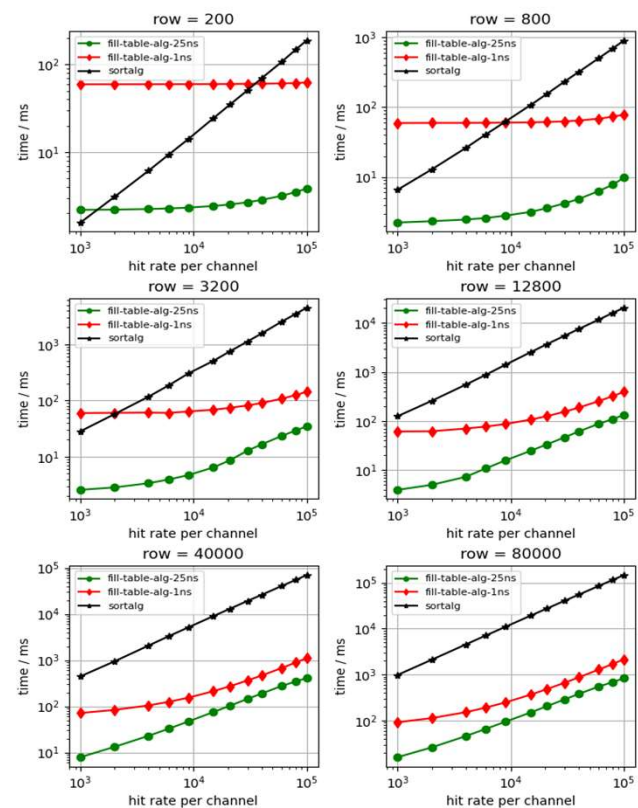
蓝色虚线指JUNO CD-LPMT TQ数据的数据量, 从图中可见, 改进算法能节约一个量级的CPU计算资源

# 算法性能测试 & 总结

## 报告总结:

1. 通过触发表替换排序，我们得到了一种更适合JUNO实验TQ在线处理的软件触发算法实现。
2. 通过对不同通道数和计数率TQ数据的模拟测试，我们发现使用触发表的算法实现很适合在高通道数和高计数率的场景中使用。
3. 多重数触发是很多实验里常用的算法，希望新的优化实现能帮助更多实验节省资源。

江晓照，将在12日14:00点核电子学与探测分会场，分享使用GPU加速多重数触发算法的报告。





**Thanks**