Online Data Quality Monitoring

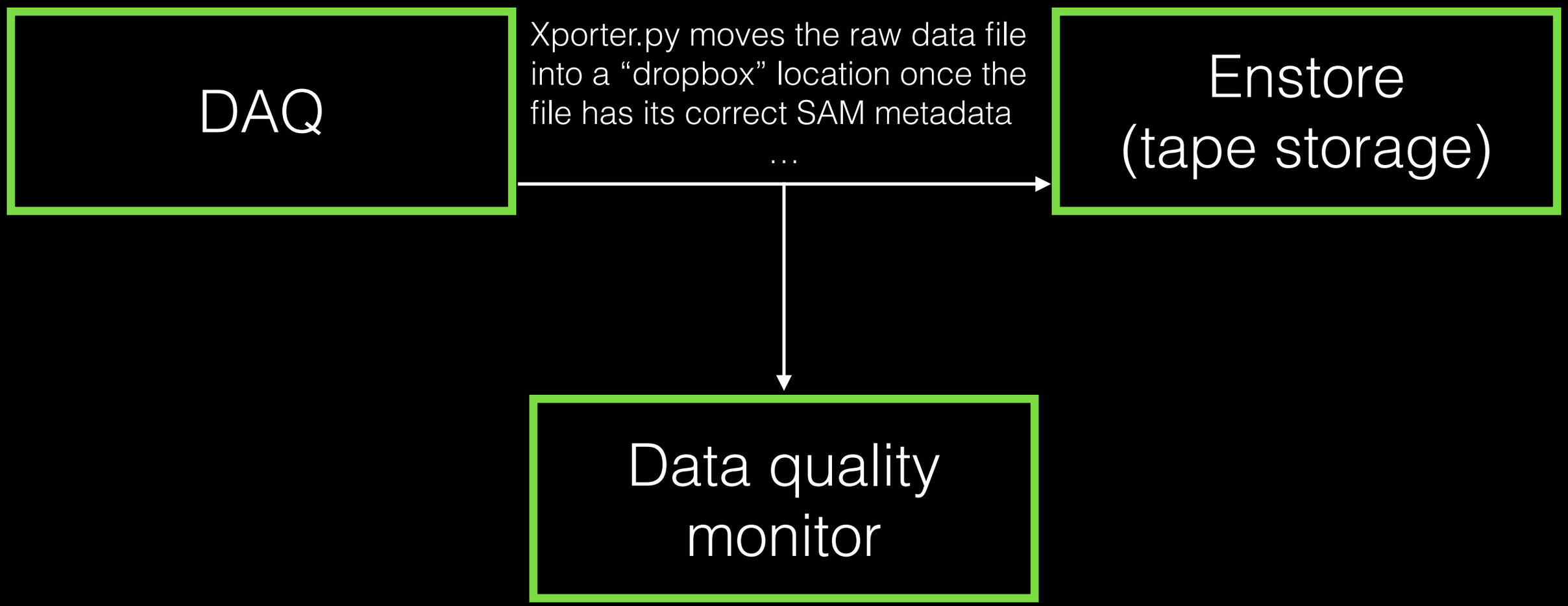
Johnny Ho
LArIAT Operational Readiness Review
13 October 2015

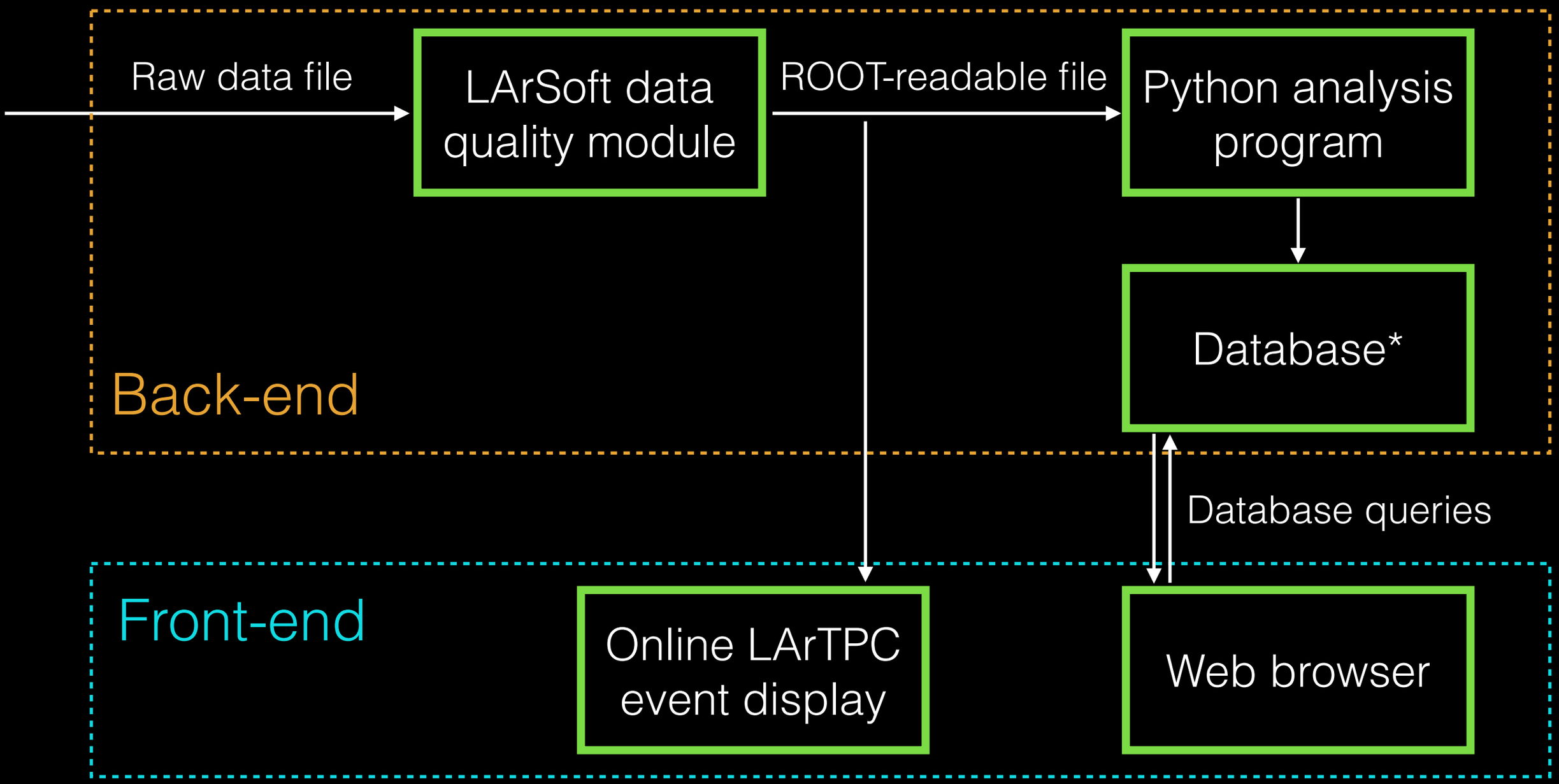
Online data quality monitor

- The online data quality monitor (DQM) lets us check the quality of the data we are writing to disk in near-real-time
- It lets us view a lot of low-level information in the data as we are running such as
  - RMS noise* and pedestals in the readout electronics
  - Performance of detectors (TPC, time-of-flight, wire chambers, etc.)
- Allows us to debug our beam and electronics right away if there is something wrong

* This was actually done offline during Run I. It will be automated in the online DQM for Run II.

Online data quality monitor





* Private database hosted on the LArIAT DAQ cluster for Run I.
We have requested a database from SCD for Run II.

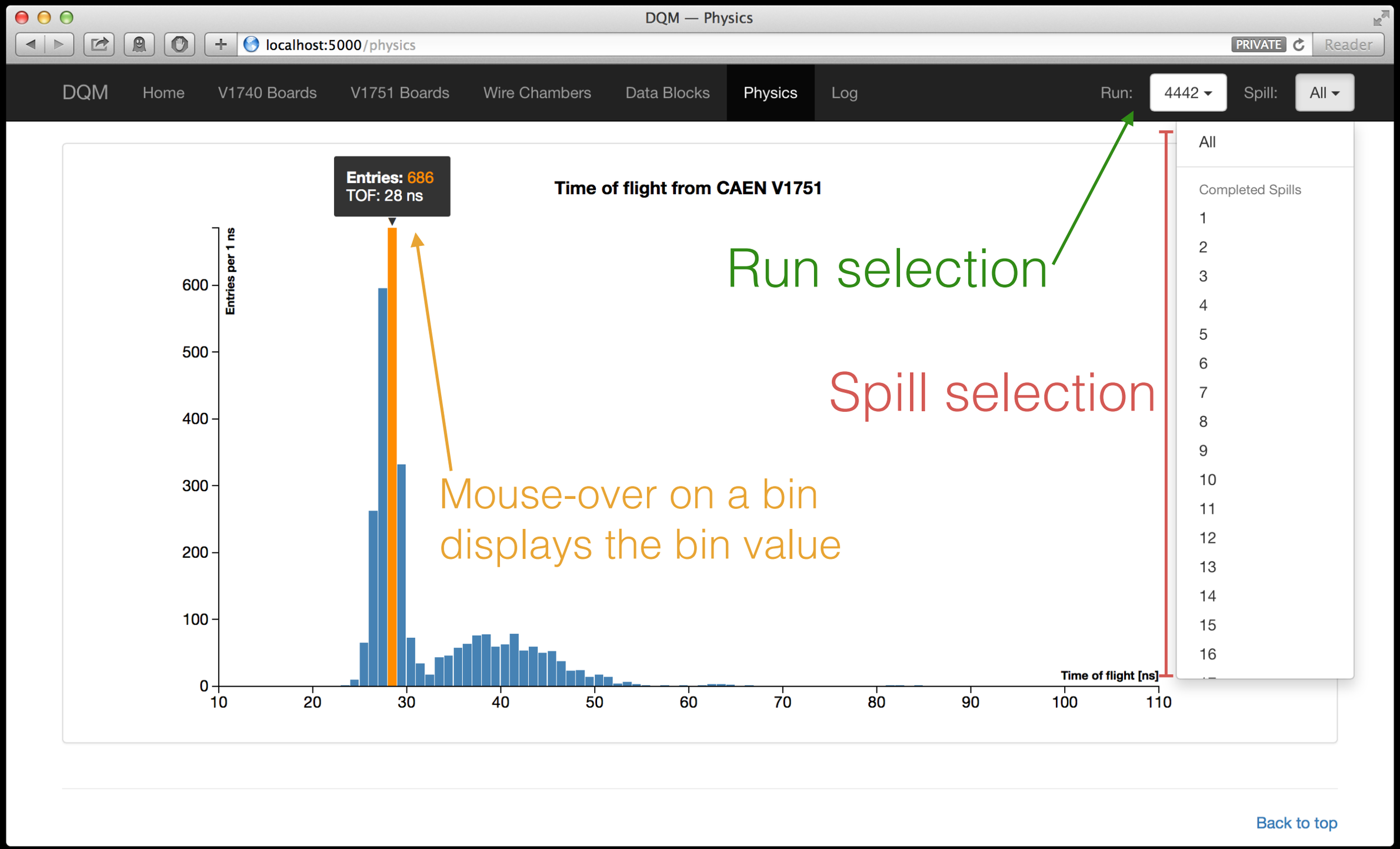
Data quality monitor front-end

- The main front-end of the online data quality monitor is an interactive website

- The website displays a set of low-level plots for each run or spill (this is user-selectable)

- If the user is interested in looking at the current run, the plots are automatically updated as the data comes in (updated every minute when there is an ongoing run)

Front-end: time of flight



Front-end: number of data blocks

DQM — Data Blocks

localhost:5000/data-blocks

DQM | Home | V1740 Boards | V1751 Boards | Wire Chambers | Data Blocks | Physics | Log | Run: 4442 | Spill: All

Number of data blocks from each device

#	Device	Number of data blocks
1	CAEN V1751 board 0	4358
2	CAEN V1751 board 1	4358
3	Multi-wire proportional chambers	4179
4	CAEN V1740 board 0	4005
5	CAEN V1740 board 1	4005
6	CAEN V1740 board 2	4005
7	CAEN V1740 board 3	4005
8	CAEN V1740 board 4	4005
9	CAEN V1740 board 5	4005
10	CAEN V1740 board 6	4005
11	CAEN V1740 board 7	4005
12	Wave union TDC	1496

Wire chambers are not reading out if this number is not increasing.

These numbers should be the same.

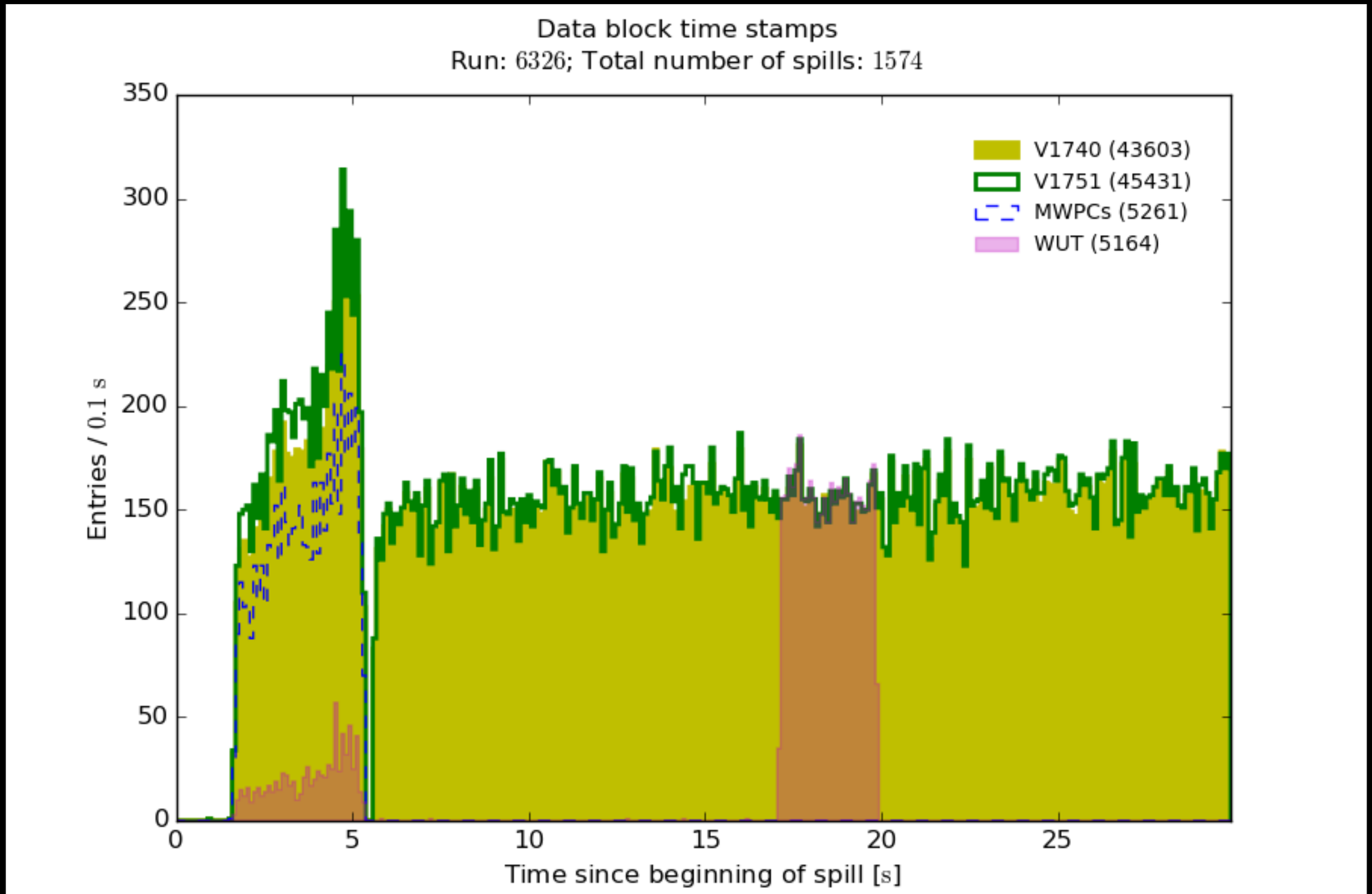
These numbers should be the same.

Data block time stamps

260 s

MWPC

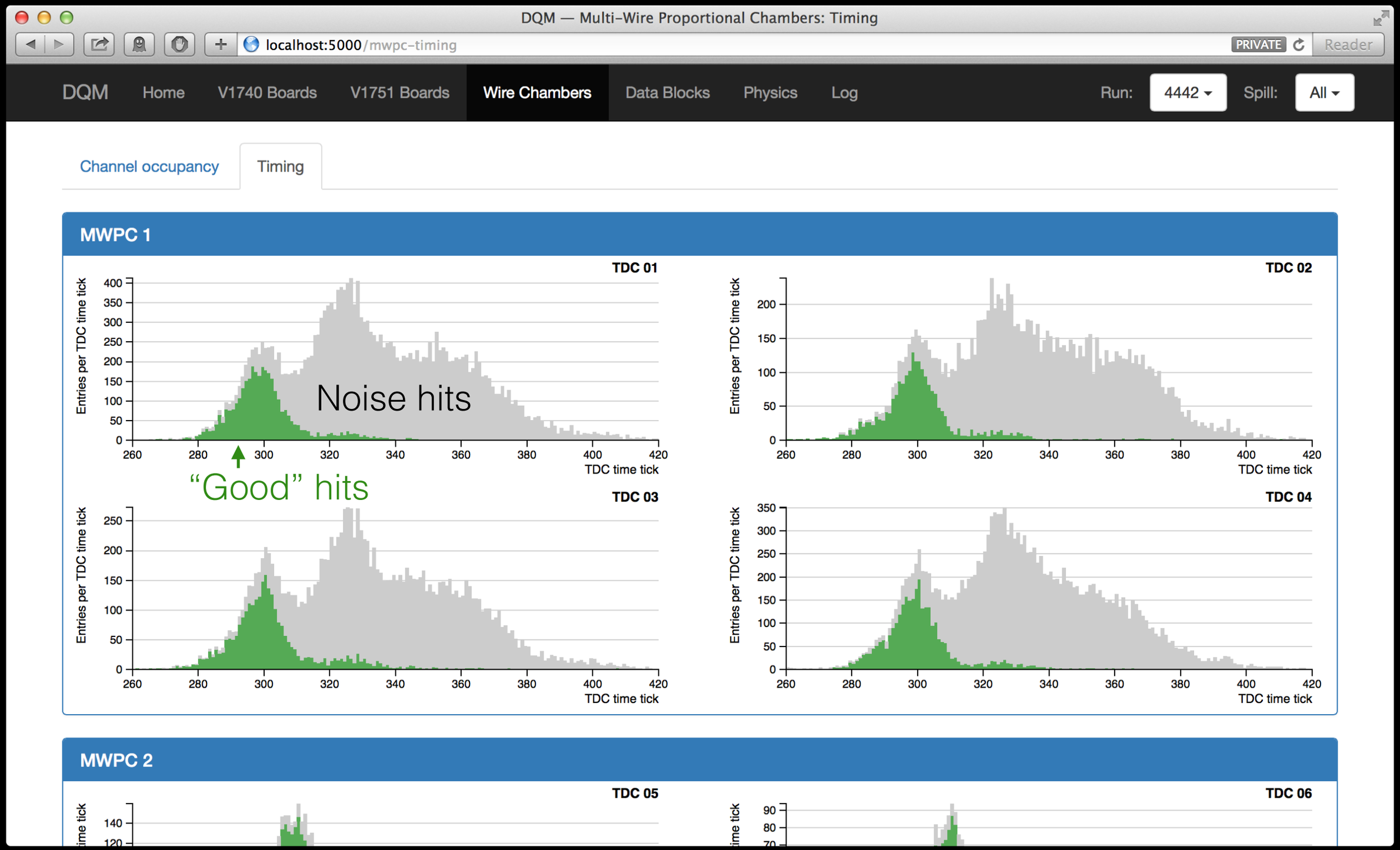
Front-end: timing of data blocks in the super-cycle



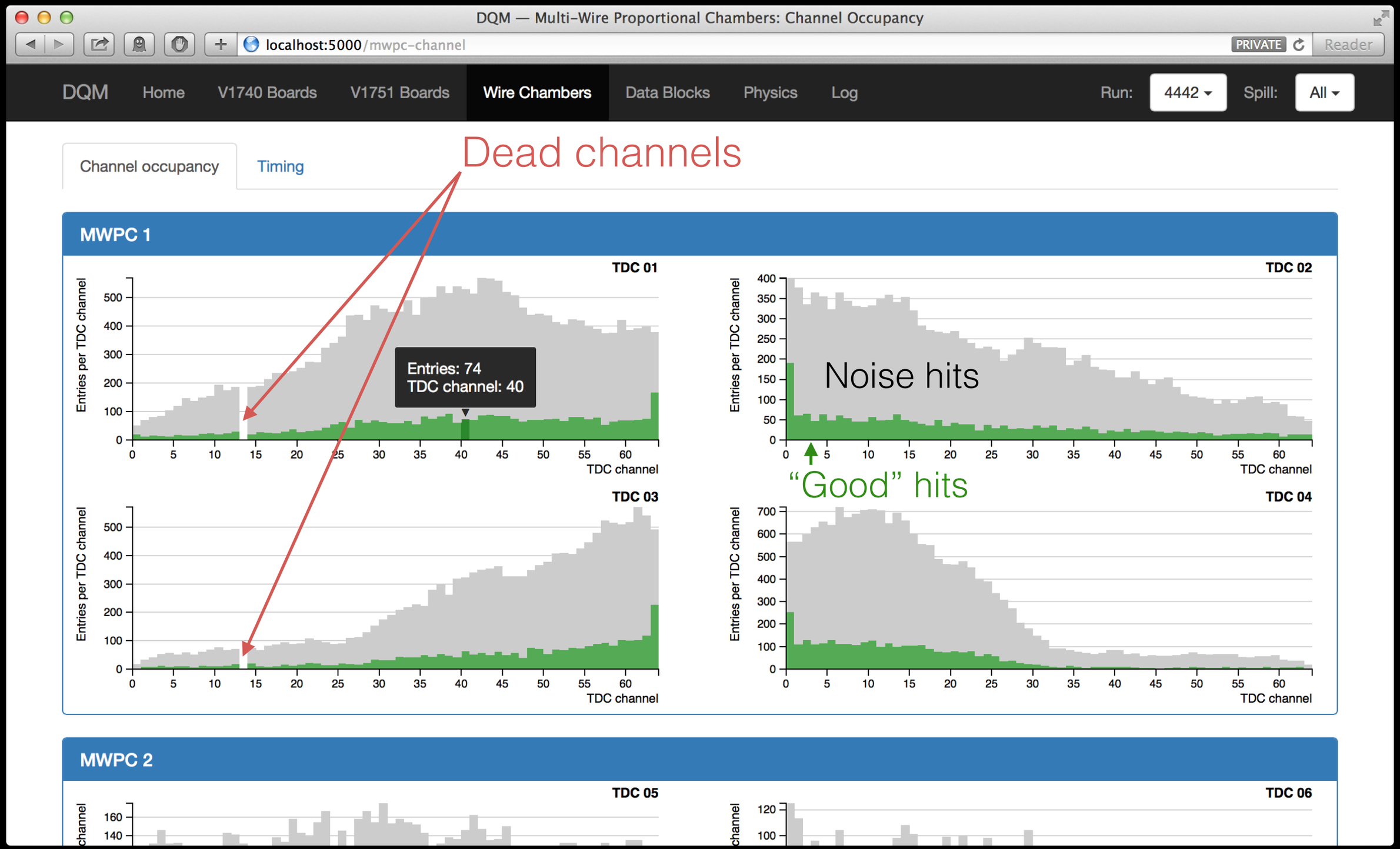
BEAMON

COSMICON

Front-end: wire chamber hit timing

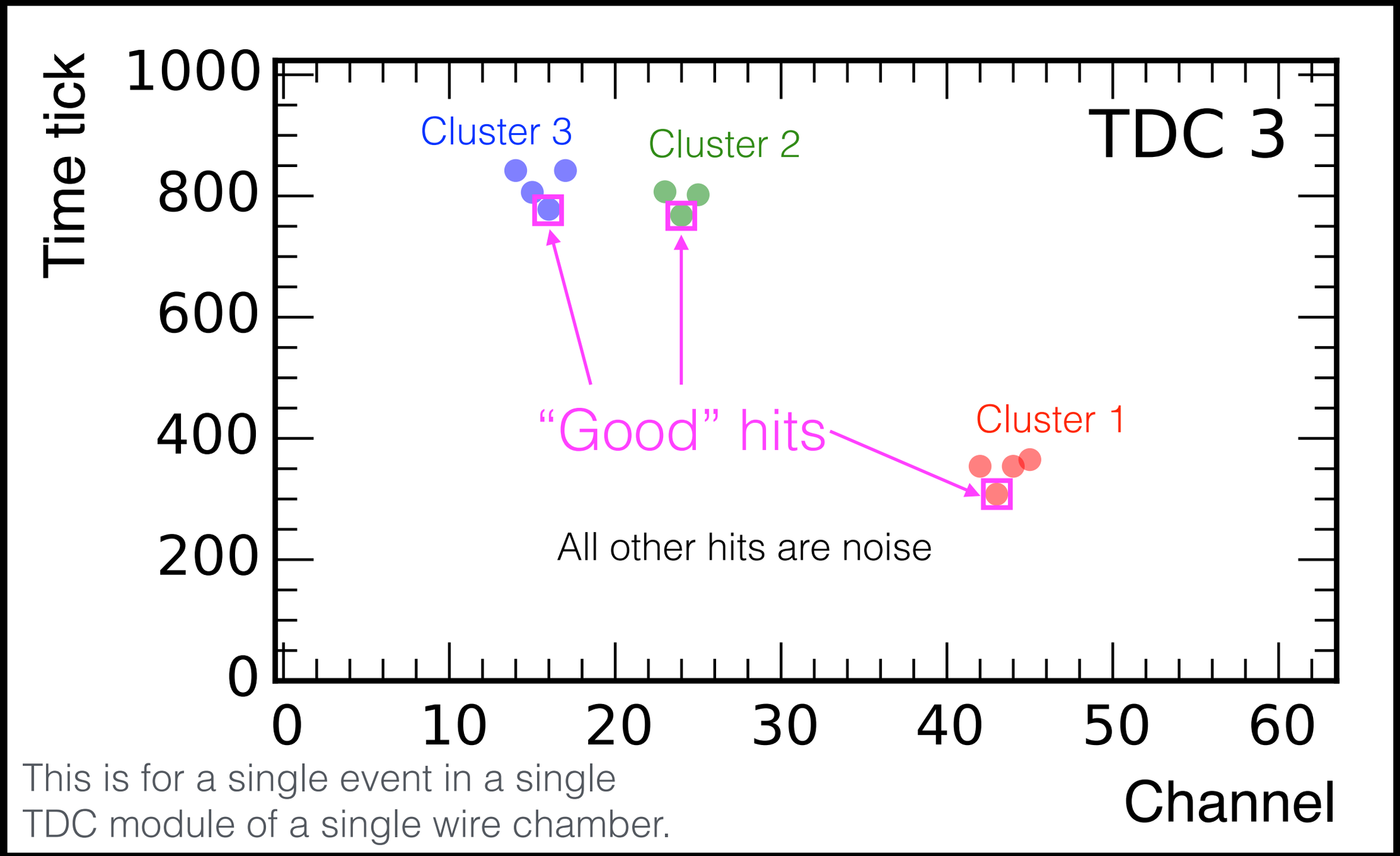


Front-end: wire chamber hit channel



Small peek of the back-end:

Clustering hits in wire chambers for noise removal



This is for a single event in a single TDC module of a single wire chamber.

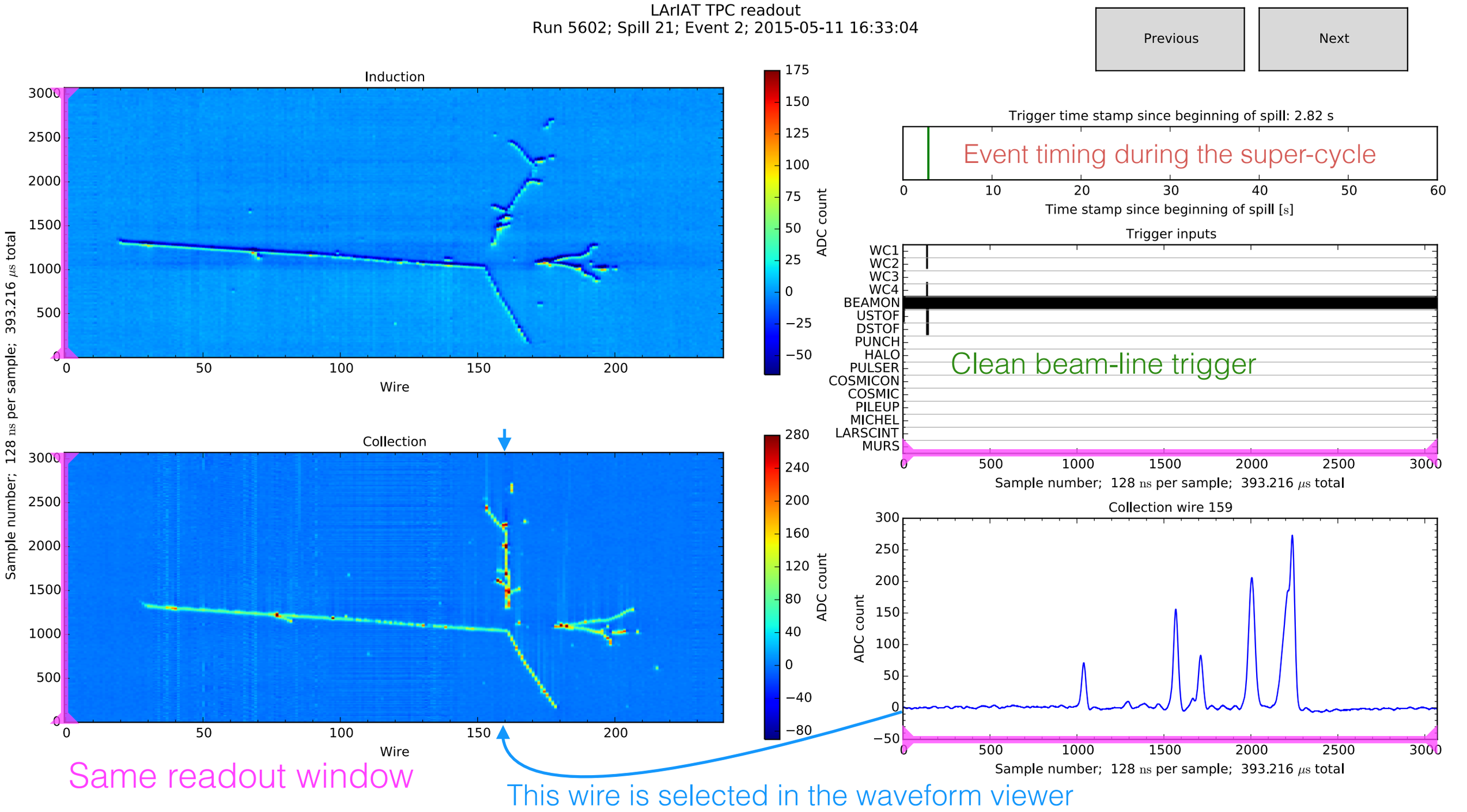
FTBF now knows how to deal with this noise in the wire chambers!

Online LArTPC event display

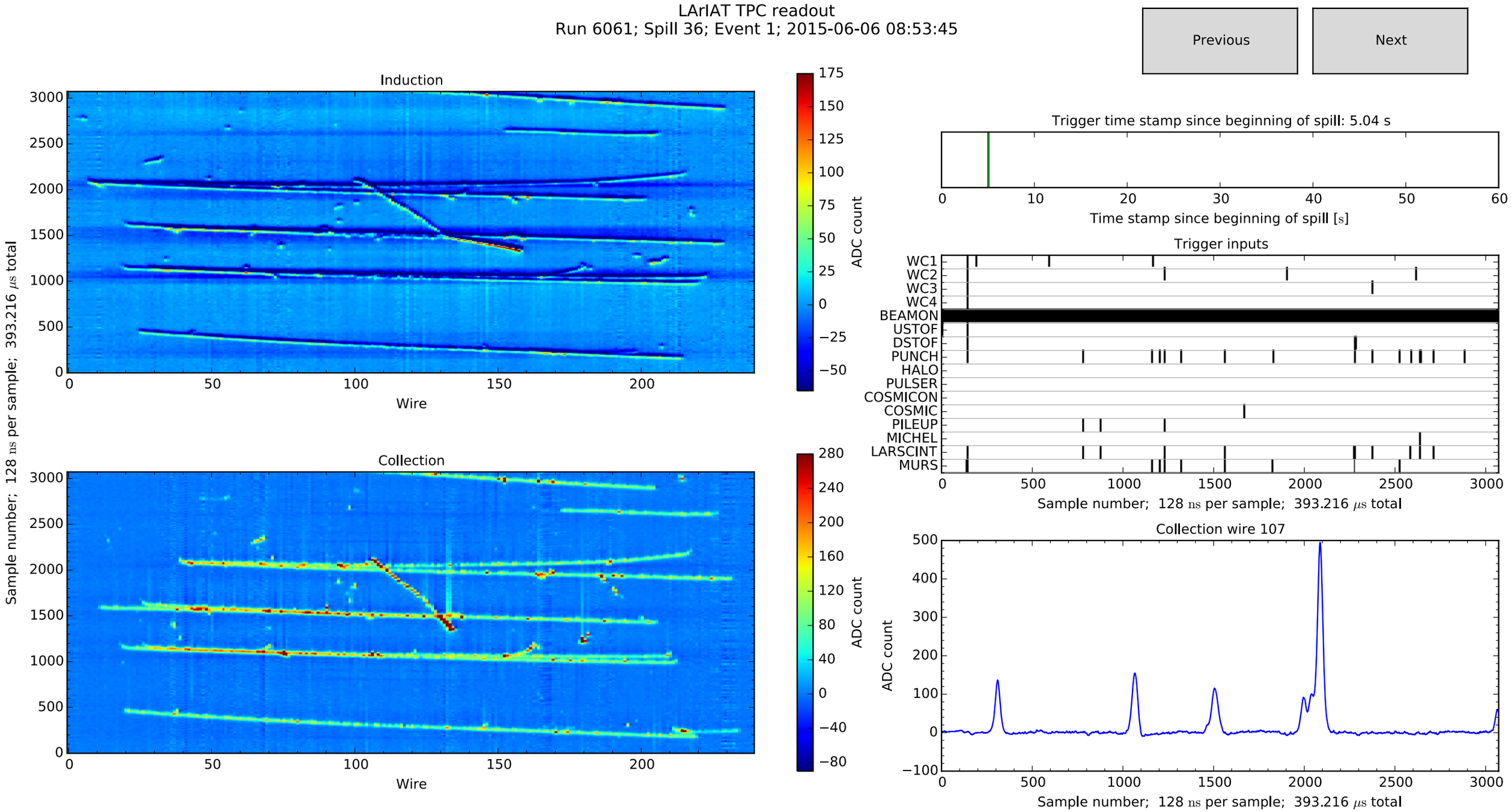
- The online TPC event display helps us decide whether we are getting good events in the TPC, i.e. no beam pile-up

- The display also shows what triggered the TPC readout, and helps us get feedback on our trigger configurations as we modify it

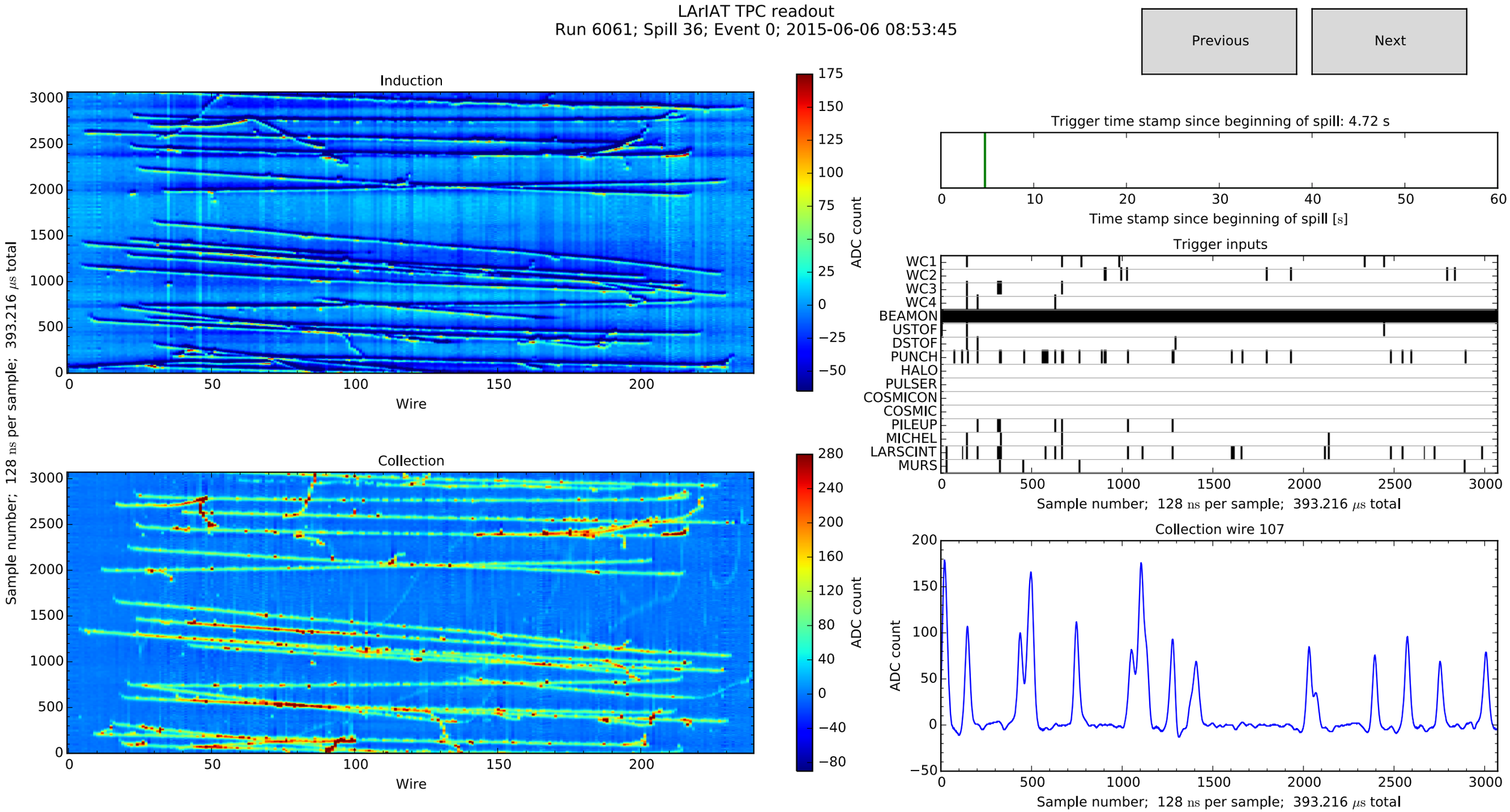
Online LArTPC event display:
Clean event, pion single charge exchange candidate



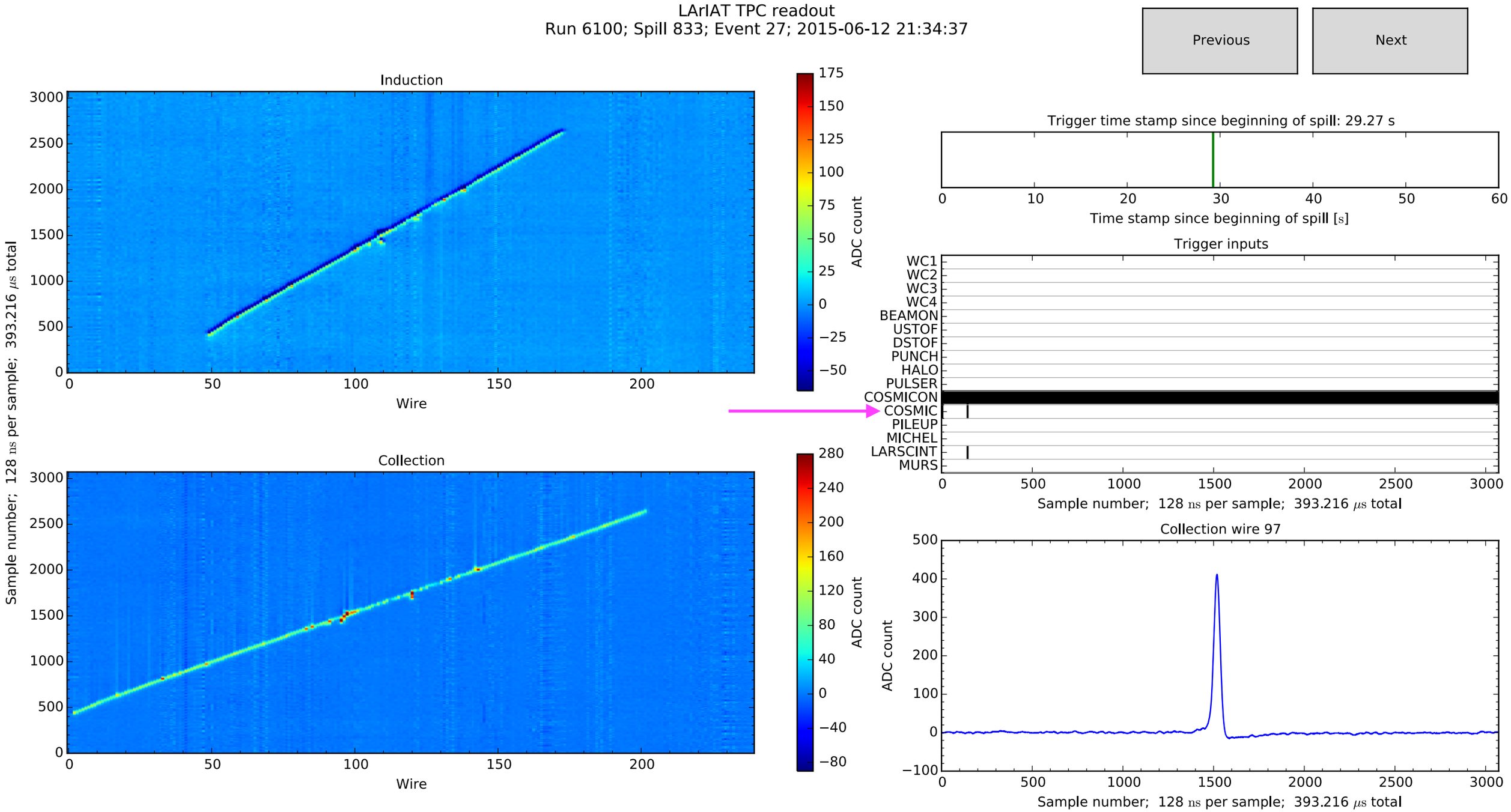
Online LArTPC event display: Pile-up



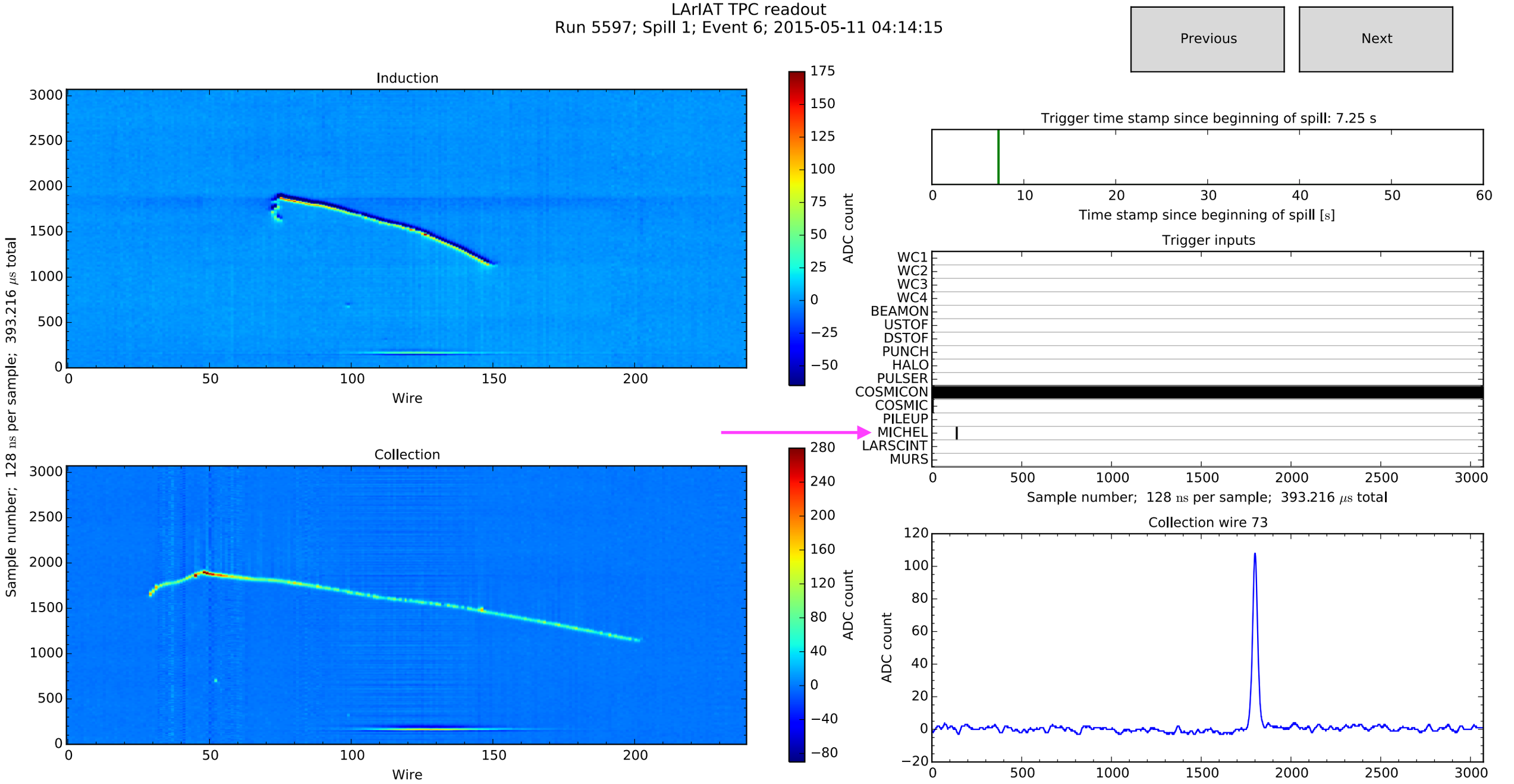
Online LArTPC event display: More pile-up



Online LArTPC event display: Through-going cosmic muon candidate



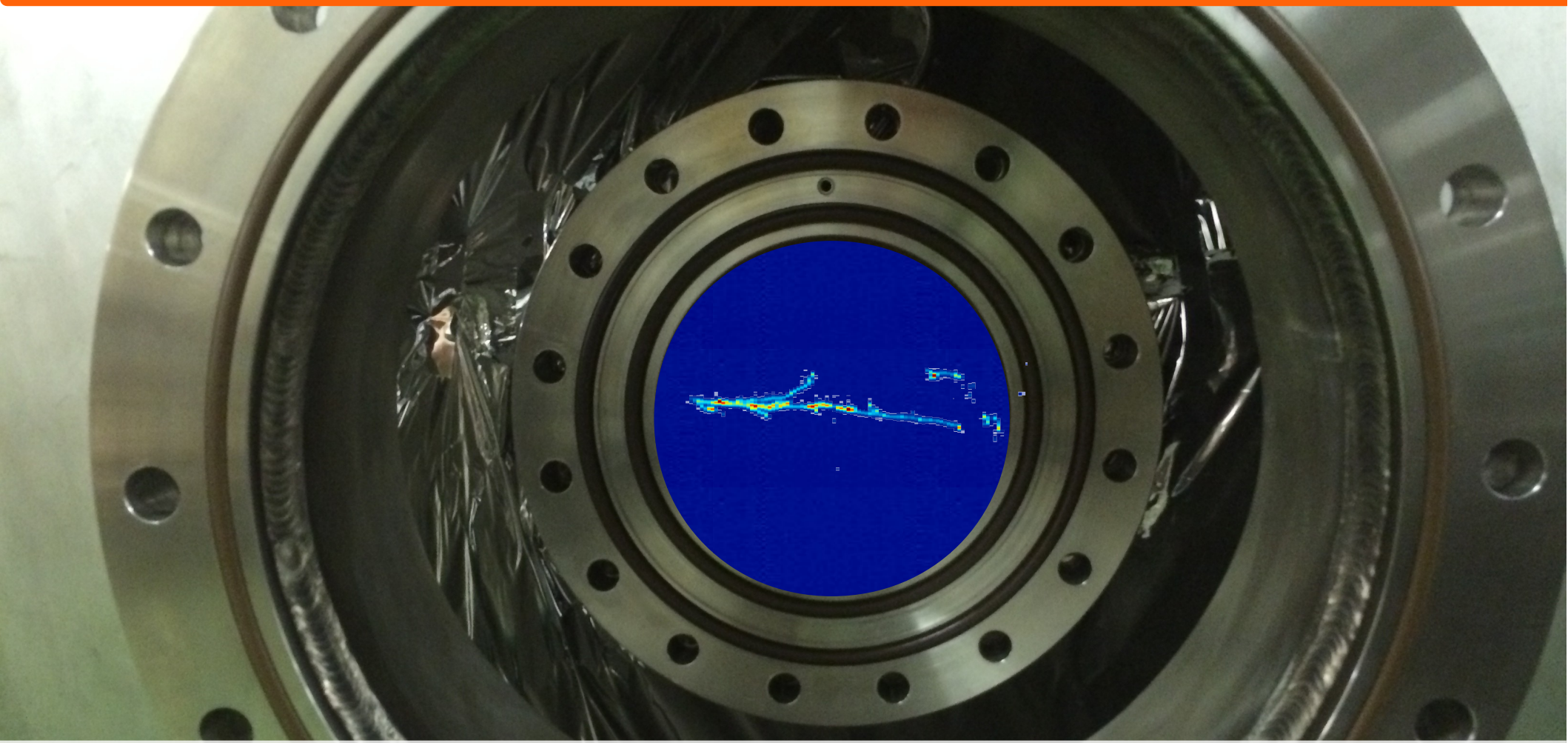
Online LArTPC event display: Michel decay candidate



Conclusion

- These data quality tools are extremely helpful in giving instant feedback on whether or not we are getting good, useful data as we are running
- Electronics behaving abnormally, poor beam conditions, etc. can be spotted right away so that the problems can be alleviated without wasting our precious liquid argon and beam time!

Liquid Argon in a Test Beam (LArIAT) Experiment



Offline Infrastructure & Data Processing

Jonathan Asaadi
University of Texas Arlington

Offline Infrastructure

- **LArIATsoft** is a collection of software modules built on liquid argon software package (**LArSoft**) for analyzing data collected by the LArIAT experiment
  - All of which is built upon the **art** framework
  - And within are many more tools used for accessing the data and running our code

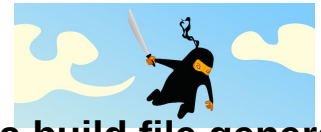
LArIATSoft
LArIAT Software Package

LArSoft
Liquid Argon Software Package

art
Event-Processing Framework

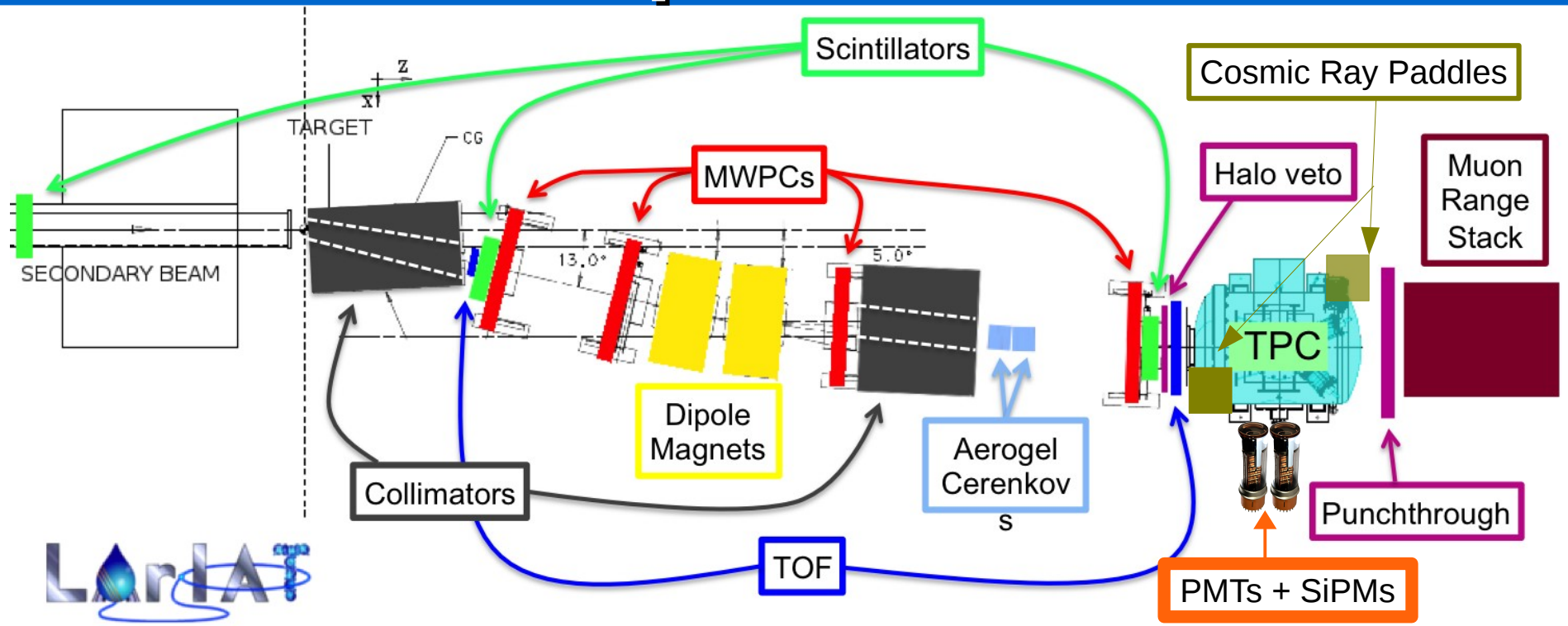
SAM
Sequential data **A**ccess via **M**eta-data

MRB
Multi-Repository Build System



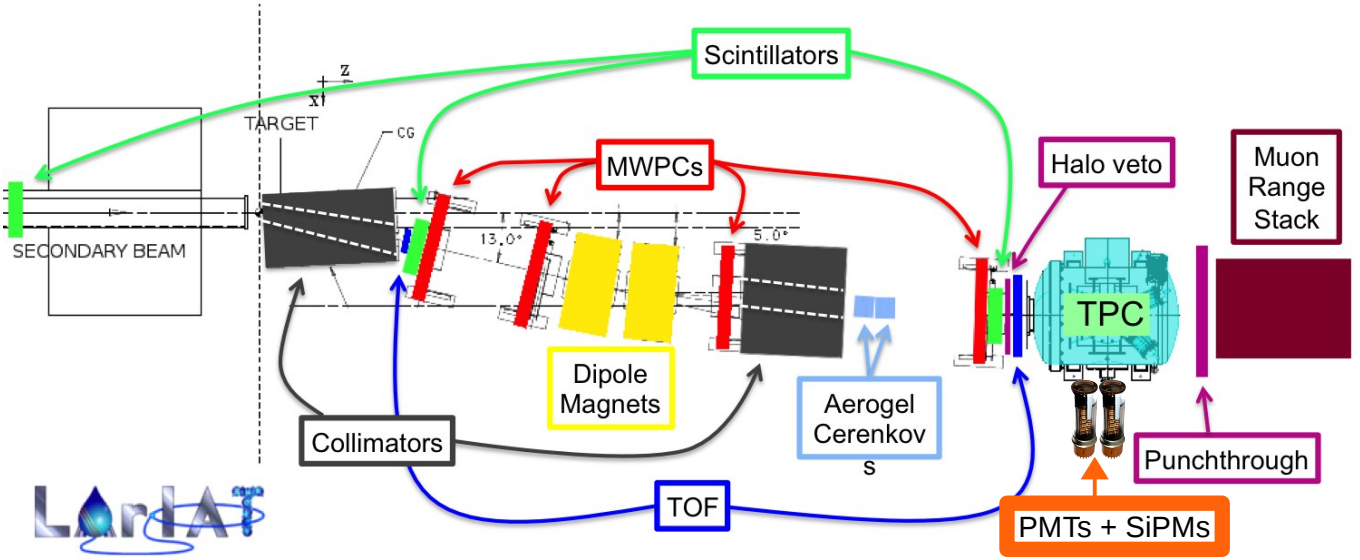
ninja build file generation system

What our experiment looks like



18 Detectors all read out in LArIAT DAQ

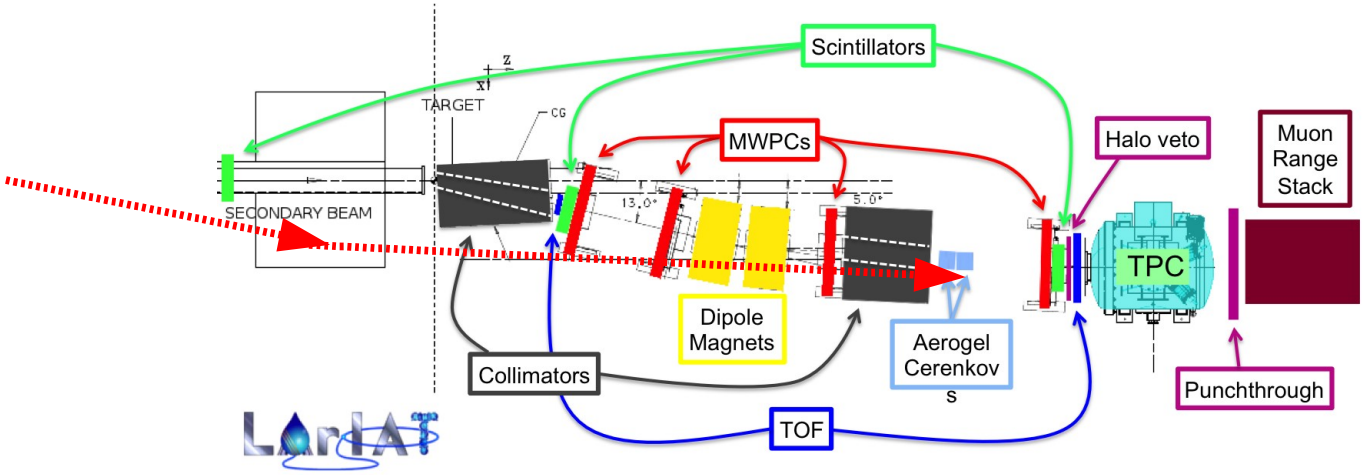
- **Two Time of Flight detectors (Upstream / Downstream)**
- **Two Cosmic Ray Paddles (Above and Below the TPC)**
- **Four Multi-Wire Proportional Chambers (MWPC)**
- **Two Aerogel Cerenkov Detectors**
- **Five LAr Light Detectors (3 SiPMs + 2 PMTs)**
- **One Muon Range Stack (16 Scintillator Paddles)**
- **Two Beamline Paddles (Halo Veto + Punchthrough)**
- **One LArTPC (480 wire channels)**



The readout of these detectors are known as "Fragments" and get turned into objects we call "Digits"

- **Detector Digits**
  - Auxiliary Detector Digits (AuxDetDigits)
  - Optical Detector Digits (OpDetPulses)
  - TPC Raw Wire (RawDigits)
  - Trigger Digits (TrigDigits)
- **Fragments from the CAEN 1751**
  - TOF, Aerogel, LAr-Light Detectors, Beam Halo-Veto
- **Fragments from the CAEN 1740**
  - LArTPC, Muon Range Stack
- **Fragments from the MWPC Controller**

What our data looks like when it comes out of the DAQ



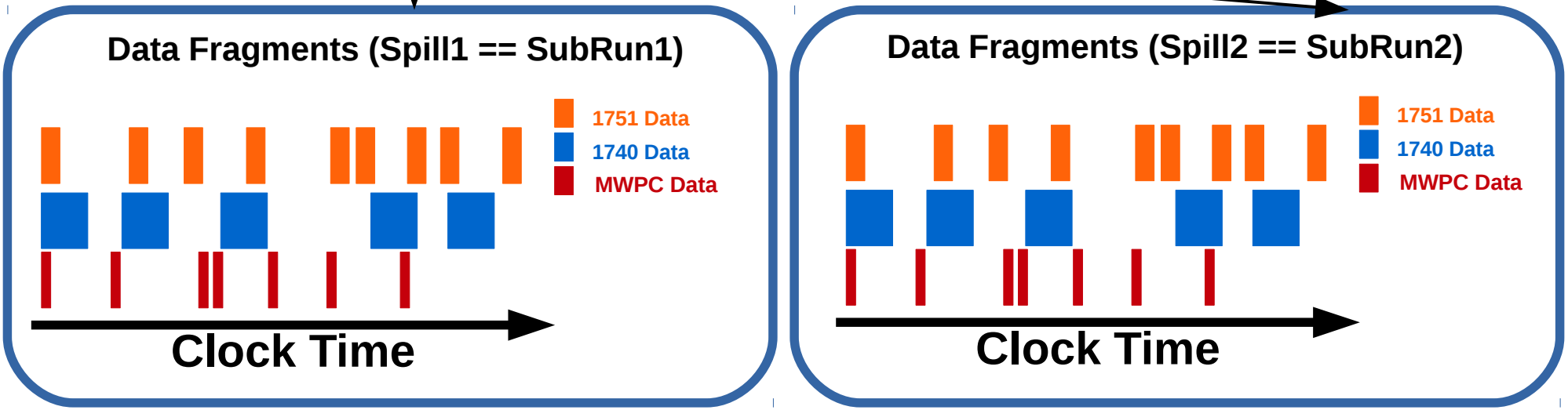
- **When we receive our beam, each 4+ second spill (along with the cosmic ray data taking period), is recorded as one long series of data fragments from the various readout**
  - The drift time of the TPC is 350 μs, meaning you can have multiple drift windows in one spill
- **Inside that one spill there are many triggers**
  - Each trigger is a predefined condition that causes the readout of of all the systems

Raw Data Structure

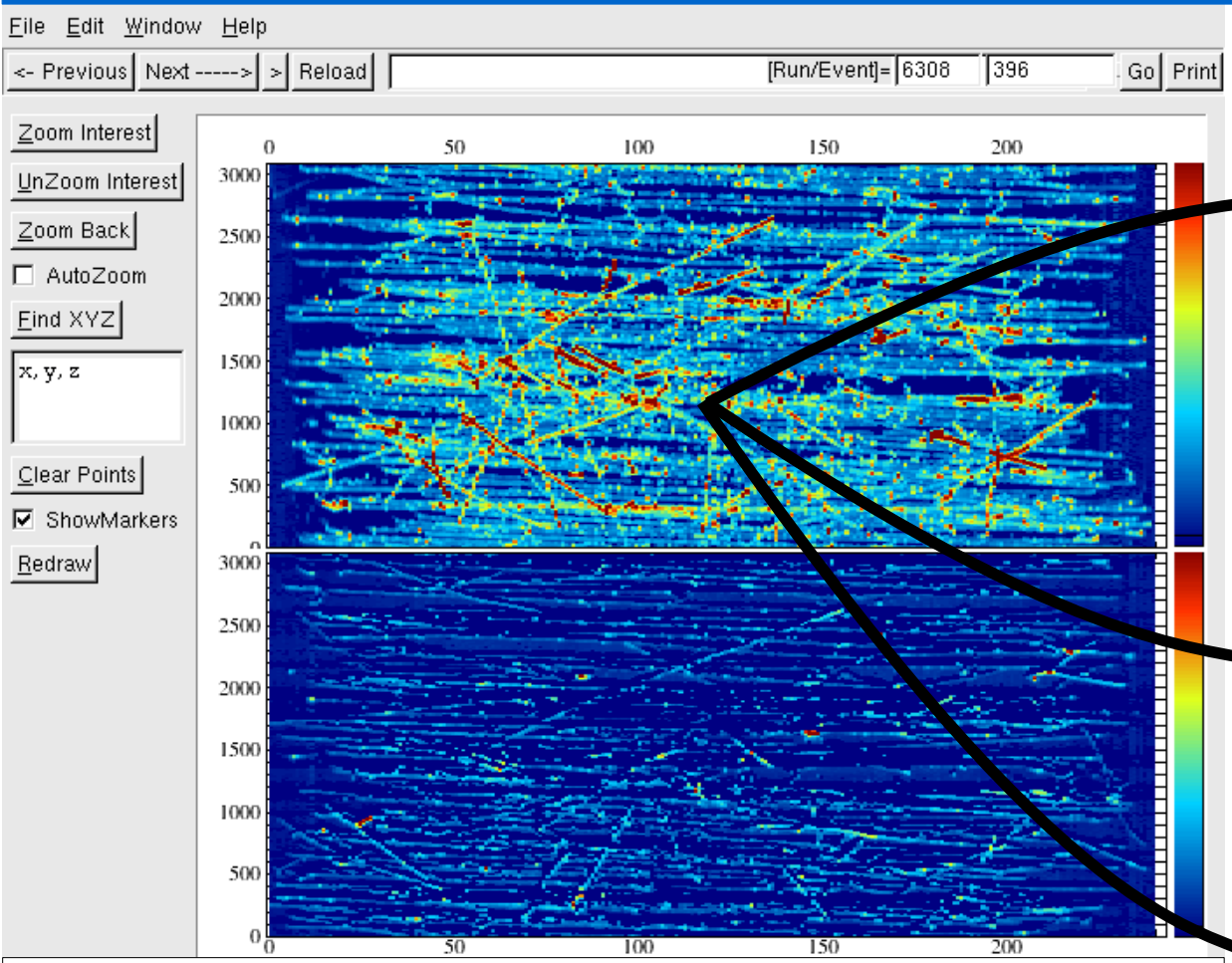
Art::DAQ
(TPC, Wire Chambers, TOF, PMT's, etc....)

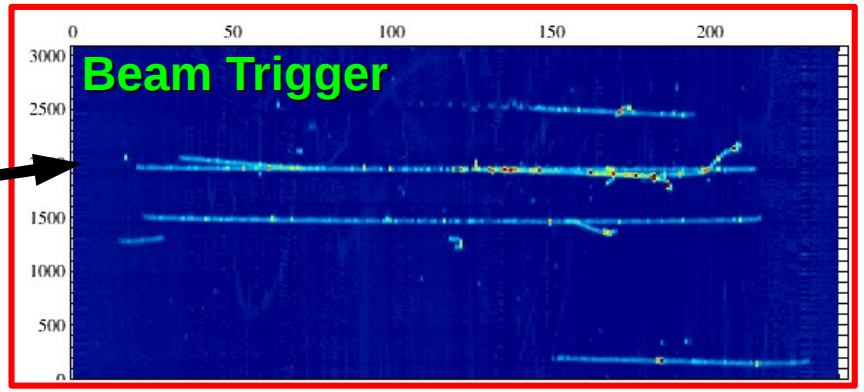
The LArSoft Line

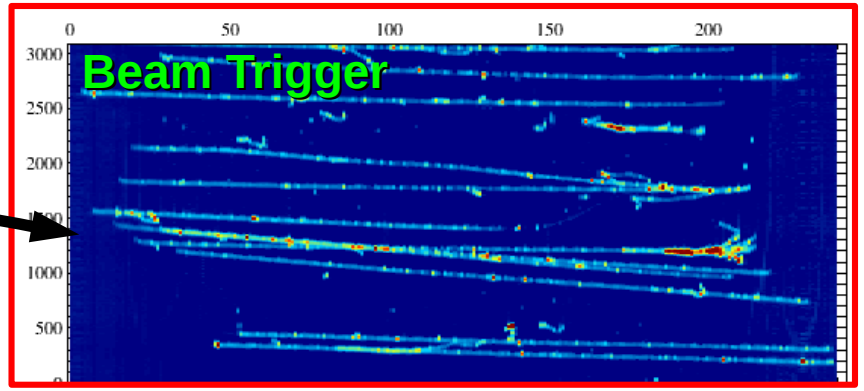
Spills recorded
(Puts together all the various subsytems into Triggers)

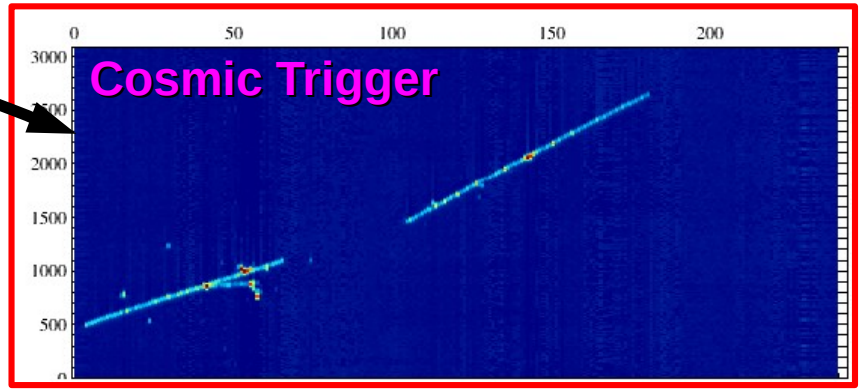


Raw Data Structure





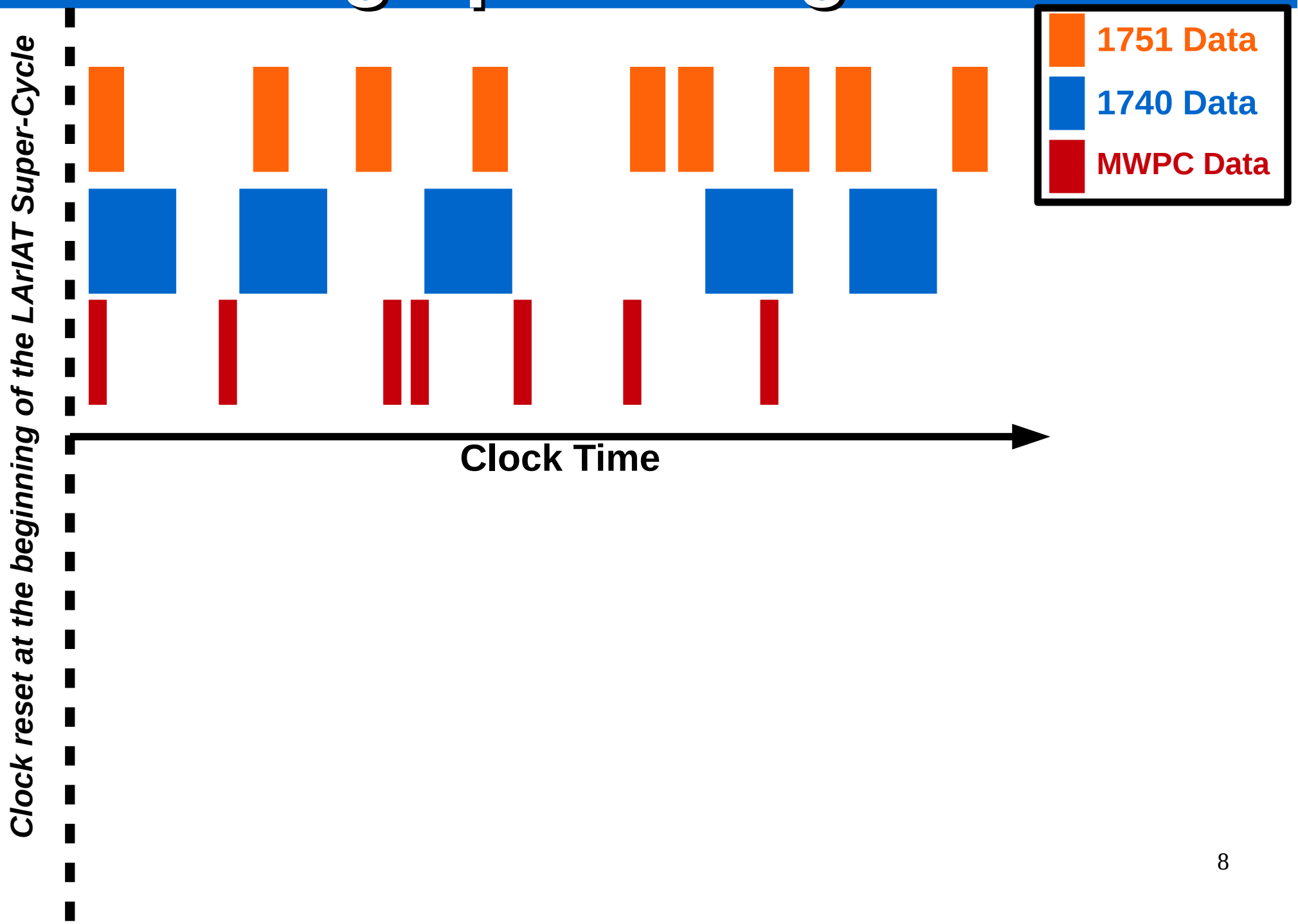




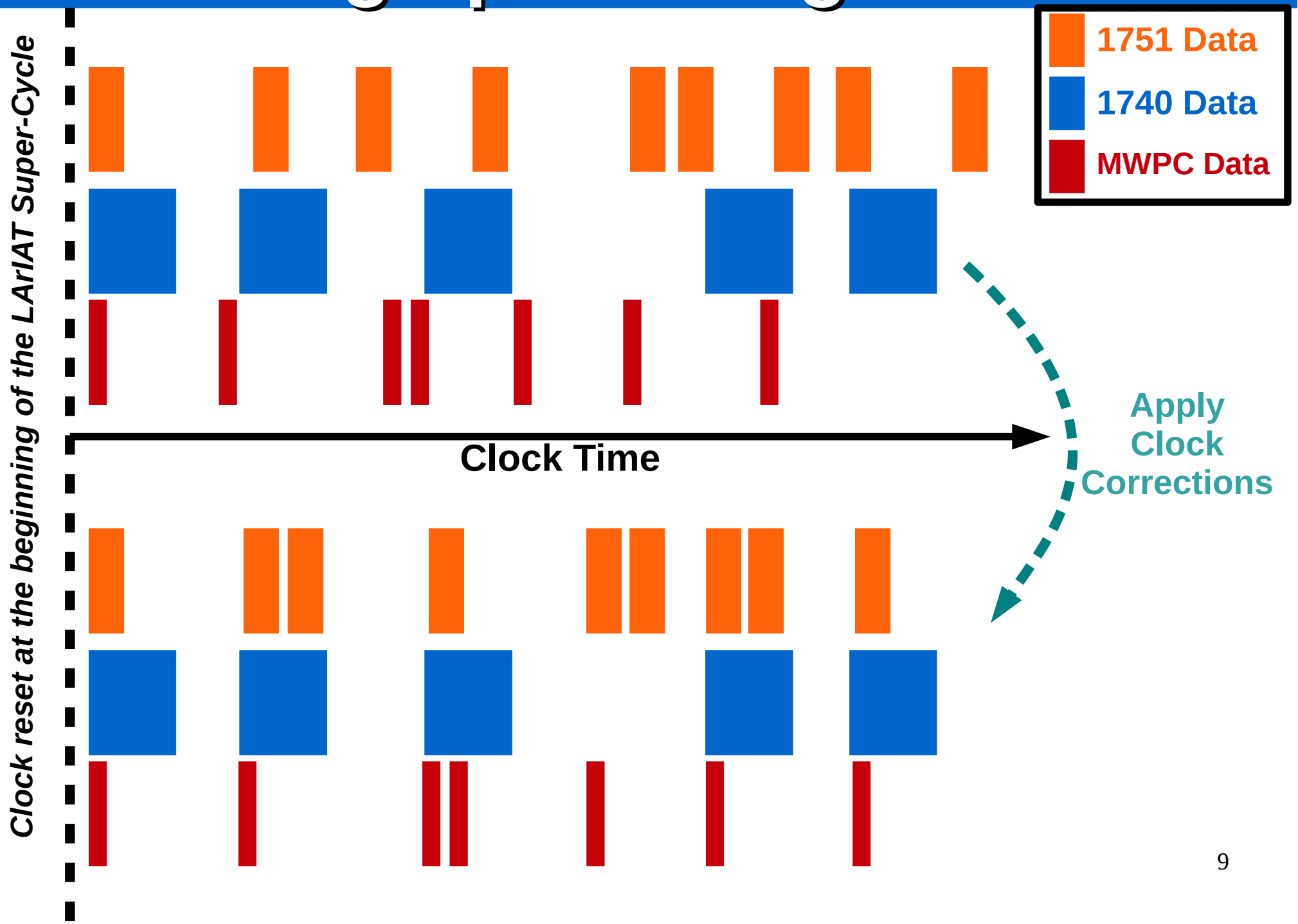
There are 40 different "triggers" within this one "data block"!

In order to make sense of this with LArIATsoft we want to restructure the data

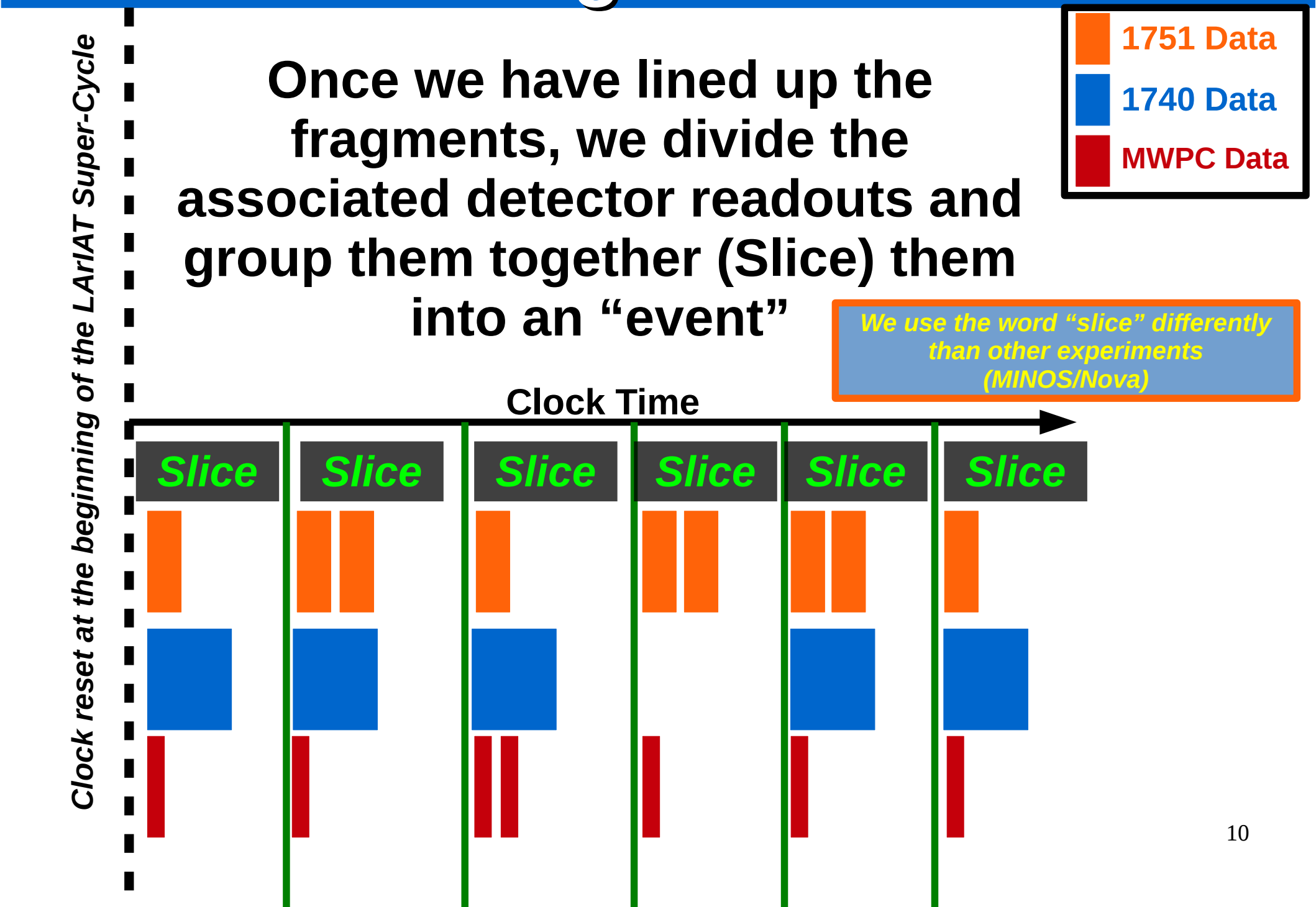
Lining up our fragments



Lining up our fragments



Slicing our data



Raw Data Structure

Art::DAQ
(TPC, Wire Chambers, TOF, PMT's, etc....)

The LArSoft Line

SlicerToDigit
(Puts together all the various subsytems into Triggers)

Run 1
Spill1 == SubRun1

Event # 1
- Trigger # 0
- RawDigits
- OpDetPulses
- AuxDetDigit (WCTrack)
- AuxDetDigit (TOF)
- AuxDetDigit (MURS)

Event # 2
- Trigger # 1
- AuxDetDigit (WCTrack)
- AuxDetDigit (TOF)
- AuxDetDigit (MURS)
- etc....

Event # 3
- Trigger # 2
- RawDigits
- OpDetPulses
- AuxDetDigit (WCTrk)
- AuxDetDigit (TOF)

Run 1
Spill2 == SubRun2

Event # 4
- Trigger # 0
- RawDigits
- OpDetPulses
- AuxDetDigit (WCTrack)
- AuxDetDigit (TOF)
- AuxDetDigit (MURS)

Event # 5
- Trigger # 1
- AuxDetDigit (WCTrack)
- AuxDetDigit (TOF)
- AuxDetDigit (MURS)
- etc....

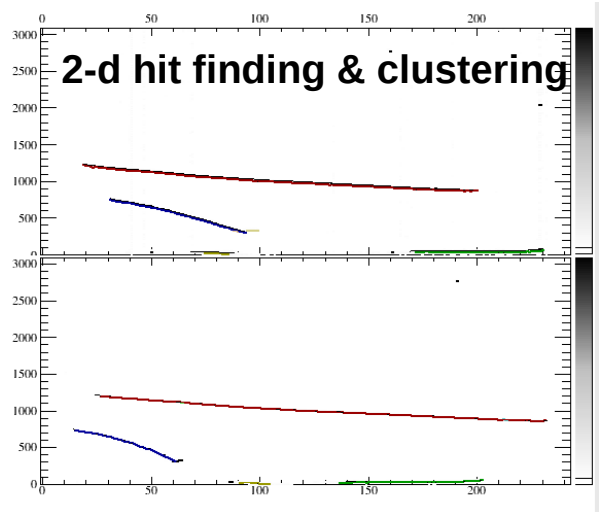
Event # 6
- Trigger # 3
- RawDigits
- OpDetPulses
- AuxDetDigit (WCTrack)
- AuxDetDigit (TOF)
- AuxDetDigit (MURS)

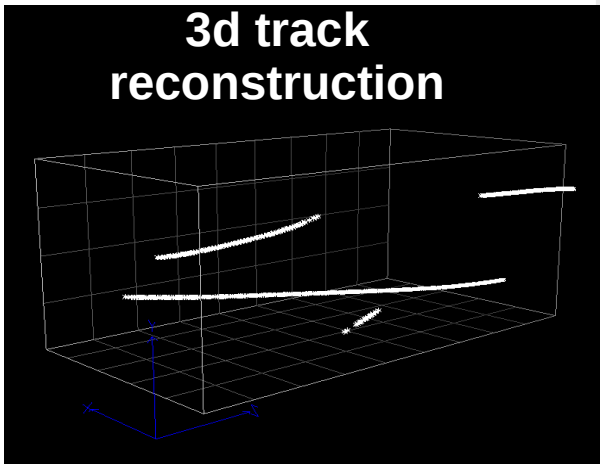
Reconstructing our data

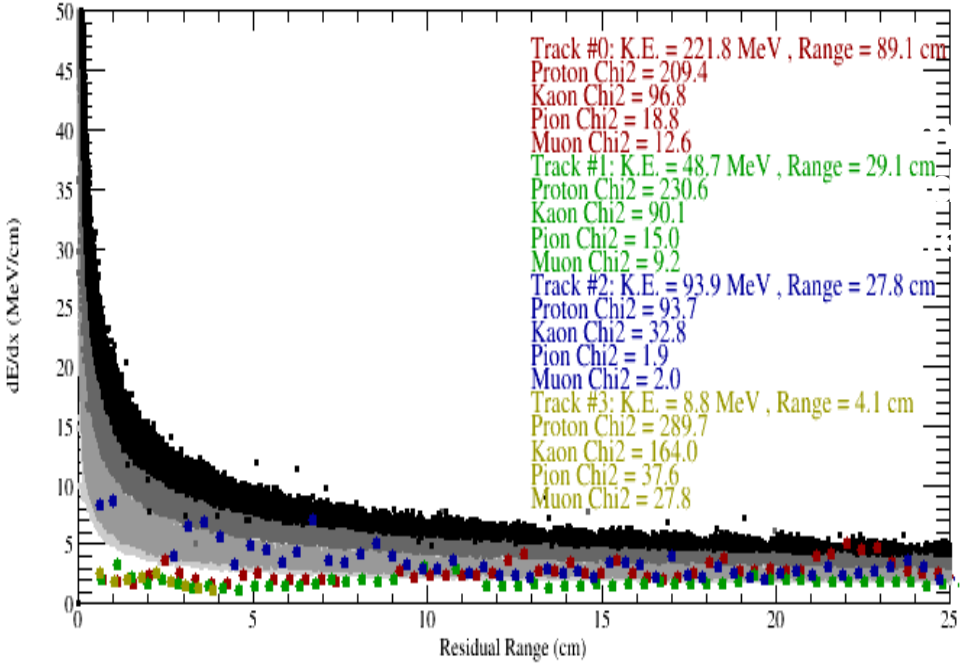
- **We use “standard” LArSoft reconstruction algorithms for TPC based information**
  - TPC Wire Deconvolution, Hit Finding, Clustering, Track Finding, Shower Reconstruction
- **For non-TPC systems (TOF, Wire Chamber Tracks, AeroGel, Muon Range Stack) we write our own modules which take in the digits for these detectors and reconstruct objects based on this information**
  - Wire Chamber Tracks, TOF Objects, Muon Range Stack Hits, AeroGel Hits
- **We can also put the non-TPC object information together to form a preliminary particle identification hypothesis for objects entering the TPC**
  - Combine Wire Chamber Tracks and TOF to separate μ/π from proton
- **Trigger decisions are also stored for users to filter per event**
  - Example: you want to require 3 of 4 Wire Chambers, the beam to have been on, and there was no activity in the halo
    - <+WCCOINC3OF4+BEAMON-HALO>
  - Example: you require no beam and the cosmic ray paddles to have fired during the cosmic readout window
    - <-BEAMON+COSMIC+COSMICON>

TPC Reconstruction

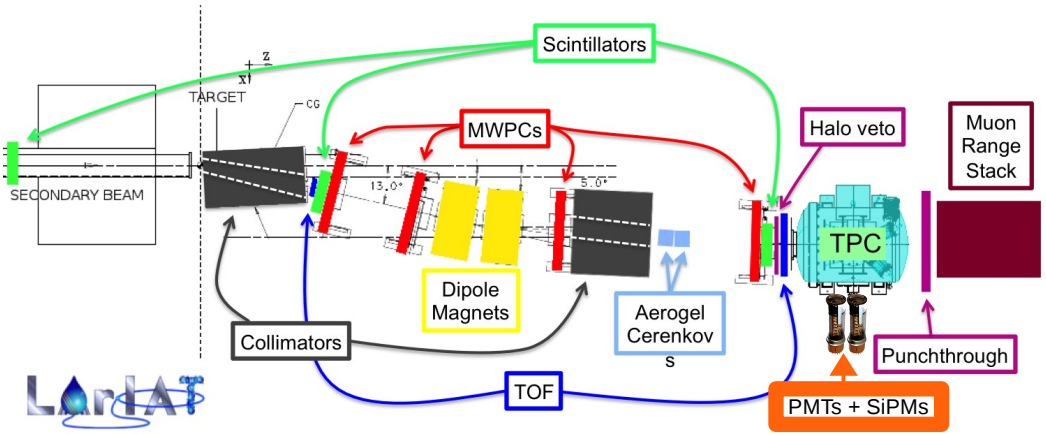
- **Utilizing LArSoft reconstruction modules (tuned for application to LArIAT) we are able to take the TPC information from 2d → 3d reconstruction**
  - 2d hit finding and clustering
  - 3d track and shower reconstruction
  - Track calorimetry and particle ID
- **Tuning of reconstruction parameters and modifying producers to be the most useful for LArIAT still underway and an active area within our analysis teams**

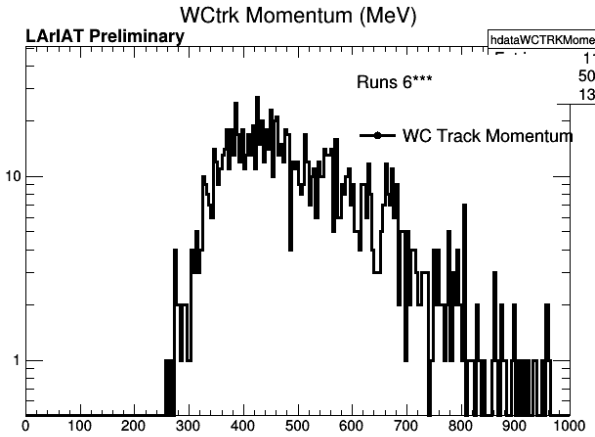


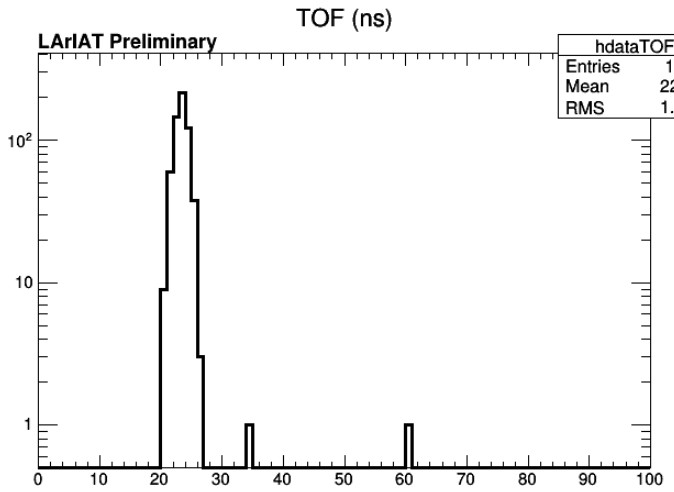




Non-TPC Reconstruction



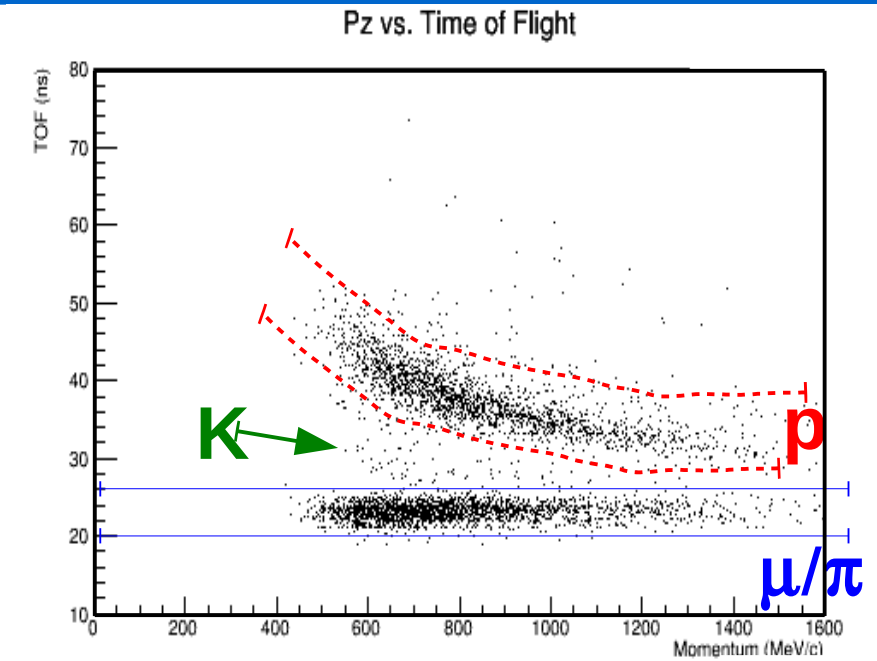


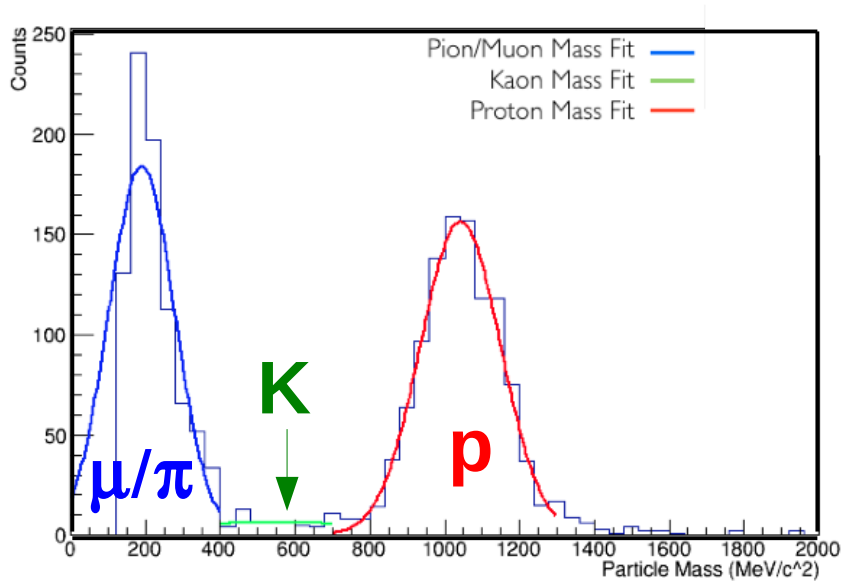


- **Utilizing our own algorithms we can reconstruct relevant beamline information**
  - Wire Chamber Tracks
    - Momentum
    - Projection onto the front face of the TPC
  - Time of Flight
    - Can correlate the TOF with the wire chamber track

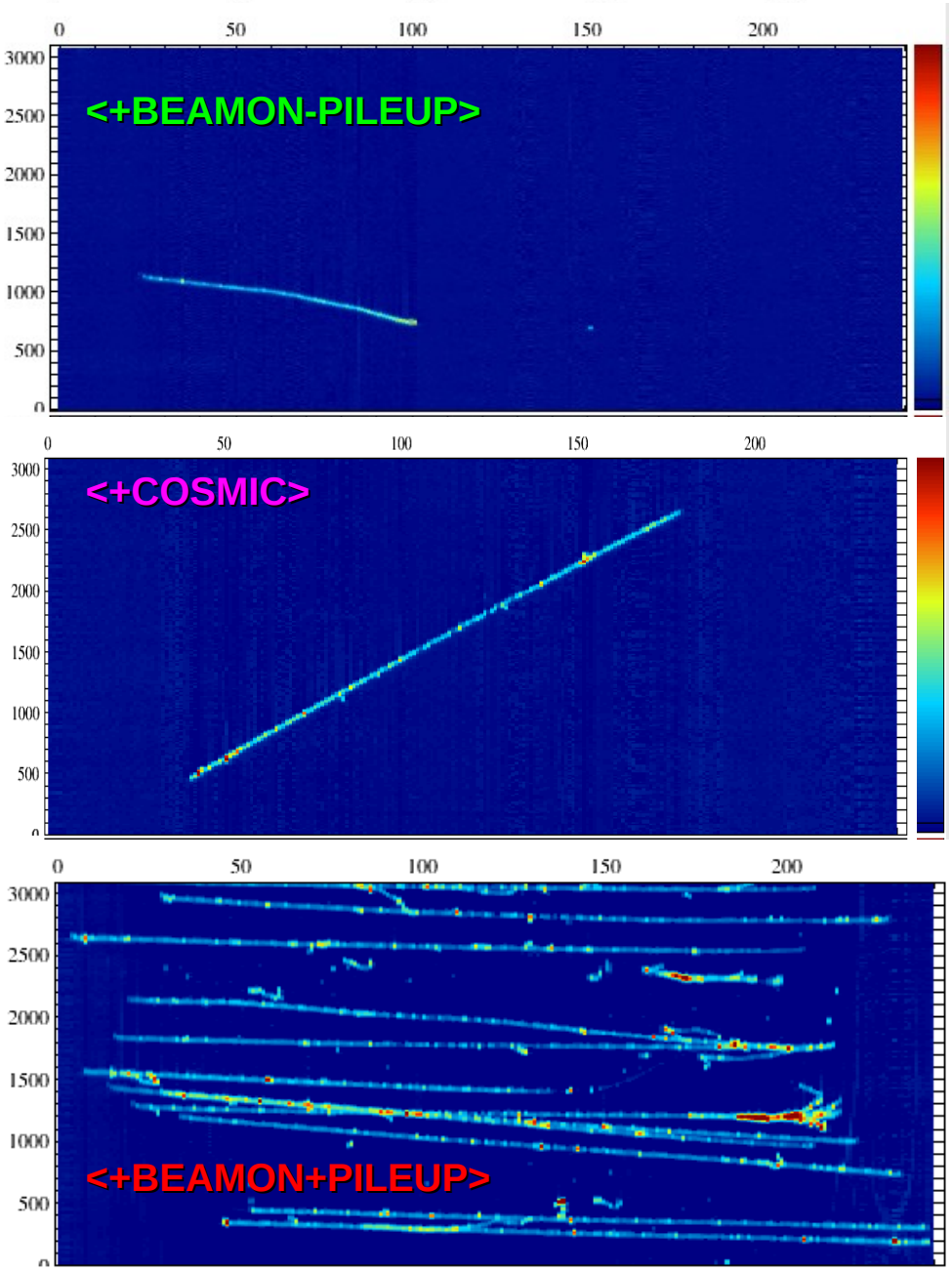
Beam line Particle ID

- **Utilizing the beam line instruments you can begin to separate particles incident to the TPC based on a preliminary identification hypothesis**
  - Right now we use TOF and Wire Chamber Track Momentum to form a particle ID hypothesis
  - Will expand this to utilize Aerogel and Muon Range Stack for μ/π separation
  - Also utilize TPC information for electron identification





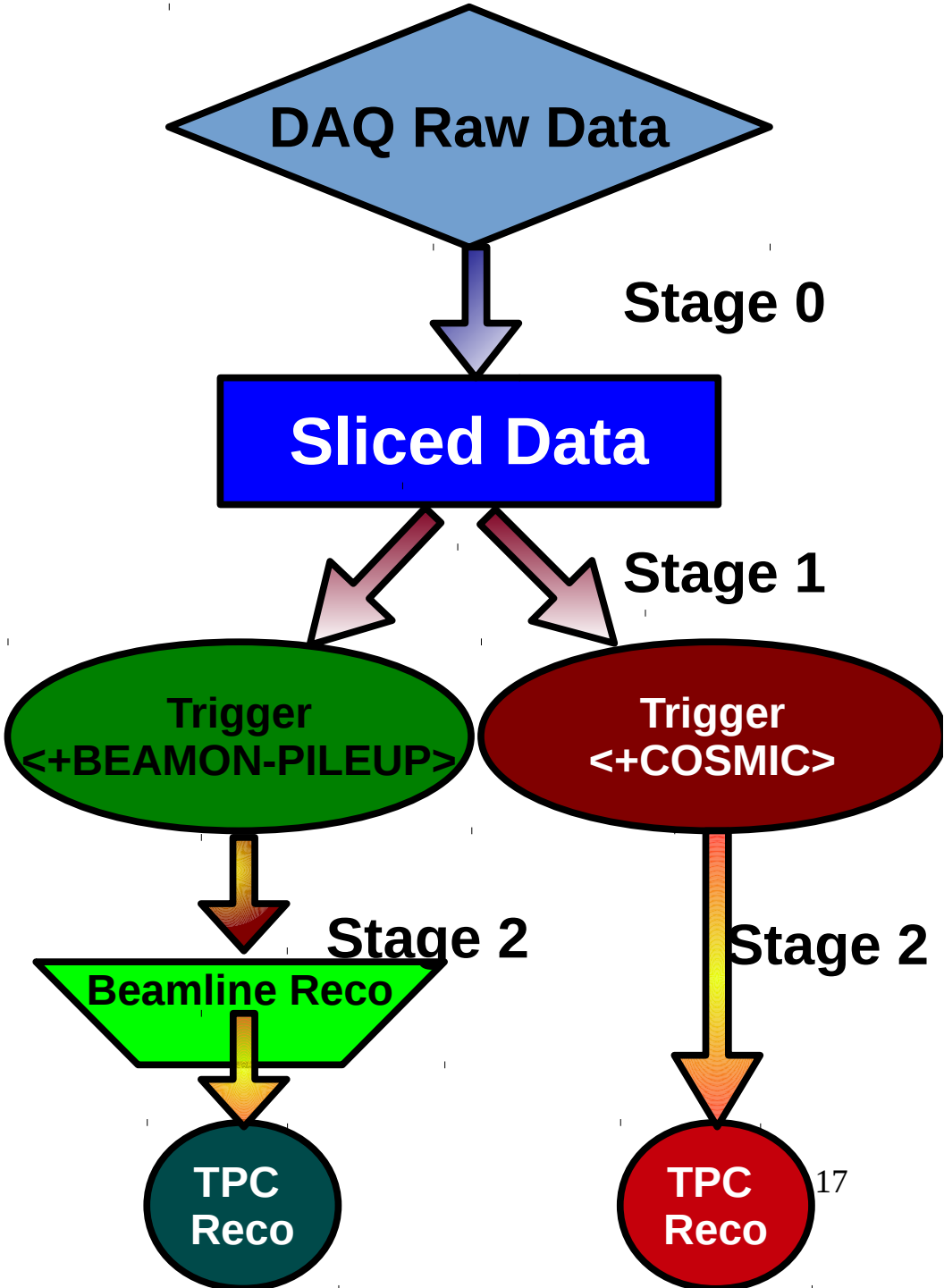
Trigger based filtering



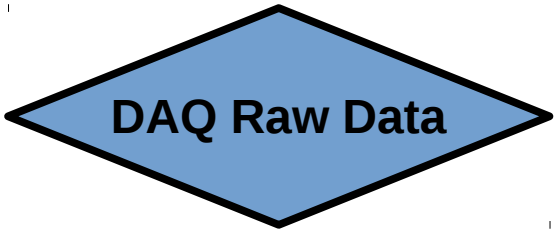
- **The conditions under which the data was read out are stored via a data base allowing us to filter on an event-by-event basis**
  - We can also filter based on running conditions via SAM Meta-data

Data Processing

- **Plans are in place for centralized processing of all the LArIAT data taken during Run-1**
  - Break the reconstruction into three stages
    - Stage0 = Slicing
    - Stage1 = Trigger Filter
    - Stage2 = Reco
- **Utilize run based data base to look up running conditions during data taking**
  - Centralize the “slicing” and “trigger filtering”



Data Processing

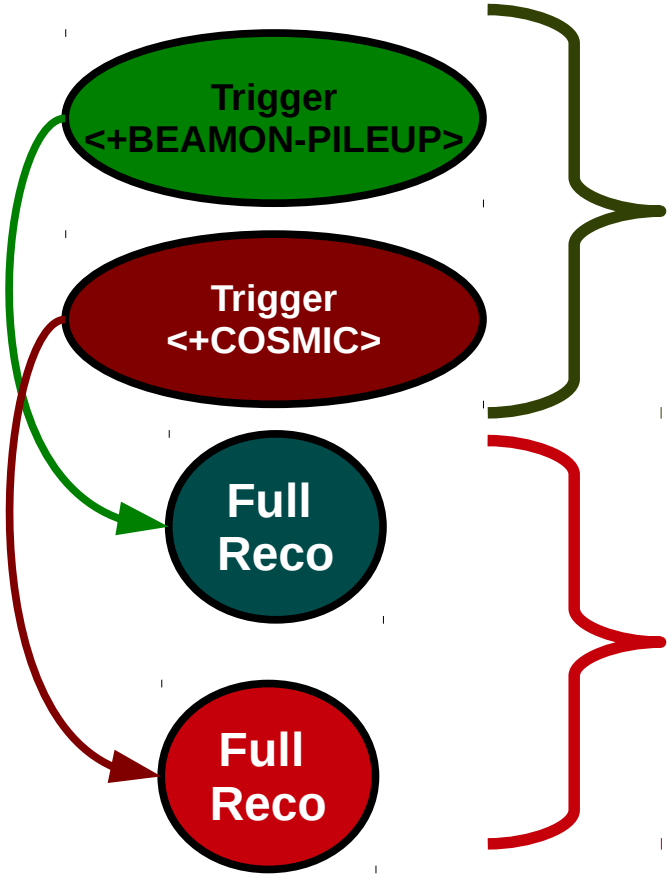


= 20 MB – 80 MB per file

Sliced Data

= 20 MB – 80 MB per file

These files are stored per Sub-Run (the number of sub-runs varies Run/Run)



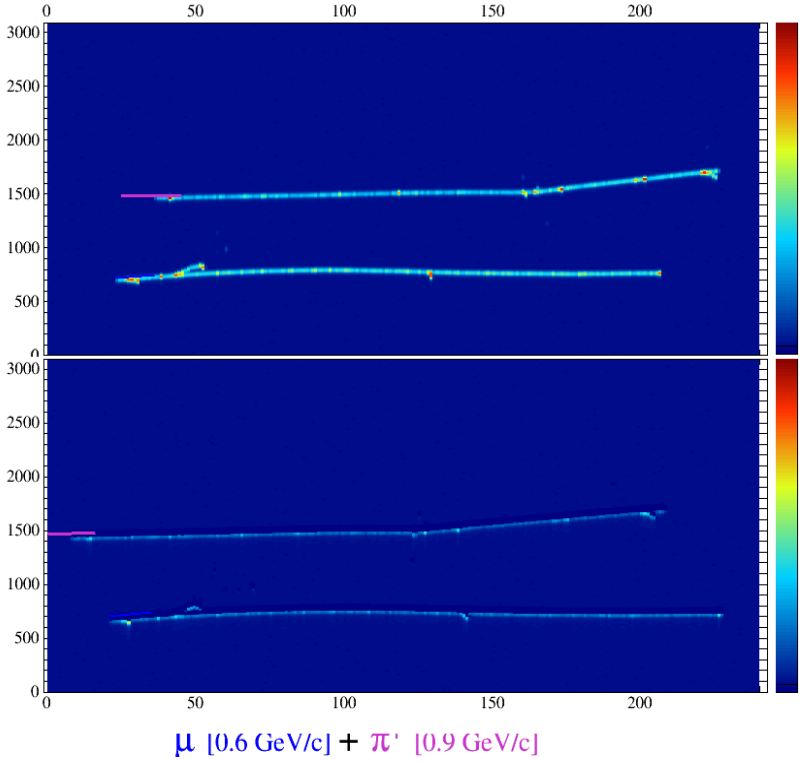
= 60 MB – 3.5 GB per file

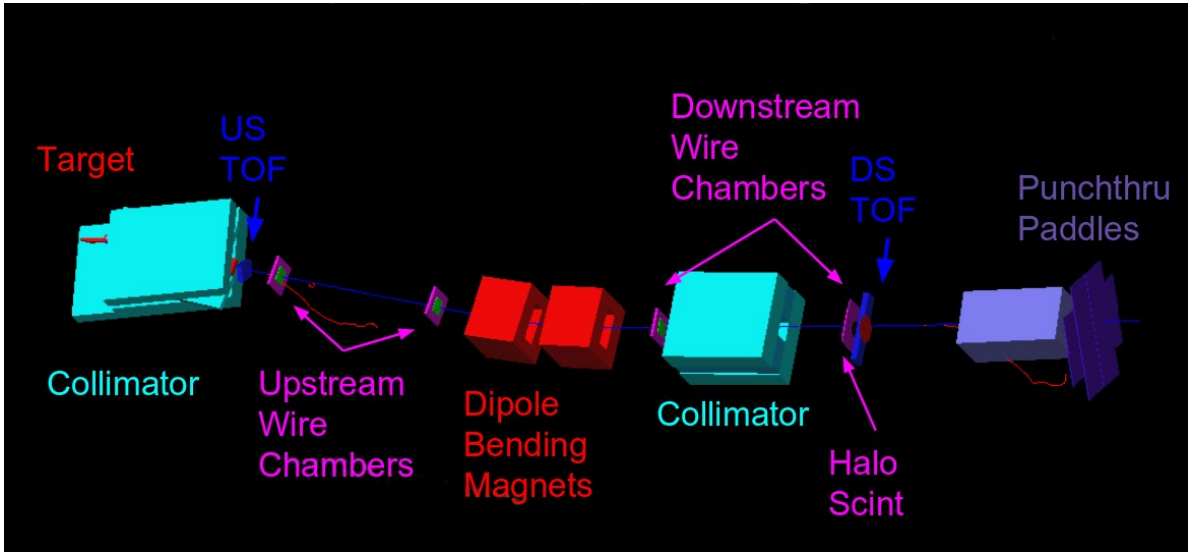
= 500 MB – 10 GB per file

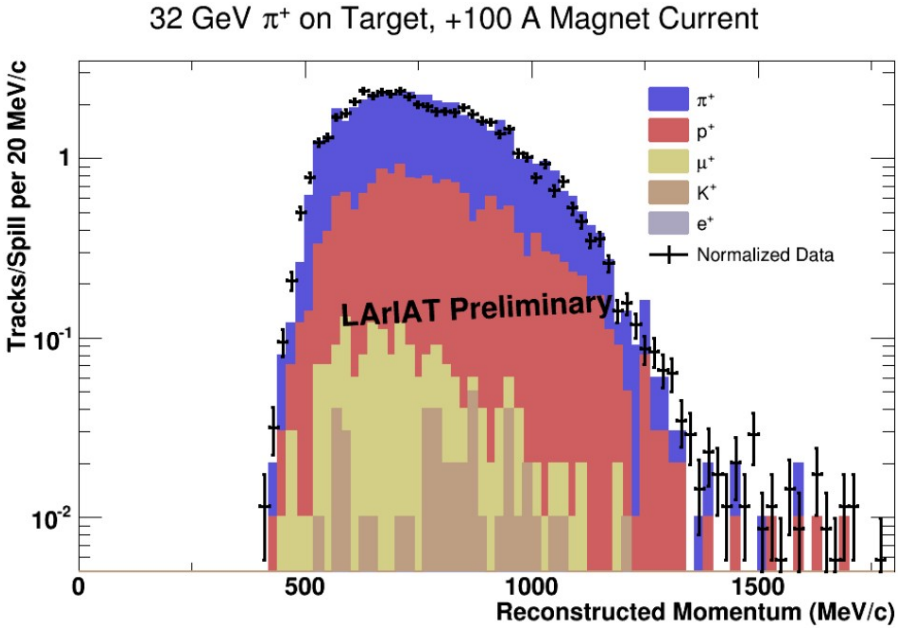
These files are stored per Run (the sub-runs are all combined per Run)

Monte Carlo Production

- **Utilizing G4Beamline simulation we simulate our particle spectrum along with our various beam line elements**
- **We also have Particle Gun Monte Carlo (standard LArSoft production) to produce dedicated TPC studies (no beamline info)**





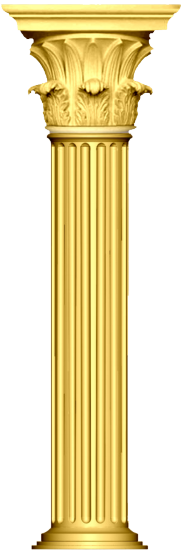


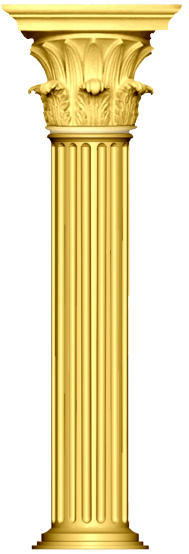
Analysis Plans

- **Inclusive Pion Cross-Section**
- **Pion Absorption Cross-Section**
- **Charged Pion Exchange Cross-Section**
- **Electromagnetic Shower Studies**
  - e.g. Electron/Photon Separation Studies
- **π/μ separation studies**
- **Calorimetric Reconstruction utilizing LAr Scintillation Light**
- **Muon Sign Determination w/o magnetic field**

High level physics analyses

- **Electron Lifetime**
- **Electronics Response Calibration**
- **Charge Recombination Studies**





Foundational calibration analyses

Analysis Plans

- **Inclusive Pion Cross-Section**
- **Pion Absorption Cross-Section**
- **Charged Pion Exchange Cross-Section**
- **Electromagnetic Shower Studies**
  - e.g. Electron/Photon Separation Studies
- **π/μ separation studies**
- **Calorimetric Reconstruction utilizing LAr Scintillation Light**
- **Muon Sign Determination w/o magnetic field**

These analyses have active teams of 2 or more people working on them right now

---

- **Electron Lifetime**
- **Electronics Response Calibration**
- **Charge Recombination Studies**

Foundational calibration analyses

Collaboration Resources

- **Three LArIAT general purpose virtual machines for data analysis**
  - We have recently added one more to accommodate for the increase in LArIAT analyzers
- **8.0 TB of disk space on /lariat/data (BlueArc)**
  - Asked this to be increased to accommodate increase use
- **Tape storage for data (/pnfs/lariat/raw)**
  - More then enough (nearly infinite)
- **2.0 TB of disk space on /lariat/app (BlueArc)**
  - Seems to be sufficient for the immediate use
- **100 slots of grid space dedicated for LArIAT use**
  - This was just recently upgraded to accommodate our forthcoming production run

“Are there robust plans for data processing and data analysis?”

- **Data Processing: Yes**
  - Data processing has been underway “piece-meal” as we tune our reconstruction and analysis
  - **Already done once over the entire data set for the lifetime analysis**
  - Large scale reconstruction is about to start over the entire data set
- **Data Analysis: Yes**
  - Three analyses have been targeted as “fast-track” analyses which have groups of people working on
    - Inclusive Pion Cross-section
    - LAr Scintillation Light Studies
    - EM Shower Studies
  - A number of other analyses are underway and build on the “fast-track” analysis work
    - Pion absorption
    - Charged pion exchange
    - π/μ separation studies
  - Analyses to extract calibration of our offline data is also continuing
    - Electron Lifetime Calibration
    - Electronics response calibration

“Have adequate resources from the laboratory and the collaboration been identified for data analysis to meet these goals?”

- **Yes**
  - As we've come to understand our file size and processing requirements the support from SCD has been very responsive
    - 100 dedicated slots on the grid
      - Can process (Slice) all the LArIAT Run 1 data in 4 hours with 25 slots
    - 16 GPVM cores dedicated to LArIAT (Three 4 core machines and one 2 core machine)
    - Increasing BlueArc storage capacity from 10 TB → 20 TB for ongoing analyses
  - The number of university based collaborators driving our analyses has been increasing to meet the demands of trying to accomplish timely publications

Questions / Comments