DUNE Data Selection

Josh Klein, Artur Sztuc,
for the DUNE DAQ Consortium's Data Selection Working Group

January 27, 2025

Penn
University of Pennsylvania

DUNE

Outline

- DUNE “Data Selection” Basics
- Trigger Design
- Performance at ProtoDUNE
- Trigger Simulation Model

DUNE DAQ Data Selection Working Group

Within the DUNE DAQ Consortium

Responsible for:

- Beam instrumentation triggering of ProtoDUNEs
- TPC-based triggering of ProtoDUNE(s) and DUNE FDs
- PDS-based triggering for ProtoDUNE (NP02) and DUNE FDs
- Calibration source triggering
- Region-of-interest selection of data for FDs
- High level "data filtering" for FDs (if any)

A lot of effort over many years by many people!

Adam Abed Abud, Manuel Arroyave, Nuno Barros, Charlie Batchelor, Thiago Bezerra, Kurt Biery, Alexander Booth, Juan Miguel Carceller, Animesh Chatterjee, Simranjit Singh Chhibra, David Cussans, David Drobner, Hamza Amar Es-sghir, Ben Harris, Ivana Hristova, Nikolina Illic, Tom Junk, Daisy Kalra, Georgia Karagiorgi, Wesley Ketchum, Josh Klein, Ben Land, Pierre Lasorak, David Last, Giovanna Lehmann-Miotto, Matthew Mann, Andrew Mogan, Alejandro Oranday, Simon Peeters, Michal Rigan, David Rivera, Marco Roda, Phil Rodrigues, Jonathon Sensenig, James Shen,Roland Sipos, Jose Soto, Artur Sztuc, Alessandro Thea, Stoyan Trilov, Deniz Uzun, Emanuelle Villa, Brett Viren, Klaudia Wawrowska…

I will focus today primarily on DUNE FD triggering and related studies.

DUNE Trigger Requirements

Raw data rate out of DUNE ~ 10 TB/s

Paraphrasing the language, DUNE's DAQ shall:

1. Trigger detector with > 90% efficiency for any interaction that leaves > 100 MeV of visible ionization energy inside the fiducial volume.

2. Trigger detector with "high efficiency" for < 100 MeV of visible energy. The lowest threshold is "expected to be" 5 MeV, but near 10 MeV "Those triggers will normally be fired using a pre-scaling factor."

3. Trigger with 95% efficiency for a SNB producing at least 60 interactions with a neutrino energy >10 MeV in 12 kt of active detector mass during the first 10 seconds of the burst

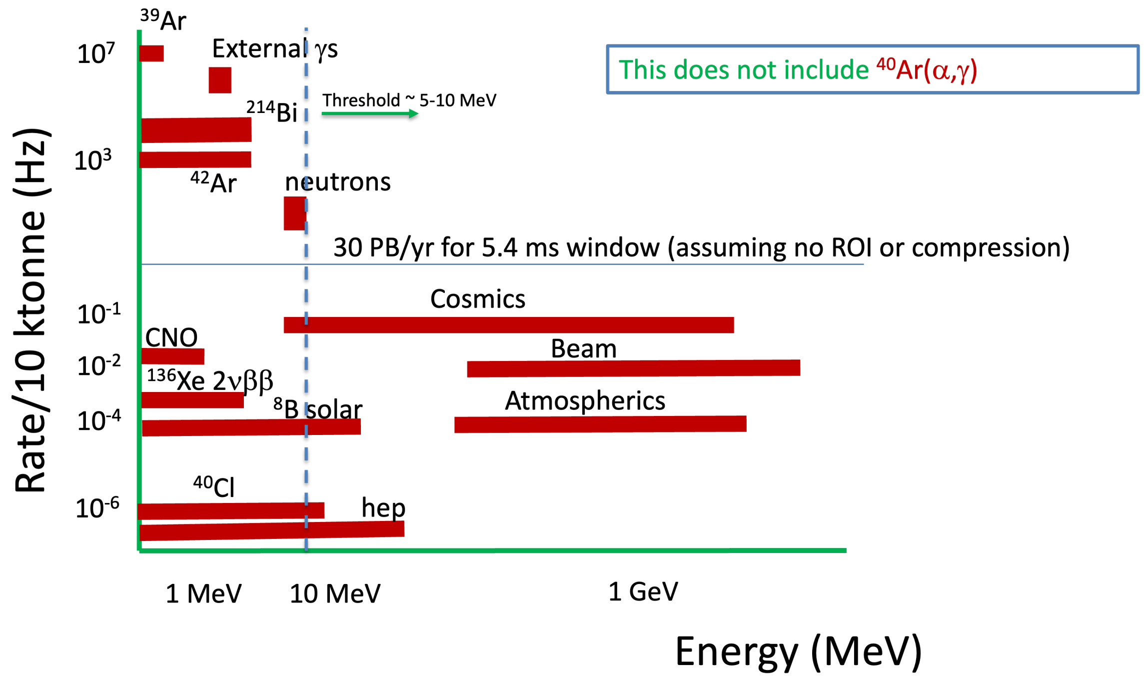
We aim to do much better than all of these.

Trigger Design Considerations

Most important component of the trigger system

4850 ft of rock means:

- Reading out every 512 ns waveform sample
- Of every channel
- For ~ 5 ms/channel
- For every muon (4000/day/10 kt)
- We would store "just" 10 PB/year/10 kt
- (< 30 PB/year allocation)

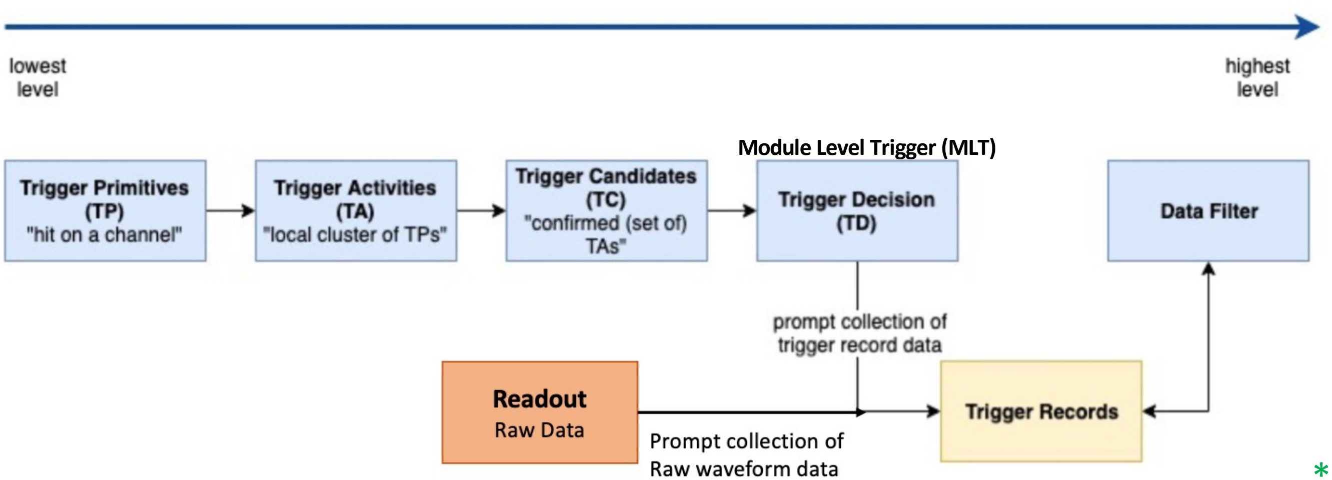


Above 5-10 MeV, little is happening but signal
Below 5-10 MeV backgrounds rise ***steeply***

Thus trigger will be inclusive—we record all "high energy" events with efficiency >99%

Trigger Design Considerations

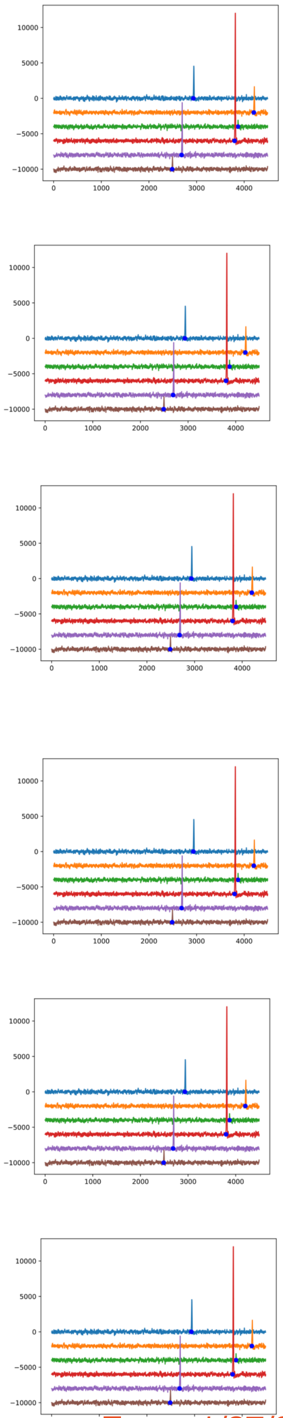
- Large DAQ buffers (10 s) and slow detector allow trigger to be (nearly) all software
- Segmentation of tracking detector allows system to be hierarchical
- Prioritize ***inclusivity*** and ***efficiency***; the trigger is not a data analysis
  - Select and record as much data as possible, let offline sort it out
  - Trigger has no "privileged" information; offline has access to everything
- Emphasize ***simplicity**** so that understanding efficiencies later is tractable
  - Include diagnostic information, prescales, randoms, zero bias, etc.
- Remain as ***flexible*** as possible to accommodate new ideas and physics
- Satisfy all other DAQ constraints given affordable hardware, storage, networking



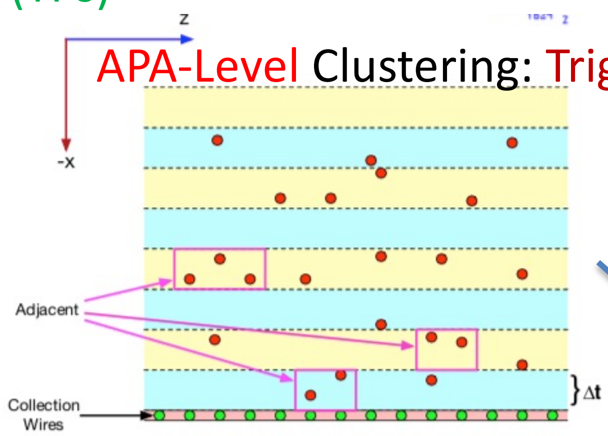
*Ockham's Razor applies: We will always use the simplest algorithm possible

DUNE Trigger Basics (TPC)

Channel Hit-finding: Trigger Primitives (TPs)

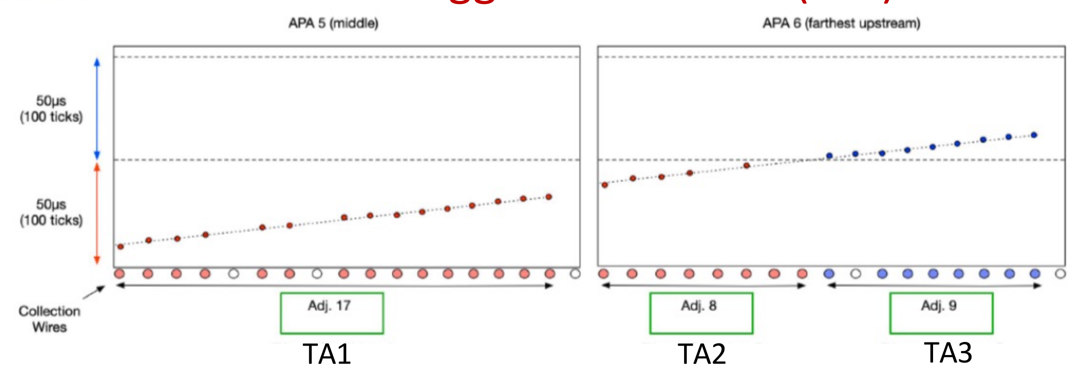


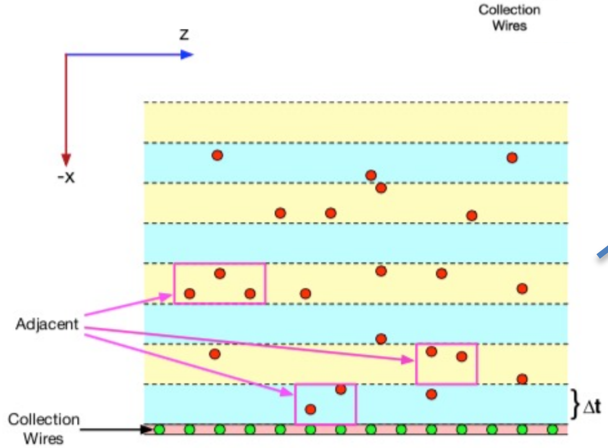
APA-Level Clustering: Trigger Activities (TAs)



Hierarchy

Module-Level Algorithms: Trigger Candidates (TCs)

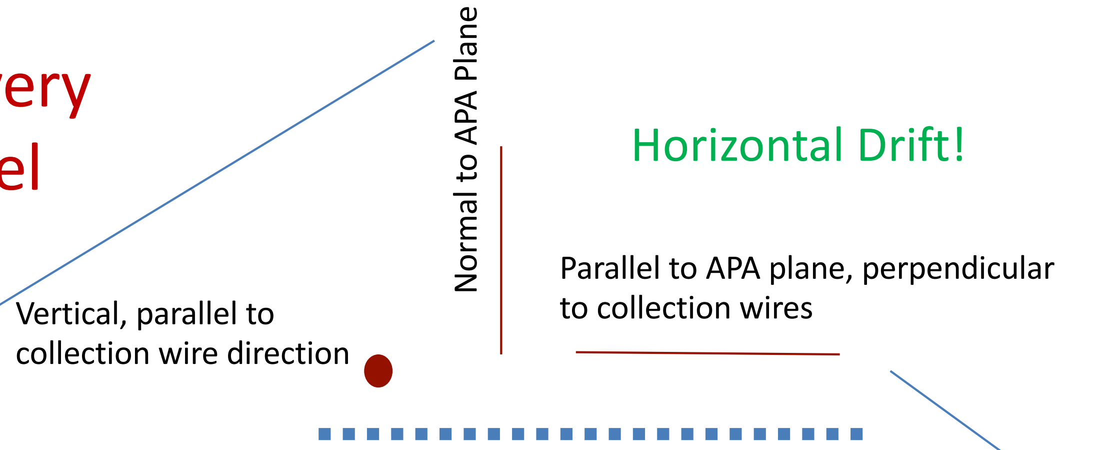


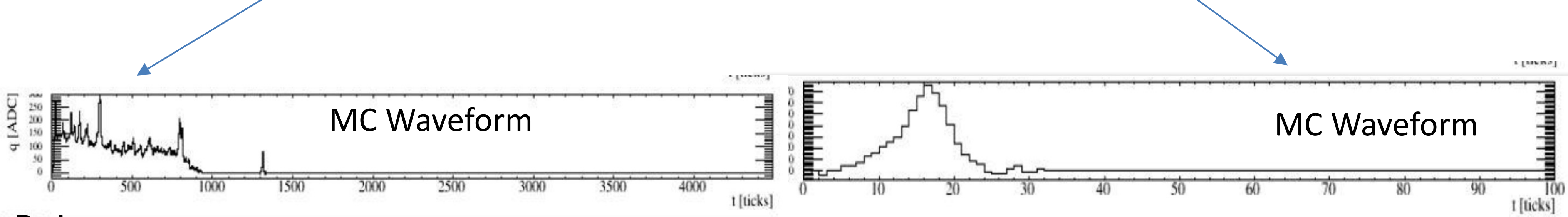


- Threshold for TPs ~ 250 keV/channel
- FD Rate is 50-100 Hz per channel
- TA Makers view ~2500 channels
- The detector is NOT event-based like MC!
  - Data never stops coming into TA finders

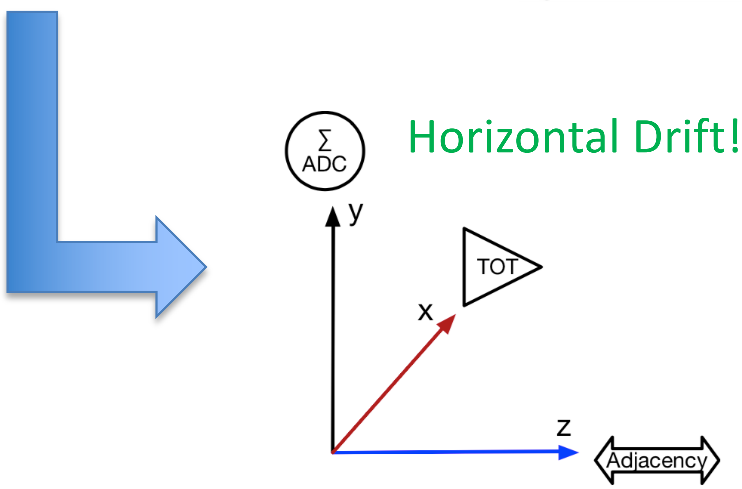
Trigger Primitives (TPs)

Different track geometries create very different per-channel waveforms





D. Last



- TPs include “orthogonal” basis of information
- We produce TPs for both collection and induction channels

D. Rivera

Trigger Primitives (TPs)

Hit finding

Two algorithms:

- Simple Threshold: Uses rms of noise and sets a threshold based on that
- Absolute Value Running Sum:

$$I_{RS}(n) = R \cdot I_{RS}(n-1) + \frac{|I_{raw}(n)|}{s}$$

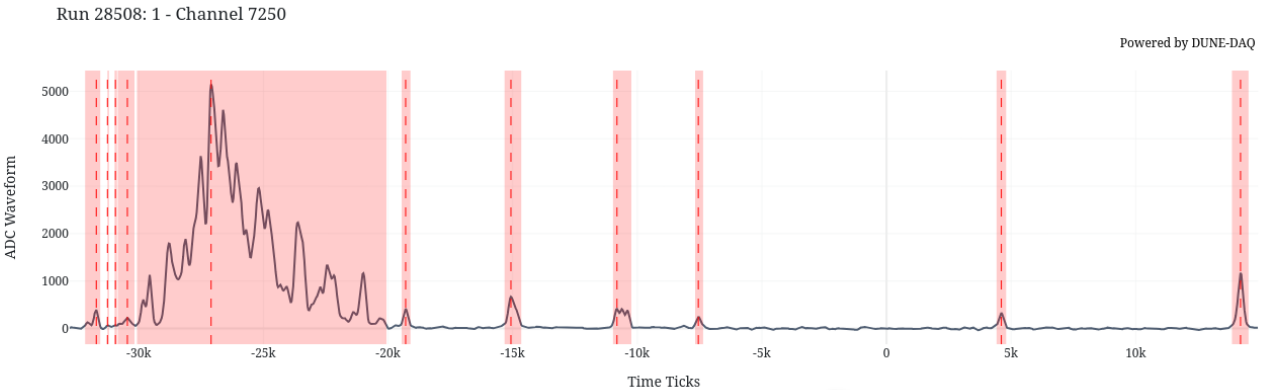
K. Wawrowska

Works on induction channels

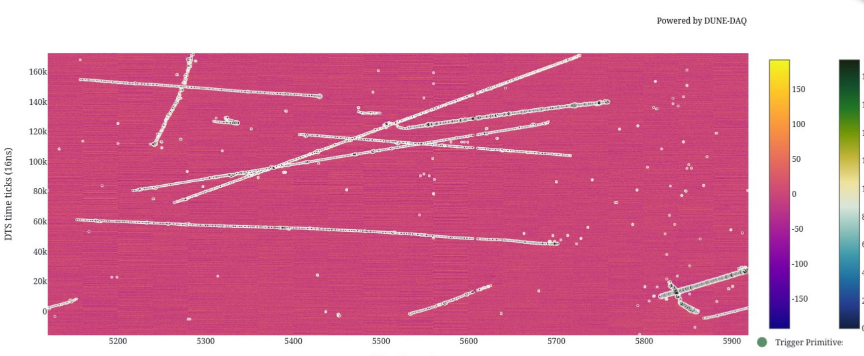
R = 0.7 (weighting factor)
s = 2 (scaling factor)

I_{RAW} = ADC value at time t
I_{RS} = running sum value at time t

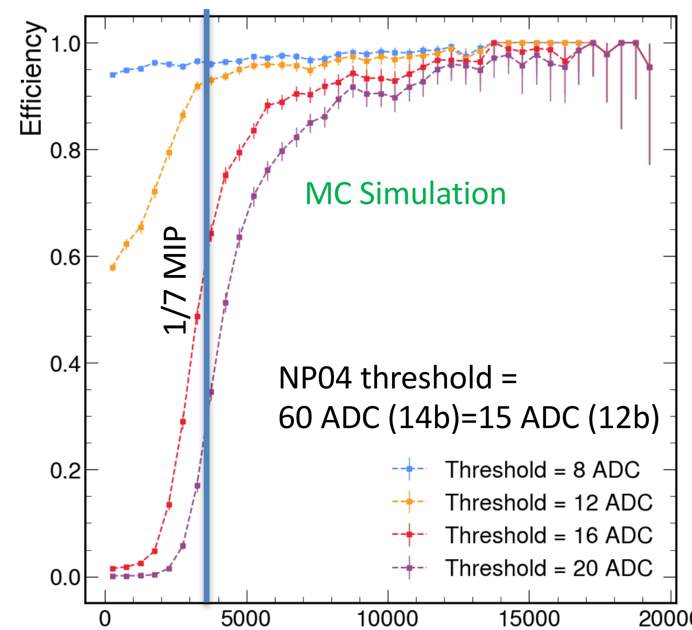
Example performance of ST at NP04:



A. Oranday



Uses AVX to be fast enough for detector---ran everything on 50% of one machine at NP04



J. Shen

Trigger Primitives (TPs)

Hit finding

Two algorithms:

- Simple Threshold: Uses rms of noise and sets a threshold based on that
- Absolute Value Running Sum:

$$I_{RS}(n) = R \cdot I_{RS}(n-1) + \frac{|I_{raw}(n)|}{s}$$

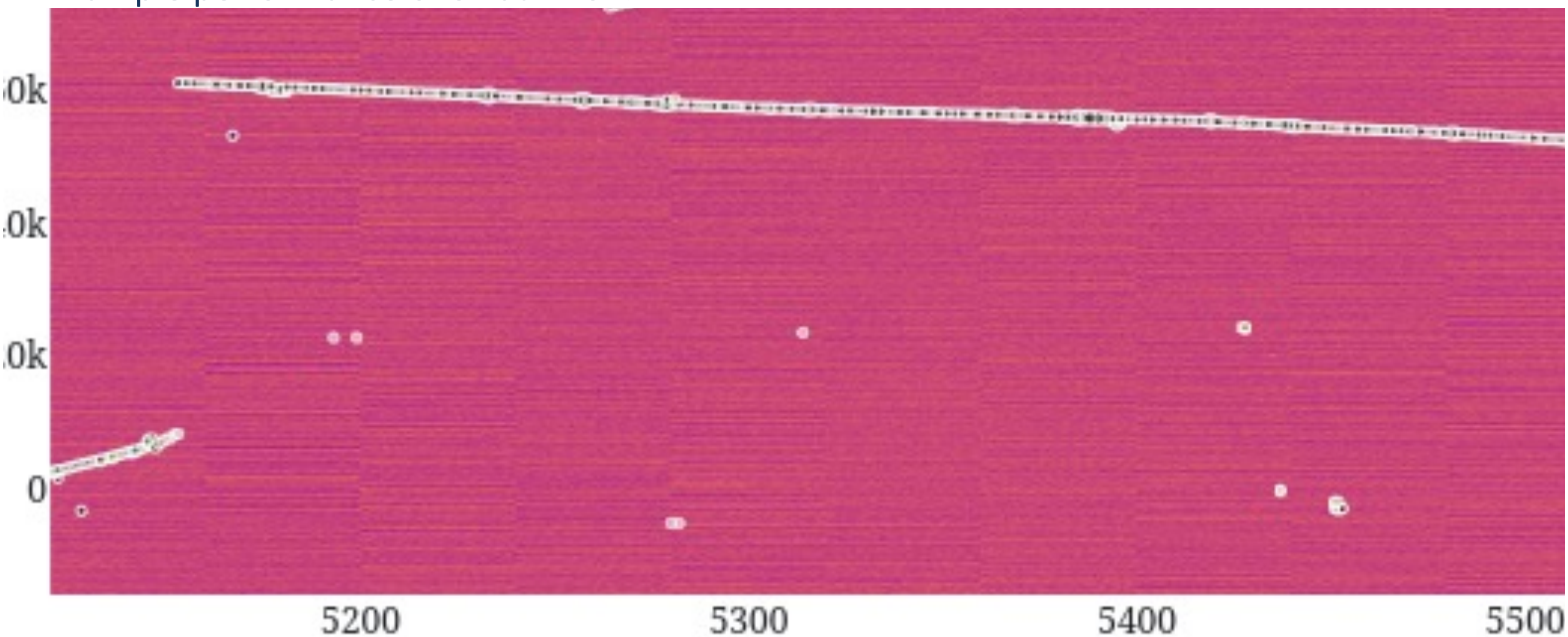
K. Wawrowska

Works on induction channels

R = 0.7 (weighting factor)
s = 2 (scaling factor)

I_{RAW} = ADC value at time t
I_{RS} = running sum value at time t

Example performance of ST at NP04:



Trigger Primitives (TPs)

The TP Stream

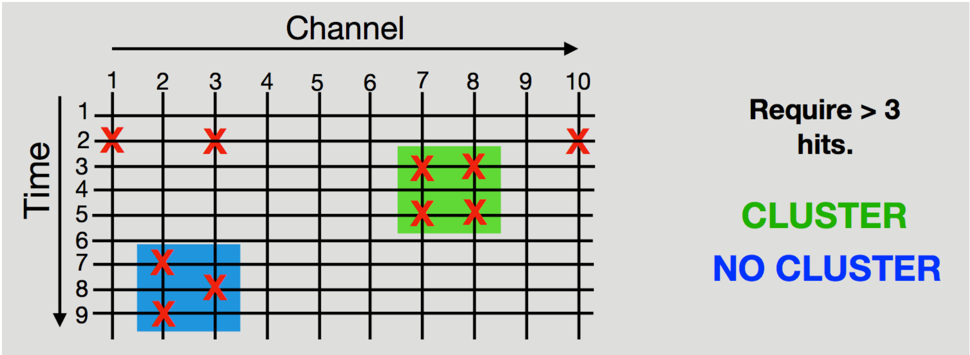
We will write out all TPs on all planes.
This is an "untriggered" data set for DUNE.

Trigger "Activity"

Clustering at the APA/CRP Level

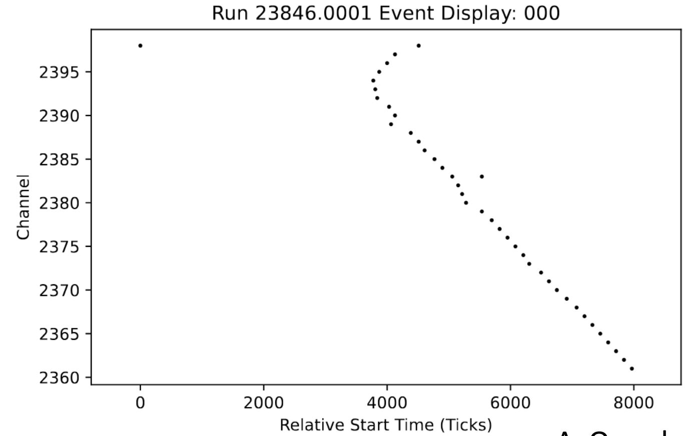
Several algorithms developed---at NP04 we have just 1 μs to process a TP

ADCSimpleWindow (cluster TPs in time and cut on total charge)



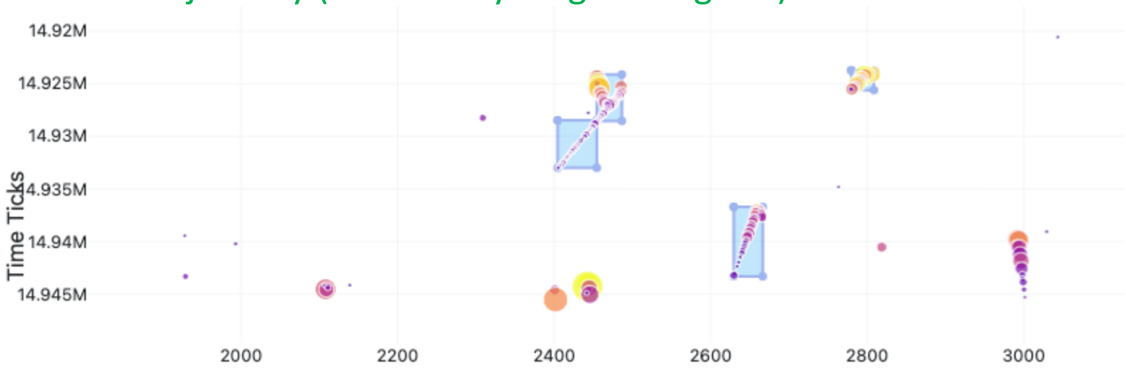
A. Booth

ChannelDistance (cut on distance between TPs)



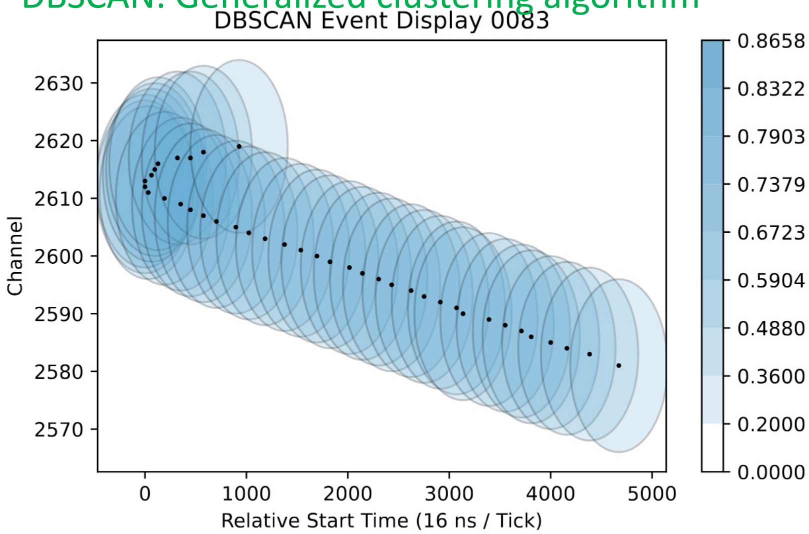
A. Oranday

ChannelAdjacency (cut on only neighboring TPs)



S.S. Chhibra

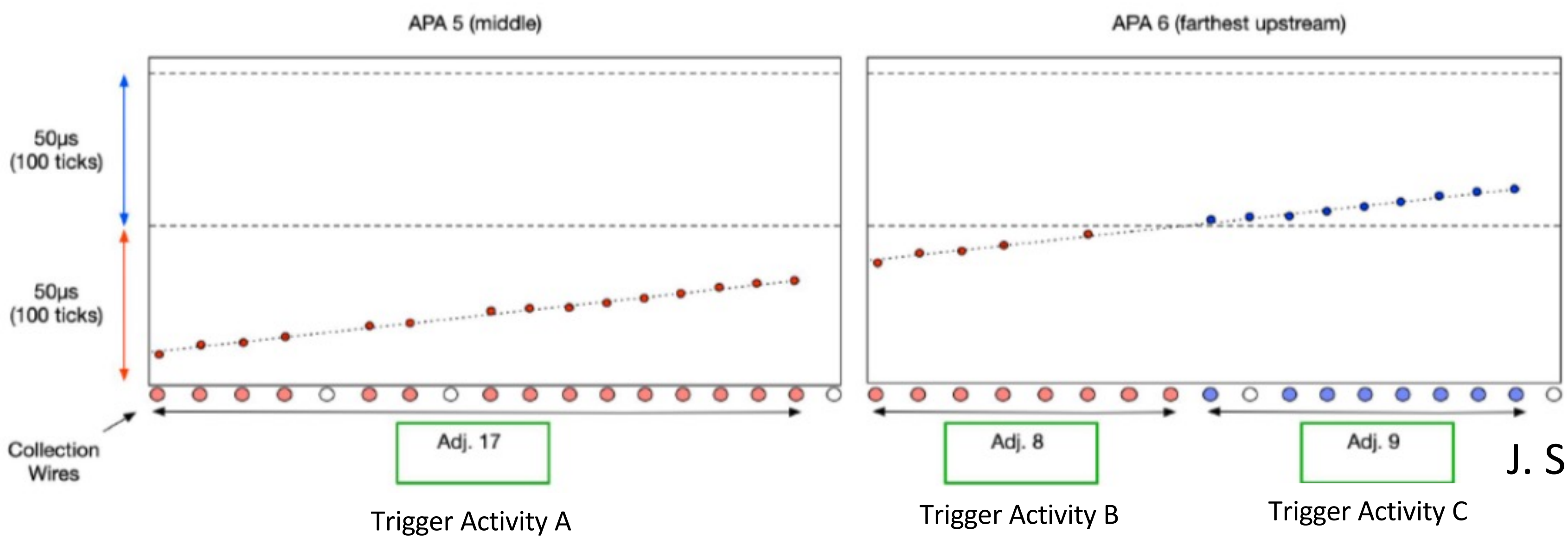
DBSCAN: Generalized clustering algorithm



P. Rodrigues, A. Oranday, A. Sztuc

Trigger Candidates

- TCMakers see full module for 1 system and algorithm
  - High E TPC, High E PDS, etc.—each a different TCMaker
- Also can see long periods of time (e.g., for SN bursts)
- For HE events, most likely process is just “promotion” of TA->TC
- But coincident TAs need to be merged into 1 TC
  - E.g., pileup of a cosmic-ray with an atmospheric neutrino
  - (Happens about once a year per module…)

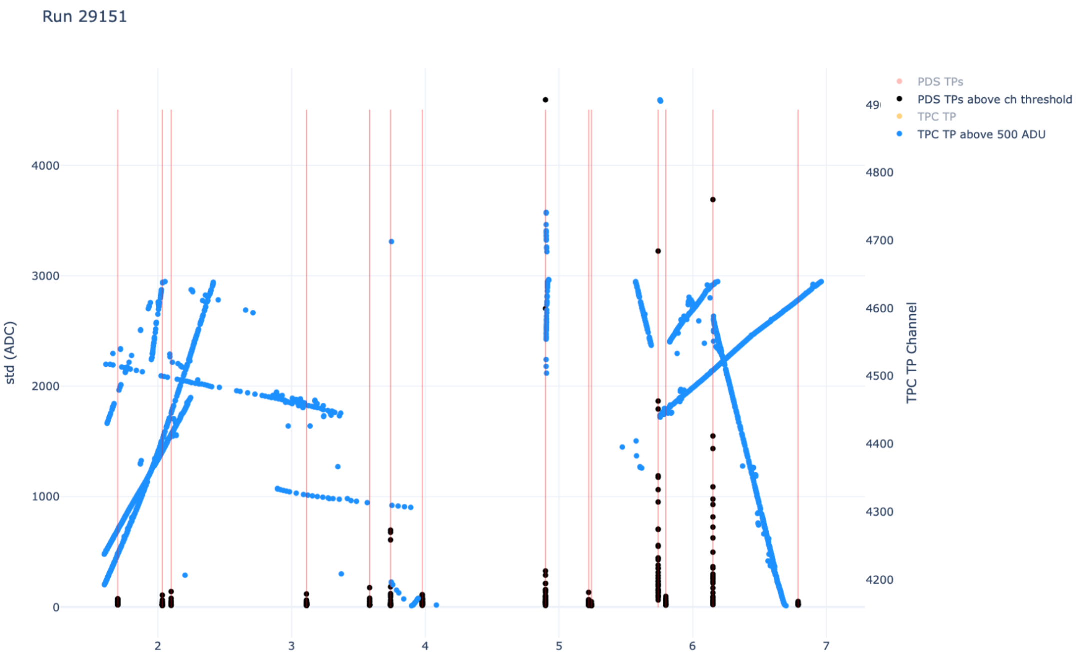


J. Sensenig

PDS Triggers

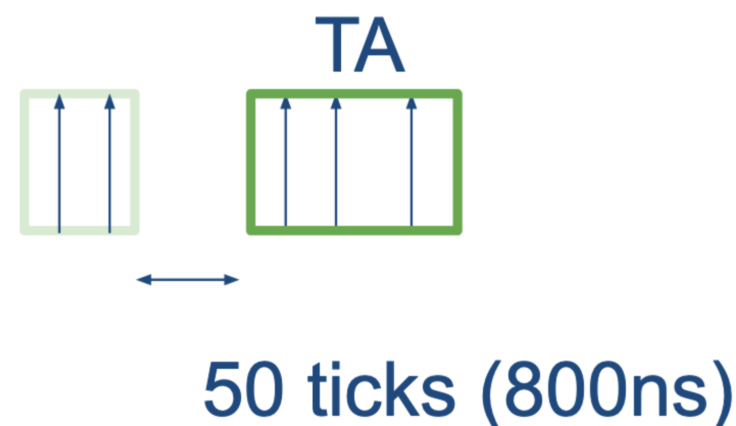
Same approach: TPs → TAs→TCs→MLT

PDS TPs generated in DAPHNE firmware

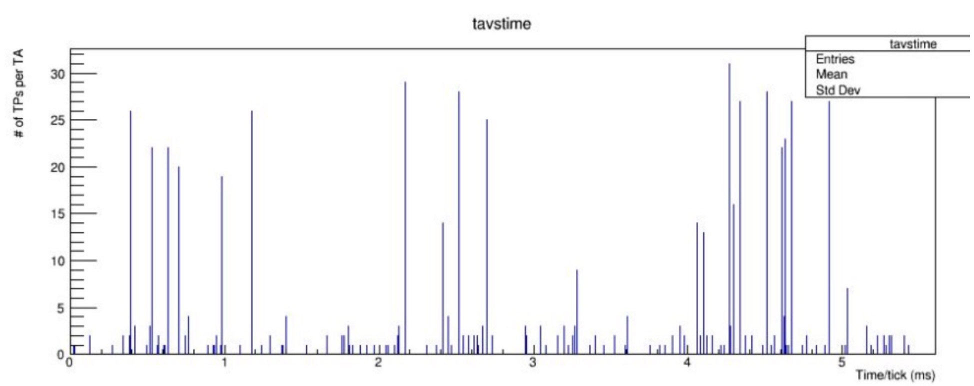


M. Arroyave

PDS TAs adapt ADCSimpleWindow algorithm



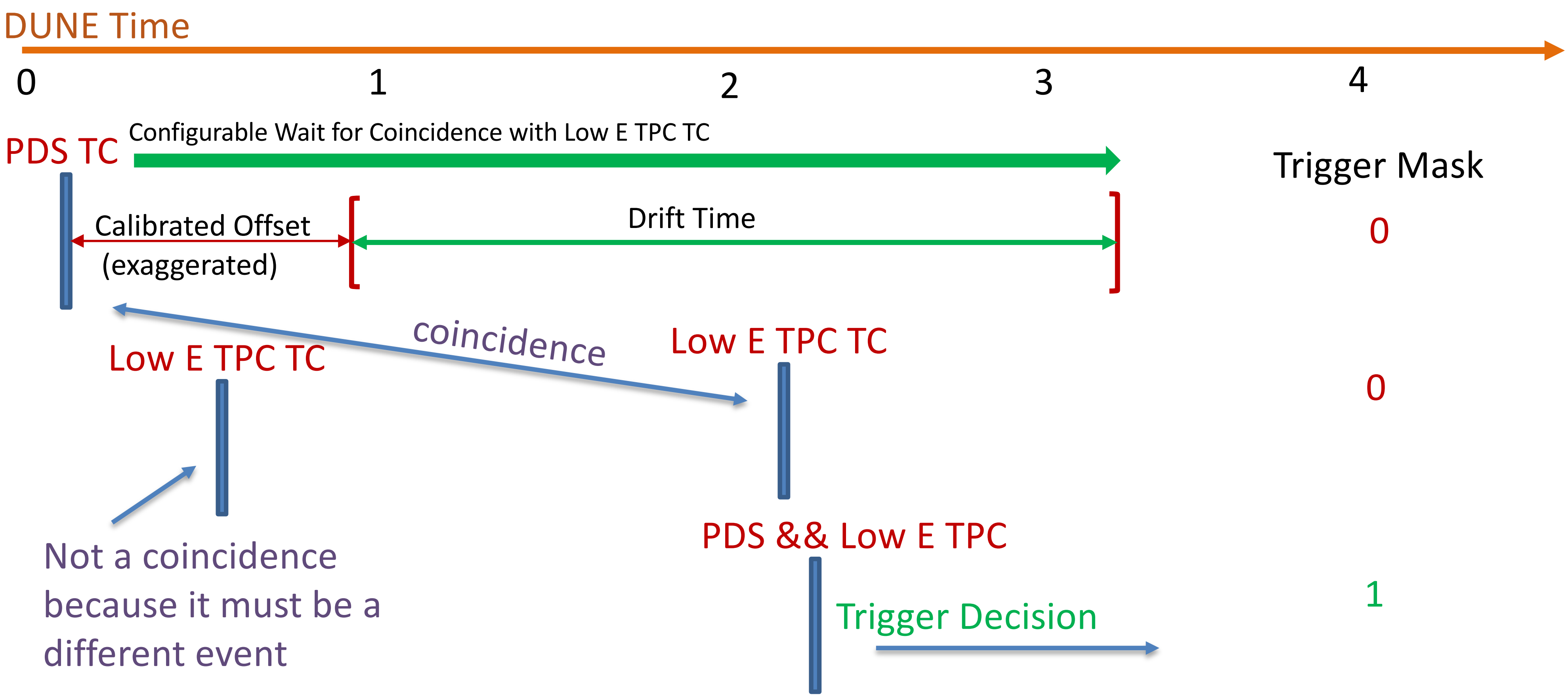
J. Soto



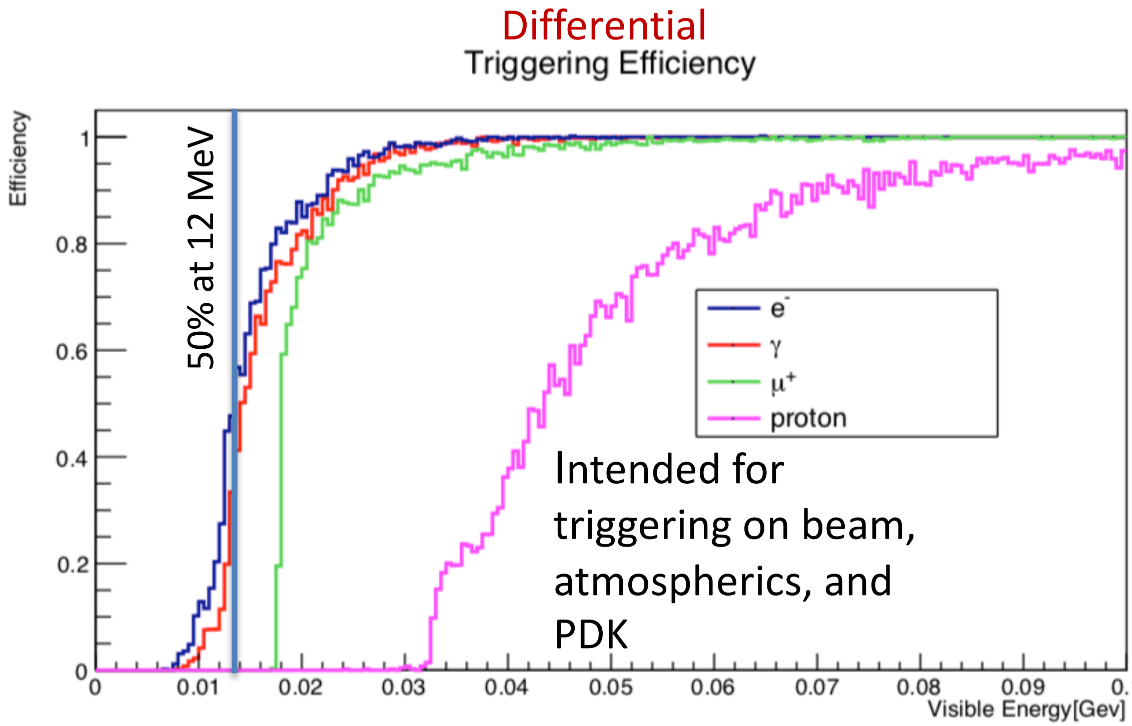
Module Level Trigger (MLT)

MLT "sees" all systems, channels, and large blocks of time

- Trigger Candidates from TPC (multiple planes), PDS, Calibrations, other modules
- Can trigger on coincidences of TCs



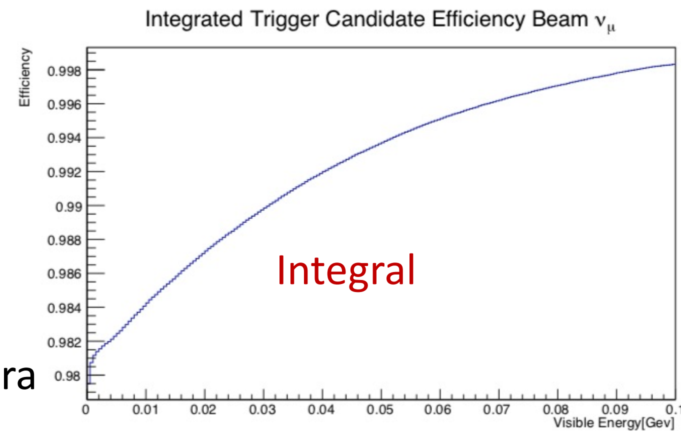
High-Energy Trigger Efficiency

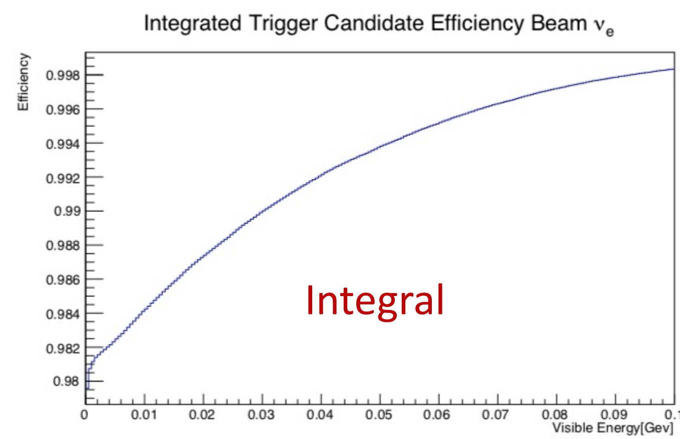


D. Rivera

- Integrated efficiency ϵ_I is given by :

$$\epsilon_I(E_{vis}) = \frac{\int_{E_{vis}}^{\infty} n_{trig}(E)dE}{\int_{E_{vis}}^{\infty} n_{evt}(E)dE}$$





D. Rivera

Threshold is set by background rate!

Threshold is set so that data volume fits within DUNE allocation if we *read out everything for 5.4 ms for every trigger*

Moving to ROI readout allows a lower threshold

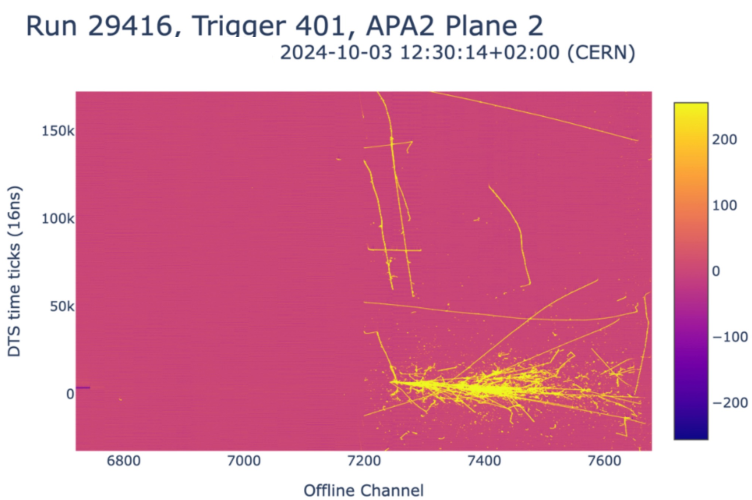
NP04 Performance

Most beam running was triggered using Central Trigger Board but tested full SW self-trigger chain:

Overall (software trigger) latencies < 500 ms << 10 s latency buffer size

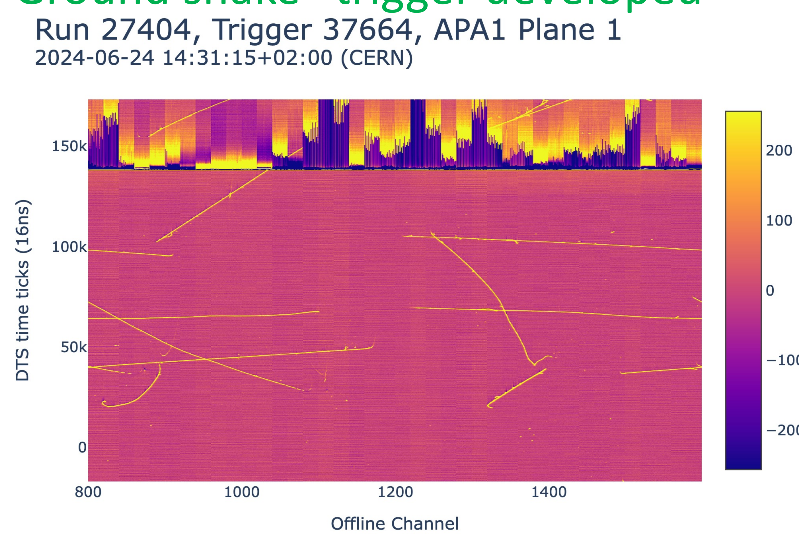
M. Rigan

High-energy “BSM” trigger developed



M. Rigan

”Ground shake” trigger developed

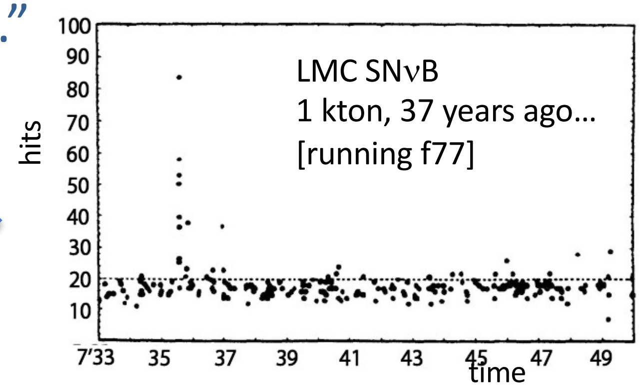


I. Hristova, W. Ketchum, M. Rigan, A. Sztuc

Supernova Bursts

There are 3 ways we trigger on neutrinos from a supernova:

1. **Standard Low E Trigger:** For a TC threshold of > 10-12 MeV, neutrinos from a supernova that produce more visible energy than that, are recorded naturally as part of the nominal "Trigger Record."



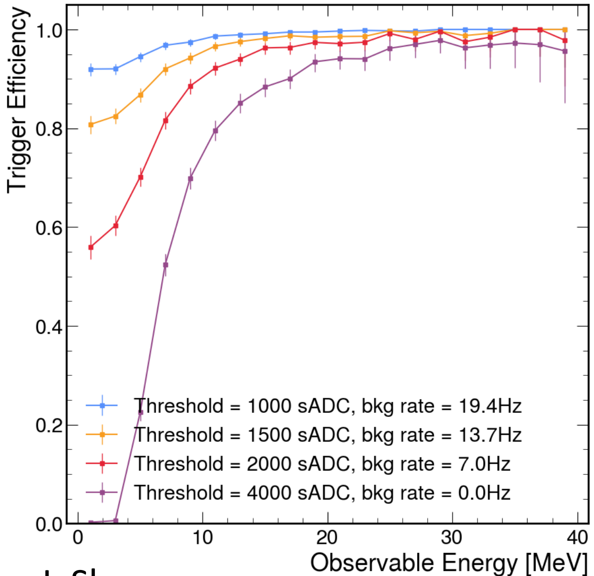
2. **Trigger Primitive Stream:** All the TPs, all the time. Zero event energy threshold

3. **SN "Burst Trigger":** When rate and spectrum of interactions is inconsistent with background fluctuations, *save every sample of every waveform* for 100 seconds. (unbiased except for front-end shaping, etc.)

(We aim for less than 1 fake SN Burst trigger per month)

Supernova Bursts

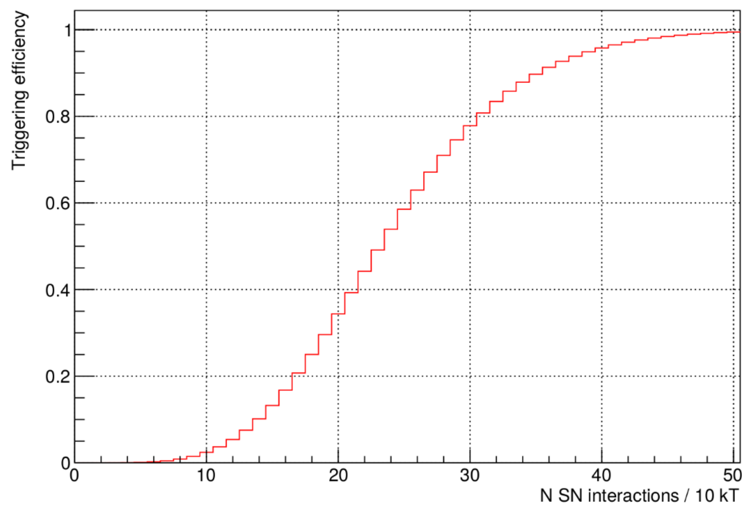
SN Burst TC efficiency



J. Shen

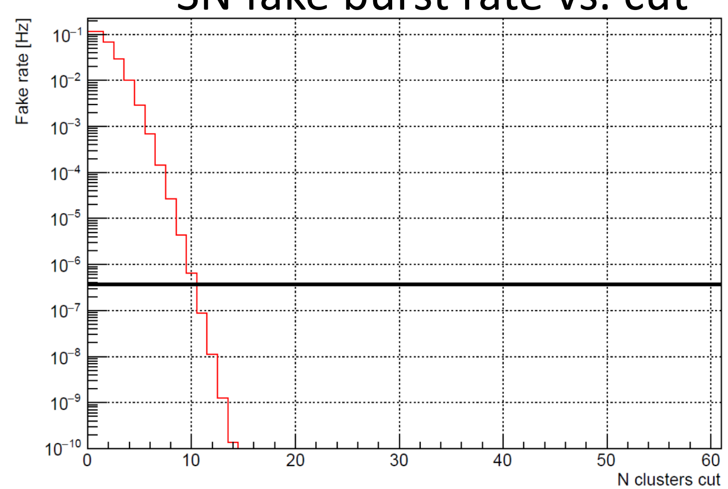
Not yet using induction TPs/TAs

SN Burst efficiency vs. interactions



T. Bezerra, P. Lasorak,.Booth, A.Borkum, J.Hartnell S.Peeters, F.Xie

SN fake burst rate vs. cut



Please note: If you have been thinking about a supernova burst trigger, it needs to be discussed within the DAQ consortium before presenting it to the outside world

Trigger Simulation!

Development of most trigger algorithms has been standalone simulation+***emulation*** tools---including replays of detector data

Very helpful for testing algorithms within dunedaq framework

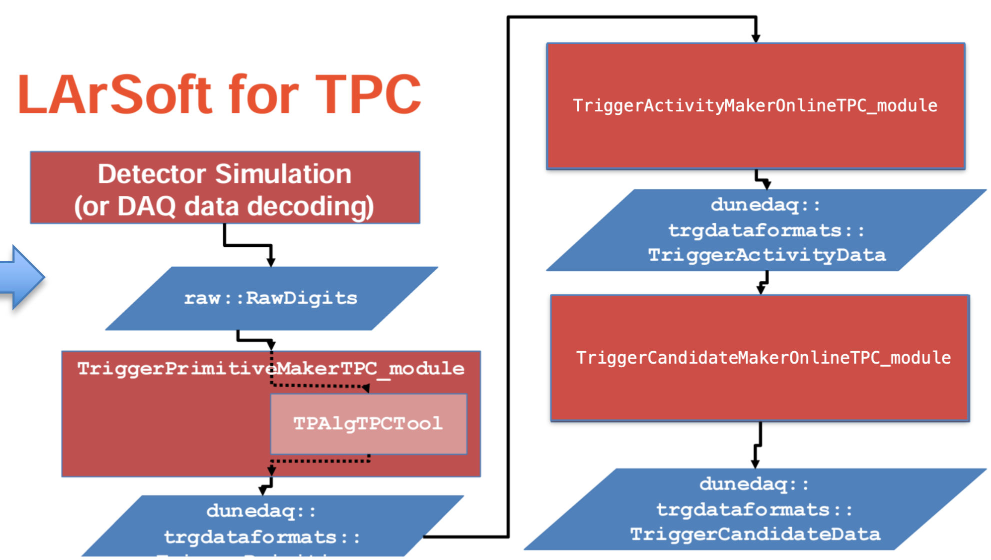
But:

- Hard to connect to physics requirements/performance
- Restricted to DAQ experts
- Hampered physics-related trigger development

Heroic effort by:

S.S. Chhibra, D. Drobner, H. Amar Es-sigir, W. Ketchum, J. Shen, K. Wawrowska and others

Trigger simulation now can be distributed with dunesw offline packages---and not just a trigger model, but the *actual DAQ code*



W. Ketchum

Trigger Simulation!

Even YOU can develop a trigger algorithm that is (nearly) DAQ-ready

- Only a few TP/TA algorithms in LArSoft
- Intention is to distribute this by default
- Run as part of standard production

DUNE DAQ Software Documentation Home

Table of Contents

How to write a trigger algorithm

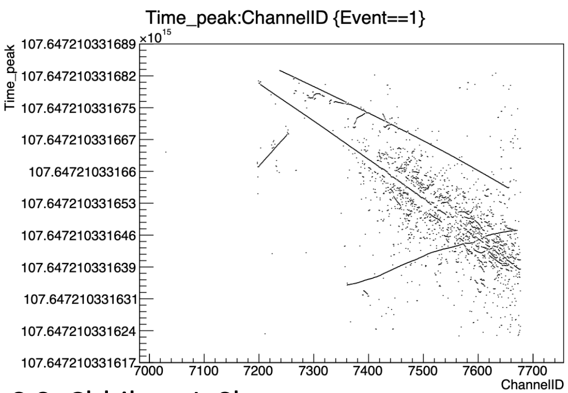
The DUNE DAQ data selection software separates physics algorithms (which make trigger activity and trigger candidate objects) from concerns about how the data objects are packaged and transported through the data selection system.

Physics algorithms are implemented in the `triggeralgs` package. To create a new physics algorithm, create a class that derives from `triggeralgs::TriggerActivityMaker` or `triggeralgs::TriggerCandidateMaker` as appropriate. I'll use `TriggerActivityMaker` for definiteness from here on: if you're writing a `TriggerCandidateMaker`, make the appropriate substitutions.

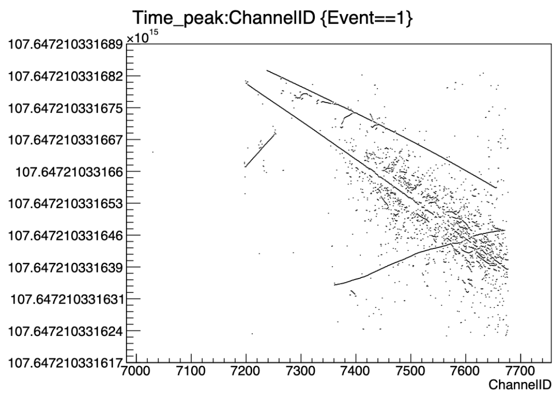
P. Rodrigues, A. Sztuc

Some early results/tests:

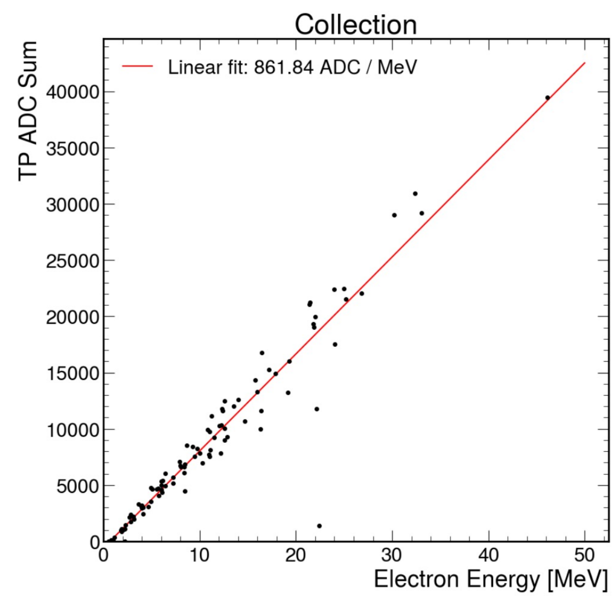
Offline



Online



S.S. Chhibra, J. Shen



Please note: If you are working on an algorithm, it needs to be discussed within the DAQ consortium before presenting it to the outside world

Trigger Simulation!

We want to hear from you if:

- Physics you want to do is not triggered on
- You have developed an algorithm that works well
  - (Also needs to be tested in emulation to see that it is fast enough)
- Something else is broken

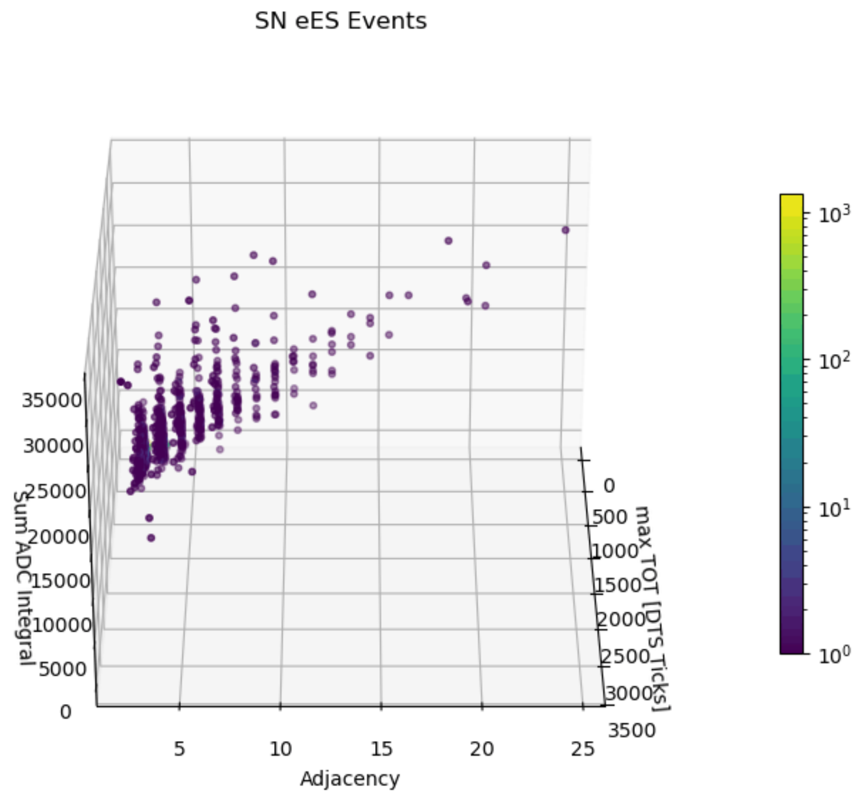
Mini-tutorial in parallel session to introduce people to running “LAr-trigger”

Summary

Trigger workshop at CERN week of April 7!

- Intro to the Data Selection/Trigger system
- Present any updated physics requirements
- New algorithm ideas and development
- "Hackathon" sessions to run both trigger and simulation

Summary



SN eES Events
35000
30000
25000
20000
15000
10000
5000
0
Sum ADC Integral
5
10
15
20
25
Adjacency
0
500
1000
1500
2000
2500
3000
3500
max TOT [DTS Ticks]
10^3
10^2
10^1
10^0

Back ups

ROI Selection and Rate

Can have a higher trigger rate if data/trigger is smaller size:

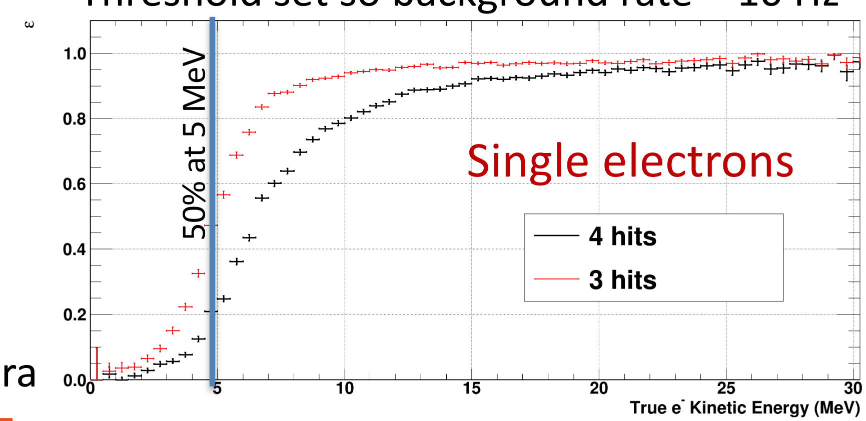
1. Halve readout window
2. Write out only APAs with trigger activity (TA)
3. Use a much narrower readout (100 μs) window around hits ("zero suppression")
4. Fully localize TA and use 100 μs window for readout

Table 2:

Data Reduction Approach	Event Size (Uncompressed)	Max Trigger Rate	Enabled Physics
Nominal	6.075 GB	0.078 Hz	Beam, NDK, Atm.
2.7 ms Readout Snapshot	3.3075 GB	0.156 Hz	Unknown
APA-Localization (Cosmics)	0.243 GB	1.95 Hz	*hep* solar ν
APA-Localization (Low-E)	0.041 GB	11.7 Hz	^{8}B solar ν, neutrons,Rn
Zero Suppression	0.040 GB	12.0 Hz	^{8}B solar ν, neutrons,Rn
TA Localization +Zero Suppression	14.6 kB	32.5 kHz	^{42}Ar, ^{40}Cl, *pep* solar ν?

Of course, ROI can depend on type of Trigger Candidate

Threshold set so background rate ~ 10 Hz



Thiago Bezerra

Frequently Asked Questions

- Why have you focused on the TPC rather than the PDS?
  - PDS triggers are in principle great! Naturally simple and inclusive, and fast!
  - TPC more advanced than PDS design, but soon PDS should catch up
  - TPC requirements are tighter than PDS; TPC *must work* as designed
  - PDS in HD has a very position-dependent response
  - But we DO plan to use PDS and in VD it could easily become most useful
- I am interested in a very rare process, how will we trigger on it?
  - If visible energy > 20 MeV, inclusive trigger should cover this
  - If visible energy 10-20 MeV, please talk to DAQ about TA/TC makers
  - If visible energy < 10 MeV, make sure neutrons, Rn, etc., don't swamp it
- How do I check that DUNE Data Selection will trigger on my physics of interest?
  - We are working on a complete trigger model for LArSoft
  - But we need help! Not very many people with the right expertise
  - And offline simulation is not really set up for this

High Energy Group Studies Valuable to DUNE Data Selection

Questions for all of you:

- We are likely to limit the readout to some region-of-interest, to save storage and streamline analysis---how aggressive is too aggressive for HE analyses? Should we read out 1 APA? Just 1000 channels? 100 microseconds/channel? What about neutrons? Guidance here would be very helpful!
- Are there "high-energy" processes that in fact lead to low visible energy (~10 MeV or so) that people want high efficiency for? (e.g., NC interactions producing a few neutrons, de-excitation γs, etc.)?
- What would you like in the offline trigger model, and are you able and interested in helping here? At the very least, once the model exists, it would be good to let us know that we are not missing anything!
- Are there any online/nearline analyses that might benefit from exclusive tagging before writing to disk?

Summary

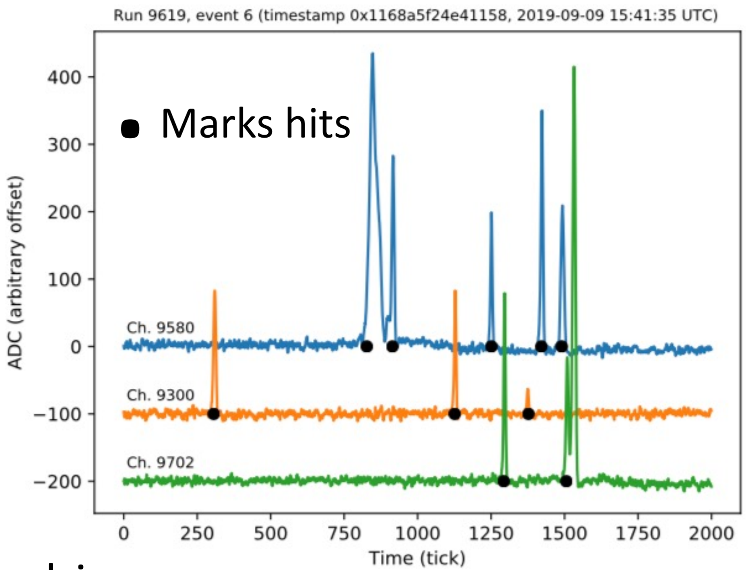
- DUNE trigger will be inclusive at high energies
  - Aim for > 99% integral efficiency above 100 MeV
- Data Selection system is hierarchical
  - Channel-level Trigger Primitives (e.g., integral, peak, ToT, etc.)
  - APA/CRP-level clusters (Trigger Activities)
  - System-level Trigger Candidates
  - Module-Level Trigger
- Region-of-interest readout is likely
  - What ROI makes sense for high-energy events, in time and space?
- Data Selection group is always open to trigger ideas that leverage new physics, please join us if you are interested!

Backups

Trigger Primitives (TPs)

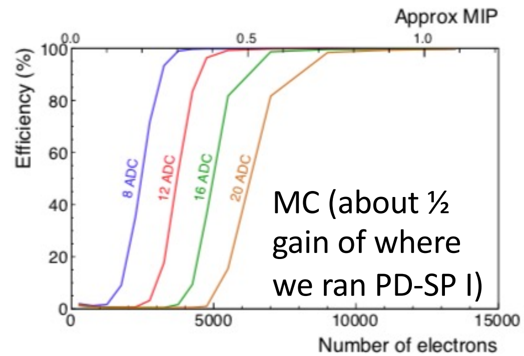
Hit finding

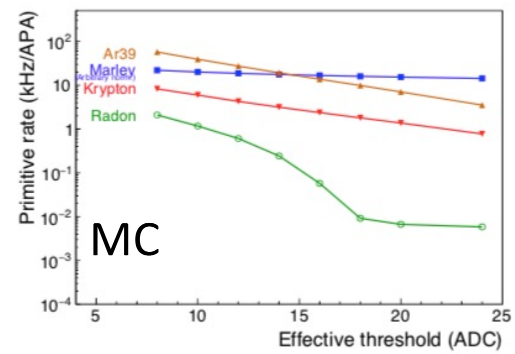
Example from ProtoDUNE-SP



P. Rodrigues

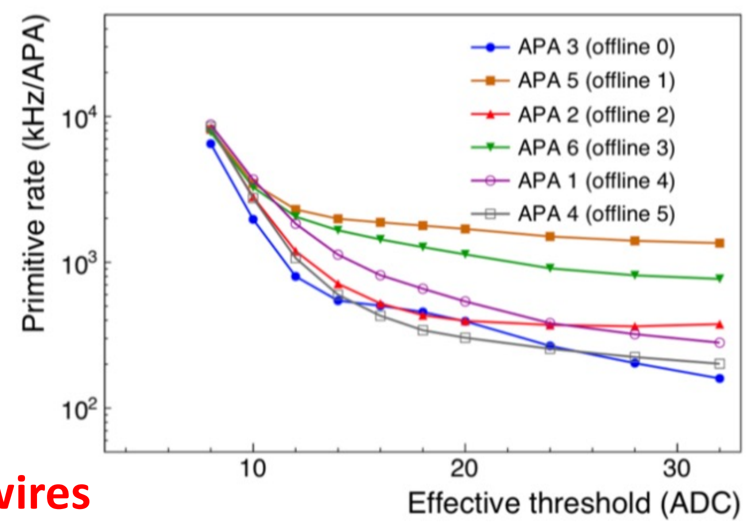
Raw noise in ENC at PD I was ~600 e (collection)





TP threshold was around 1/4 MIP-equivalent, or around 250 keV_{ME} (per wire)

Raw noise RMS in PD I was 3-4 ADC above pedestal



Has been studied also for VD strips instead of HD wires

Trigger Design Considerations

1. *Inclusive*
    - Above 20 MeV visible E we aim to trigger on everything with close to 100% efficiency (there is no “cosmic-ray trigger”, “atmospheric ν trigger”, etc.)
    - Possible that SOME exclusivity can help with radiologicals below 10 MeV
    - But don’t want different algorithms for different signals (e.g., CC vs. ES)---let the offline analysis do the hard work
2. *Simple*
    - Measuring efficiencies and their uncertainties is complicated
    - Need a reasonable latency so that we don’t run off the front-end buffers
    - We never ever want code to hang (fewest possible dependencies!)
3. *Hierarchical*
    - Trigger is hierarchical: highest level sees most information
    - This applies to space/time/systems/views/external input
    - Allows for more sophistication downstream where rate is lower
4. *Physics*
    - A trigger that rejects ^{39}Ar by 1000 doesn’t add any physics

Data Volume for High Energies

Our allocation is 30 PB/year for 2 modules

- Most conservative assumption is we read out entire detector for 2x drift time for every high-energy trigger
- At high energies, totally dominated by cosmic rate of 4000/day/ 10 kt module
- (And we want to keep every cosmic!)

For HD:

=384,000 ch * 14 bits/sample * (1/512 ns)*5.4 ms * 4000/day*365 days
=10.35 PB/year

Most importantly: This means our FD trigger can be *inclusive*

Above some threshold, we will trigger on EVERYTHING at > 99%

(At ProtoDUNE, however, this is NOT true).

Data Volume for High Energies

Our allocation is 30 PB/year for 2 modules

- Most conservative assumption is we read out entire detector for 2x drift time for every high-energy trigger
- At high energies, totally dominated by cosmic rate of 4000/day/ 10 kt module
- (And we want to keep every cosmic!)

For HD:

=384,000 ch * 14 bits/sample * (1/512 ns)*5.4 ms * 4000/day*365 days
=10.35 PB/year

- No compression assumed
- Does not include low-energy (< 10-20 MeV) triggers
- Does not include headers, trigger-specific information, PDS, or calibrations
  - (These are all 10-20% additions)

(But it is unlikely we're going to read out every channel)

Data Selection Output

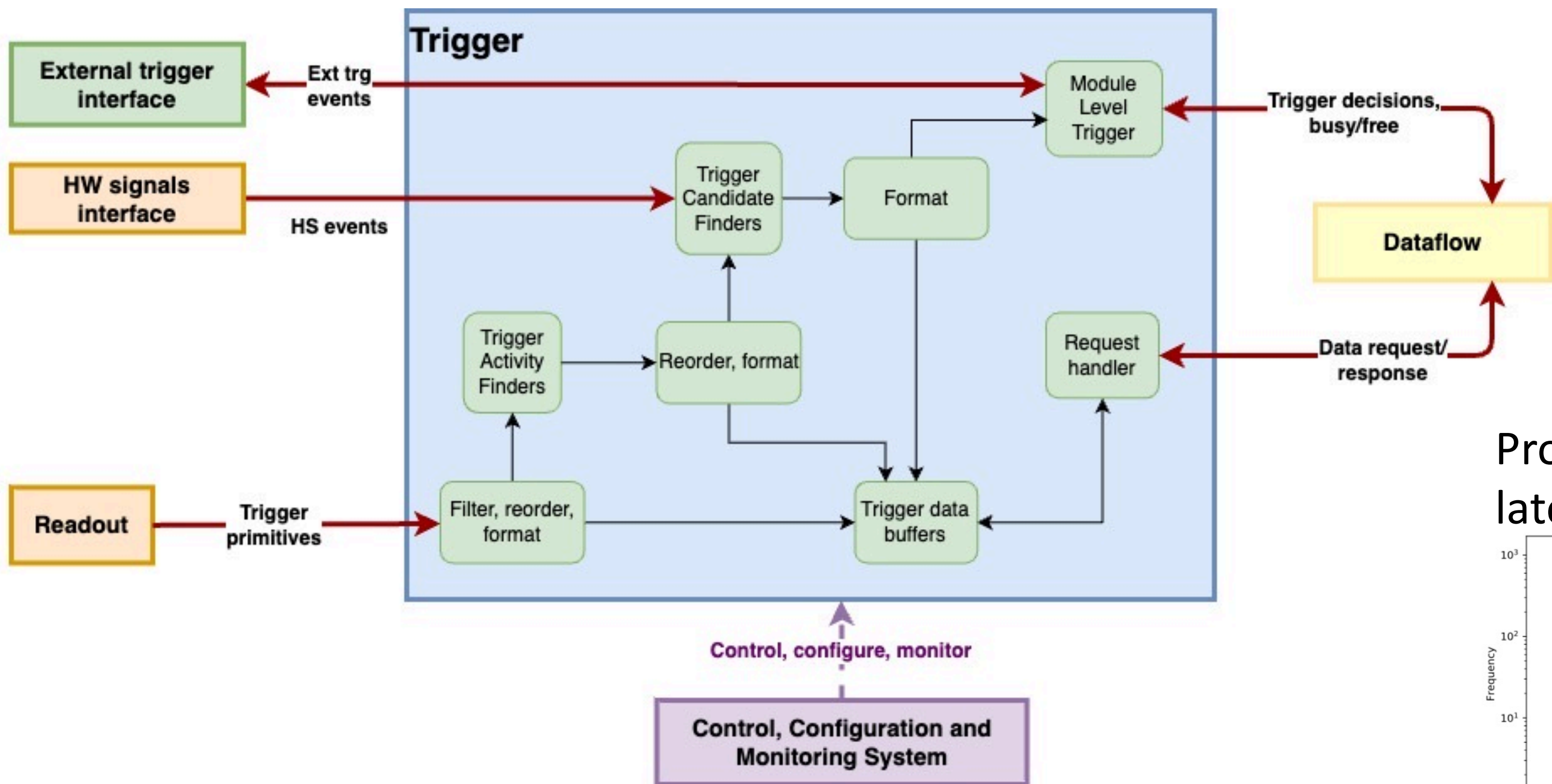
- Stream of trigger primitives (x/u/v and PDS)
  - TPC threshold ~250 keV
  - PDS thresholds 1-2 pe
  - Record time, charge integral, peak, time-over-threshold
  - Triggerless (other than channel thresholds)
- SN Burst data: 100 seconds of ***everything***
  - Target fake rate of 1/month
- Nominal Trigger Records---typically single interactions/decays
  - Limited by total storage requirements
  - Size of each trigger record depends on ROI selection (if any)
  - Strictest ROI likely on lowest energy data

Note: Supernova Burst Neutrinos are collected by all 3 paths; second path collects *everything* but time to analysis is longer than others.

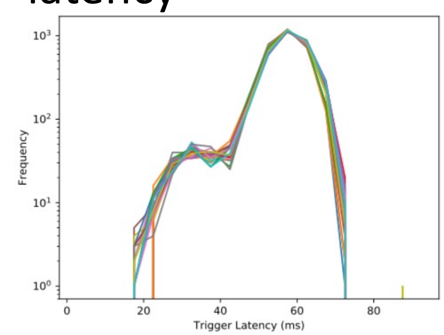
Trigger Flow Diagram

~20 servers/module

(no GPUs or other acceleration)



ProtoDUNE I trigger latency



A lot of the complexity is in dealing with a realtime system---
Making sure things are in time order, dealing with problematic channels, etc.
We do not expect algorithms to take much time.

TPC Triggering

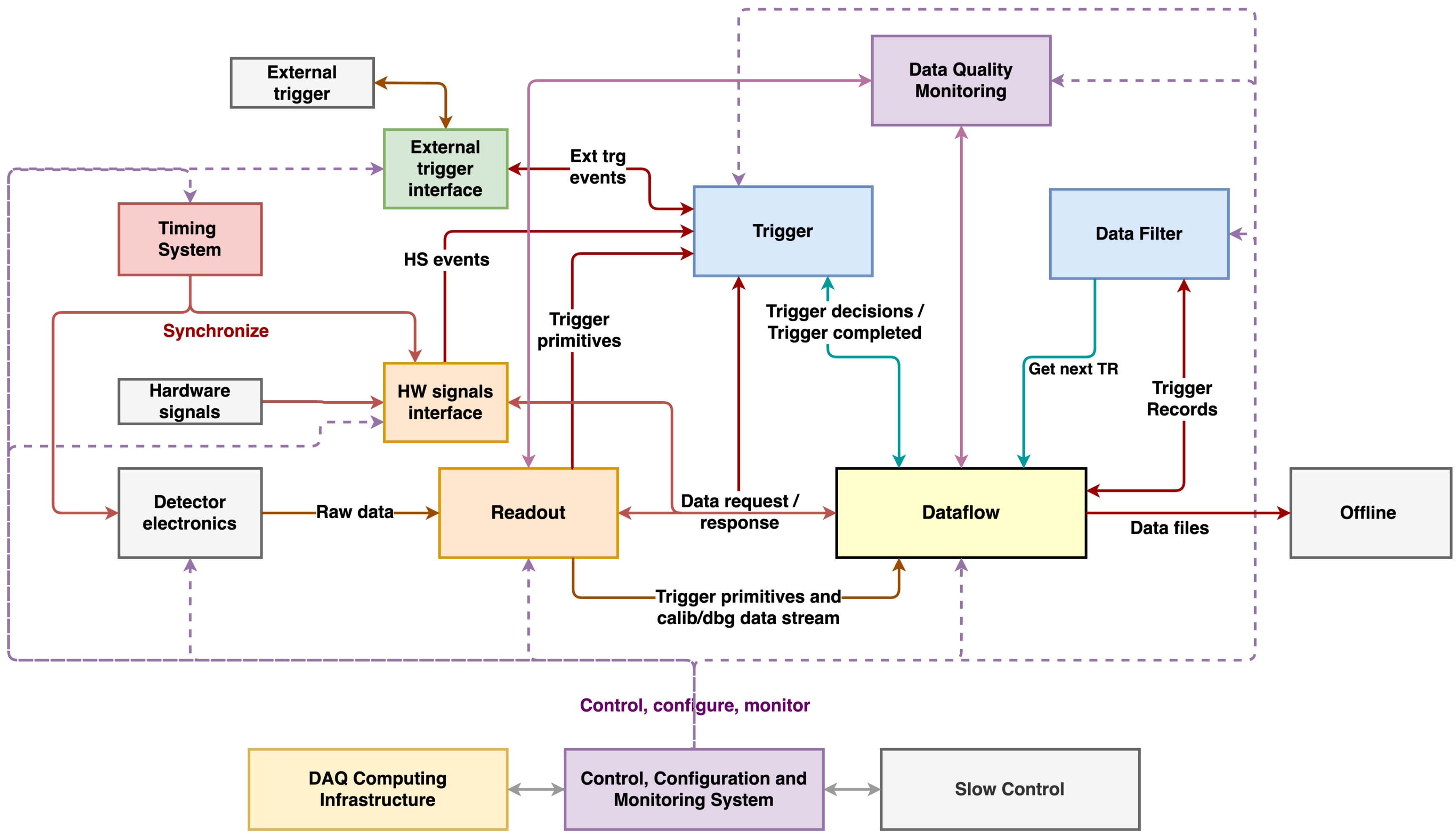
Basics

A quick note about the word "threshold"….

- Trigger Primitives have a "hit" threshold
- Trigger Decision based on various "event" thresholds
- Supernova Bursts have a "burst threshold"
  - Once a burst is triggered, the data is acquired with zero threshold for 100 s

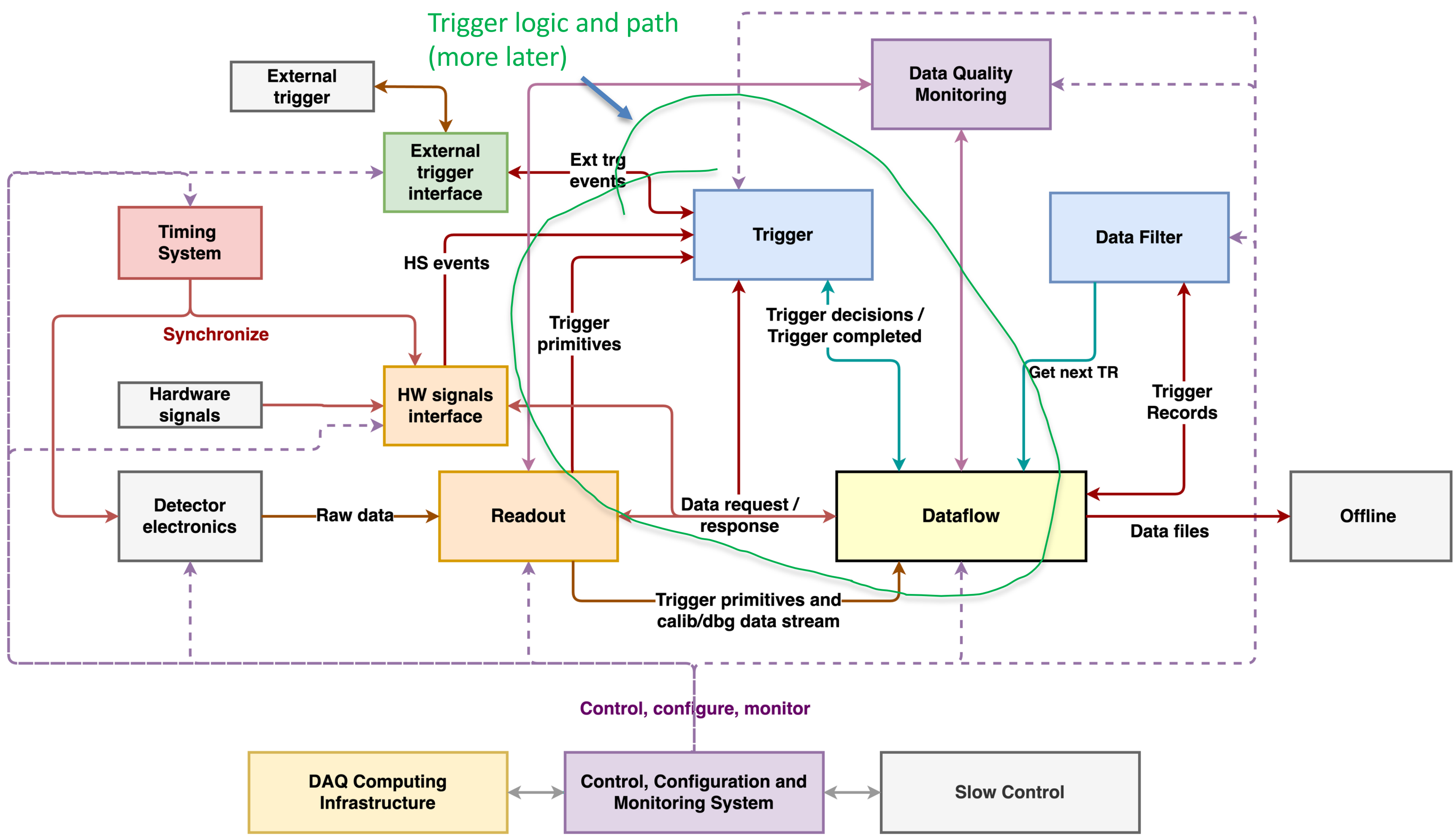
DUNE Trigger Basics

DAQ Overview:



DUNE Trigger Basics

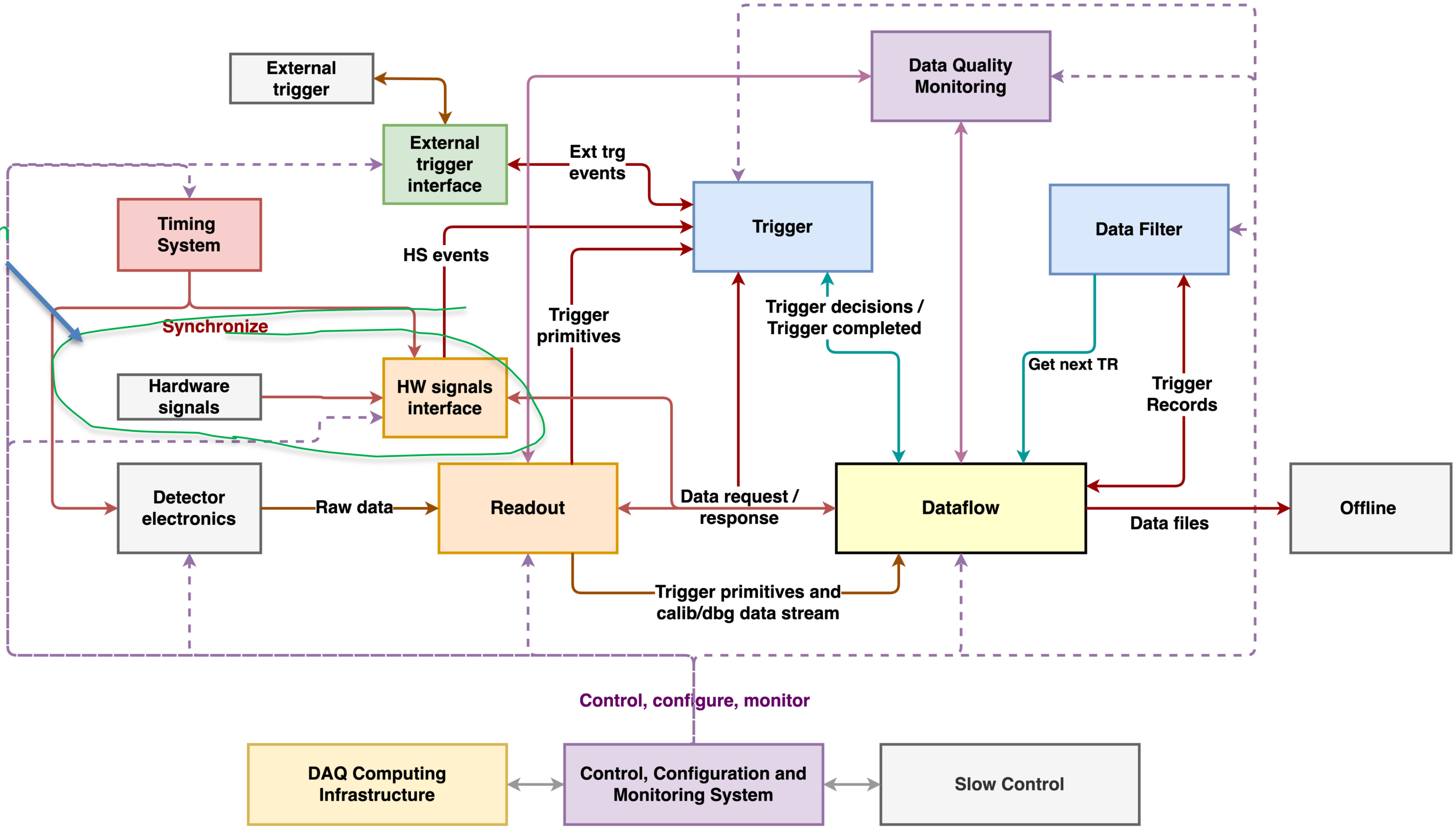
DAQ Overview:



DUNE Trigger Basics

DAQ Overview:

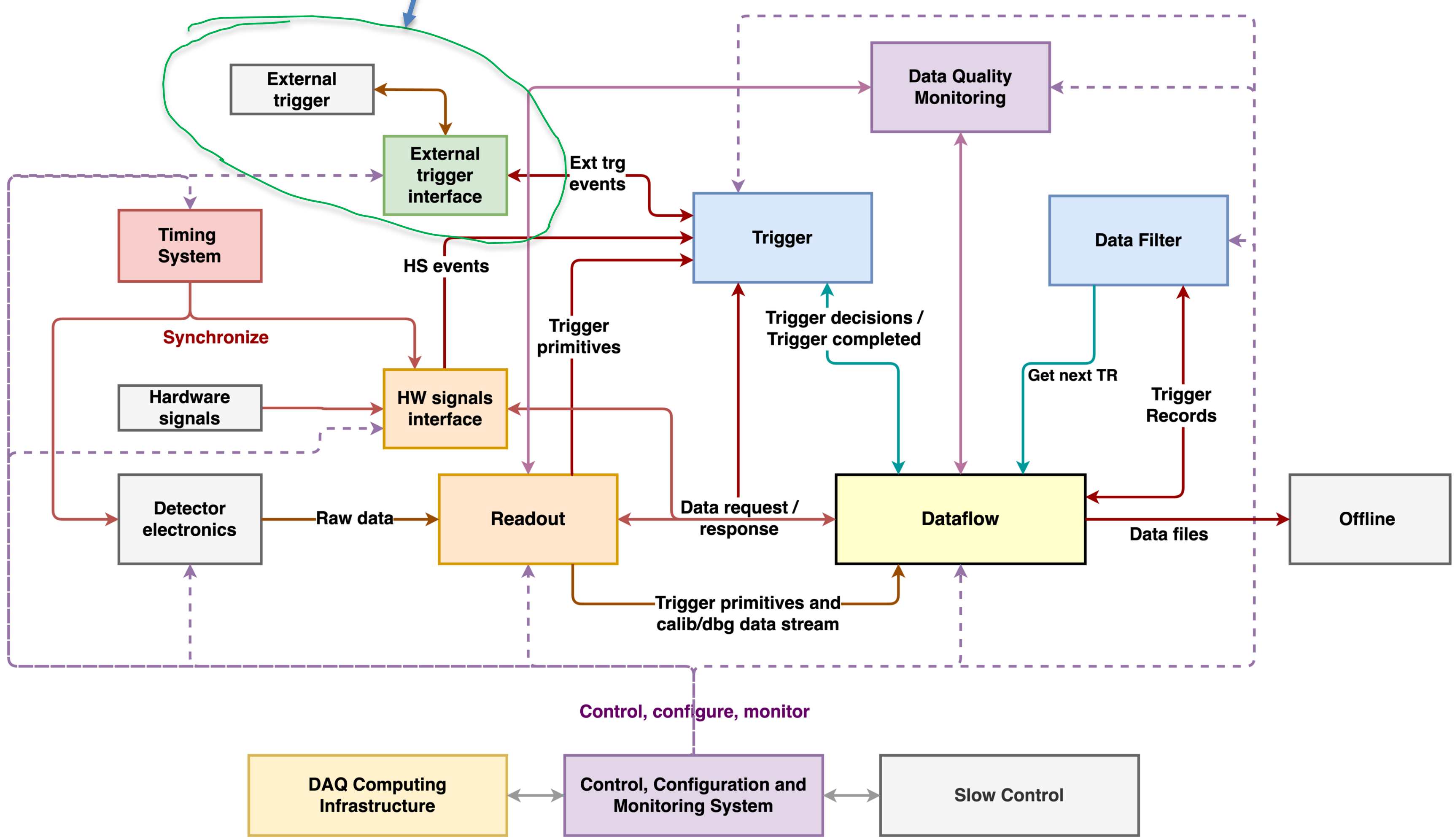
Triggers from hardware: calibrations, “test” scintillator telescope, test stands



DUNE Trigger Basics

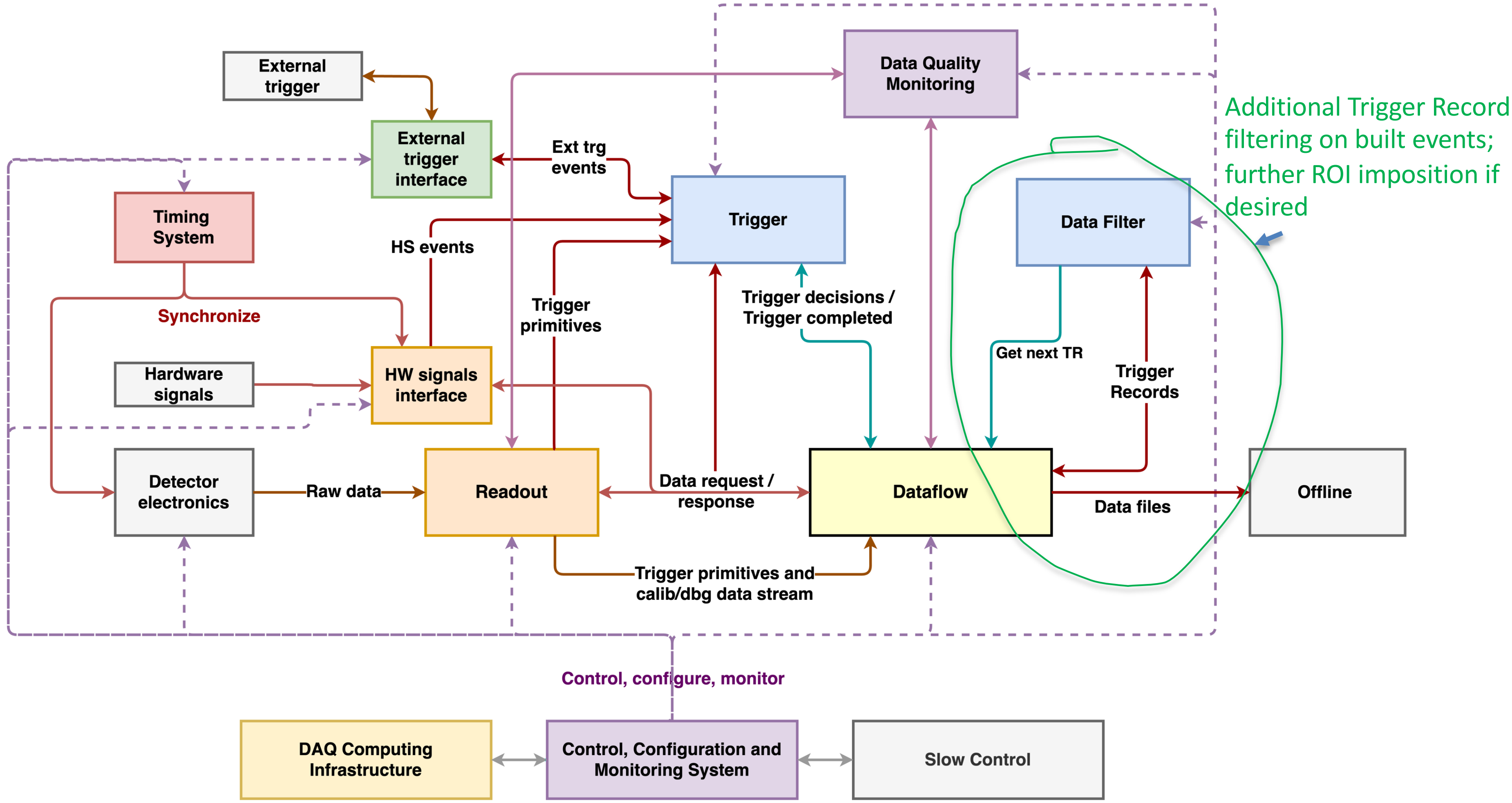
DAQ Overview:

Triggers from other modules, other experiments, beam



DUNE Trigger Basics

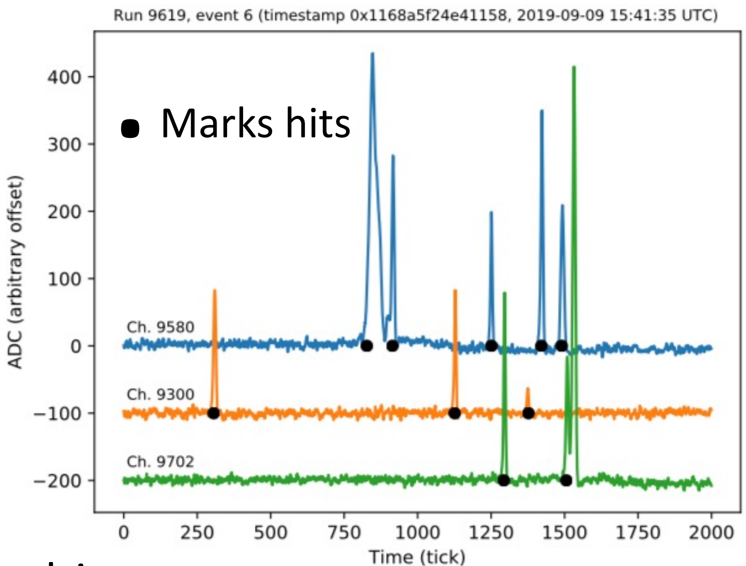
DAQ Overview:



Trigger Primitives (TPs)

Hit finding

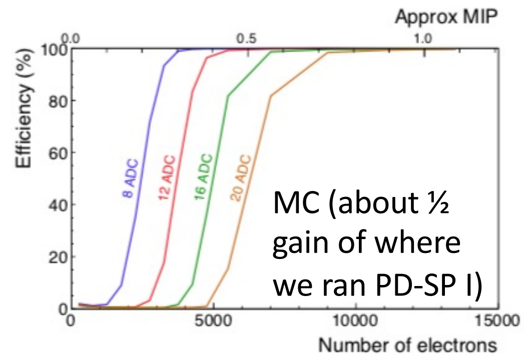
Example from ProtoDUNE-SP

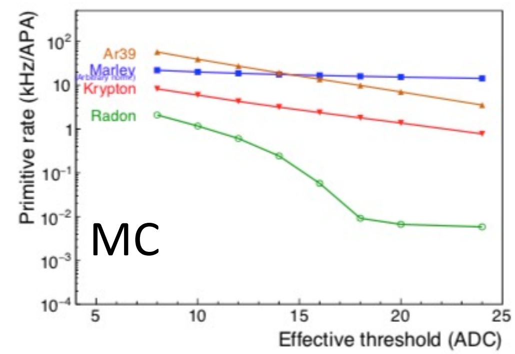


P. Rodrigues

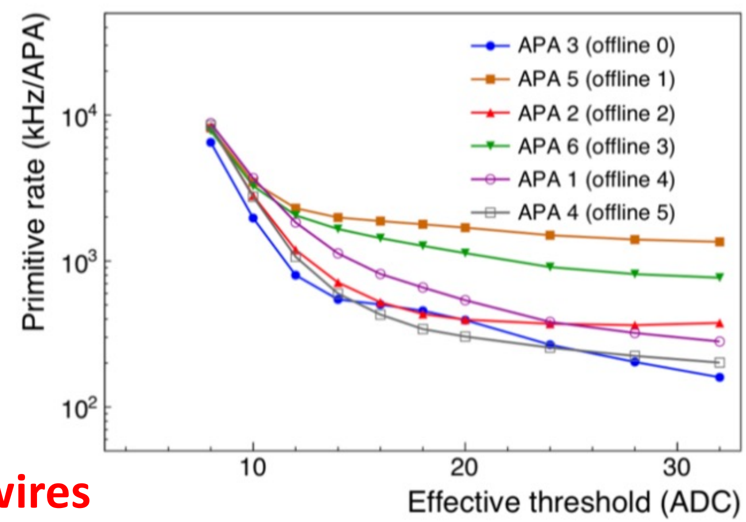
TP threshold was around 1/4 MIP-equivalent, or around 250 keV_{ME} (per wire)

Raw noise in ENC at PD I was ~600 e (collection)



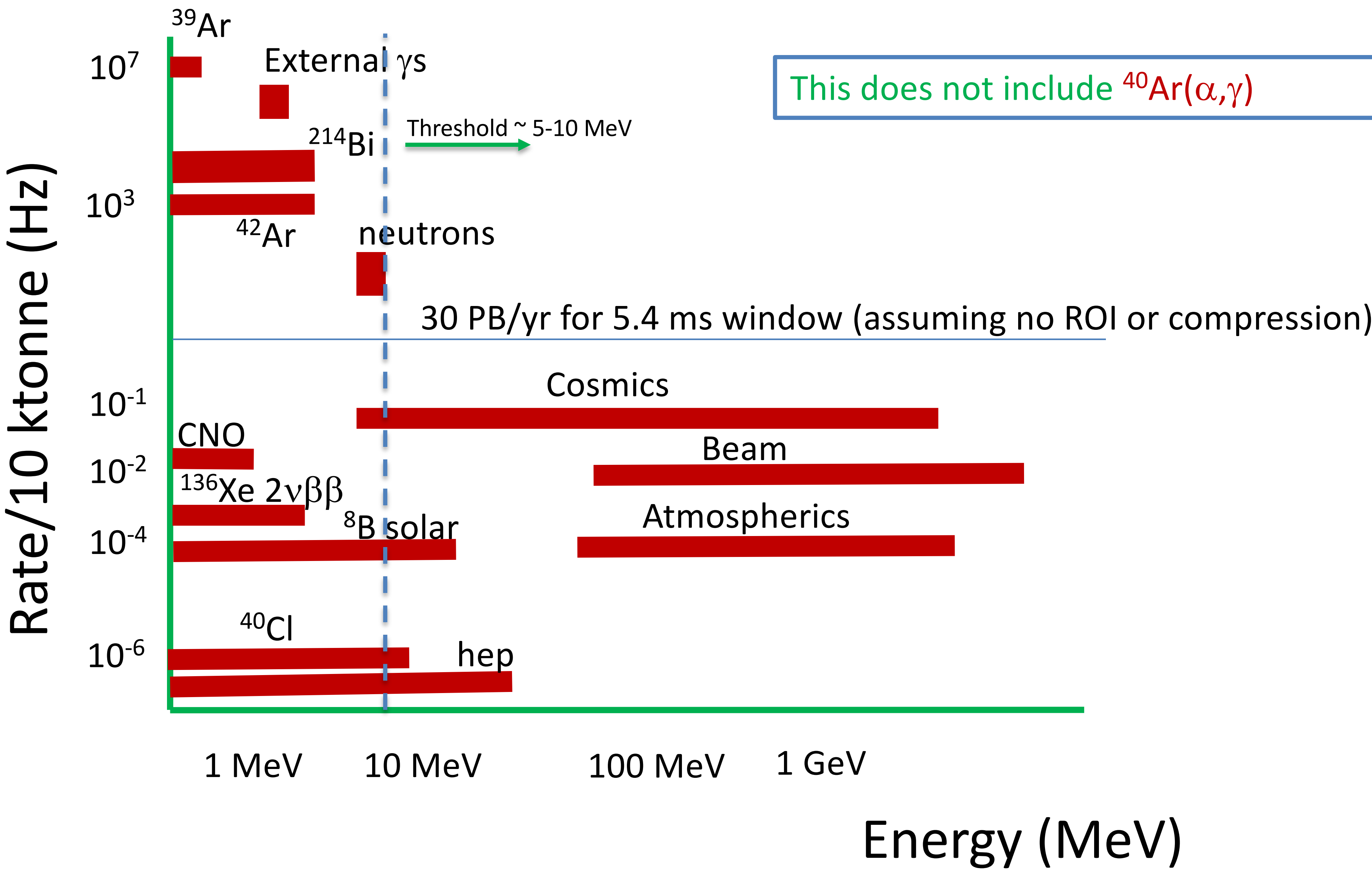


Raw noise RMS in PD I was 3-4 ADC above pedestal



Has been studied also for VD strips instead of HD wires

Background and Storage Limits for Triggering



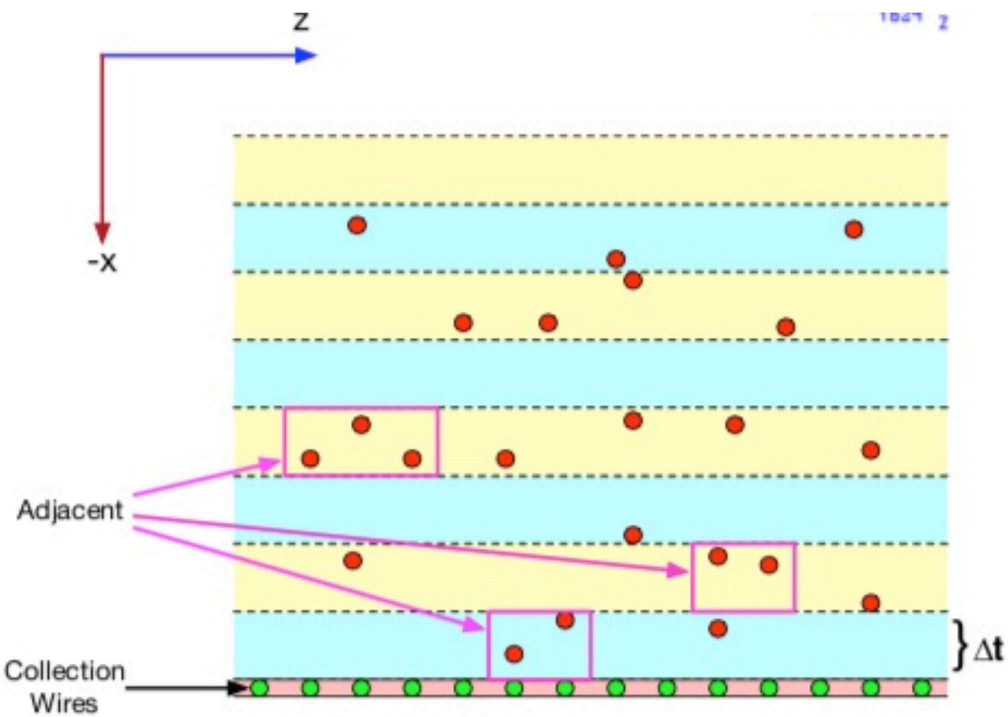
This does not include ^{40}Ar(α,γ)

Trigger Activity

Clustering and Cutting

Cannot simply sum up all charge---in 10 ktonnes and a full drift, this is about a GeV of charge.

Need some kind of clustering as first stage of triggering, then cut on number of adjacent channels, total charge, max charge, max time-over-threshold



D. Rivera

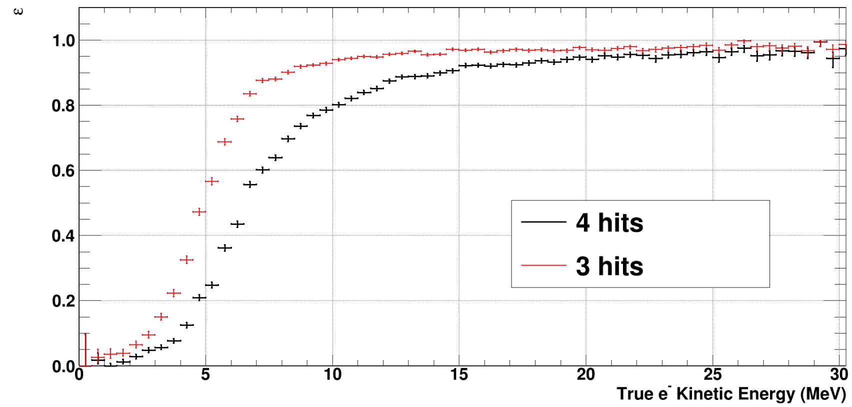
Thresholds is set conservatively assuming 5.4 ms readout of *all* channels had data rate < 25% of cosmic data rate (~1000 ev/day)

Radiological rate acceptable with:

- N_{adj} >= 8 wires
- Cluster charge sum > 7000 ADC counts
- Max integrated wire charge > 6500 counts
- Max time-over-threshold >= 45 ticks

Supernova Bursts

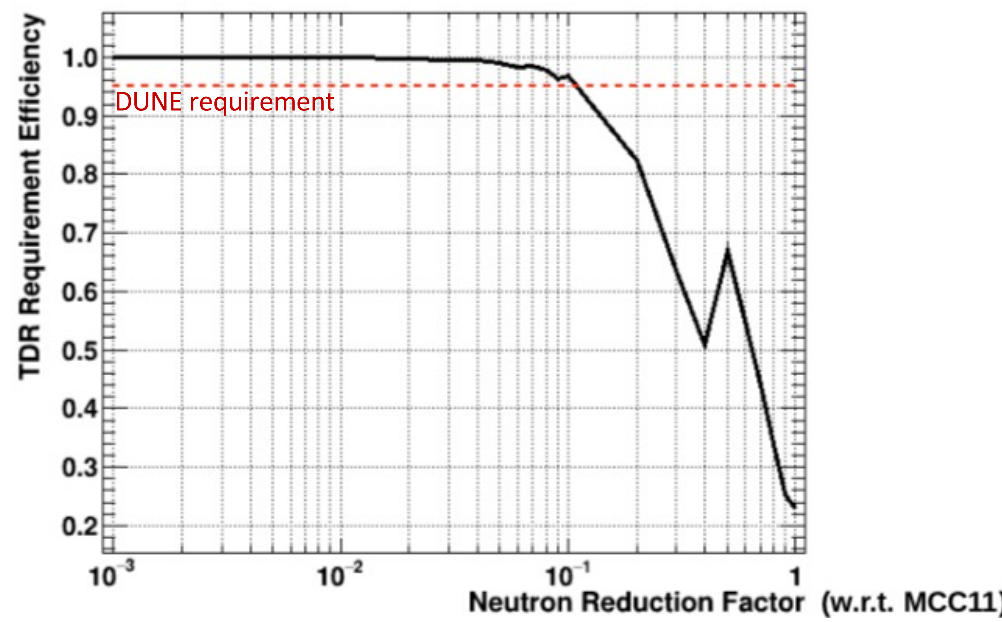
Additional handle: Time and energy profile



T. Bezerra

Can accommodate more aggressive “trigger activity” threshold but lowering single-interaction threshold hurts without energy-weighting

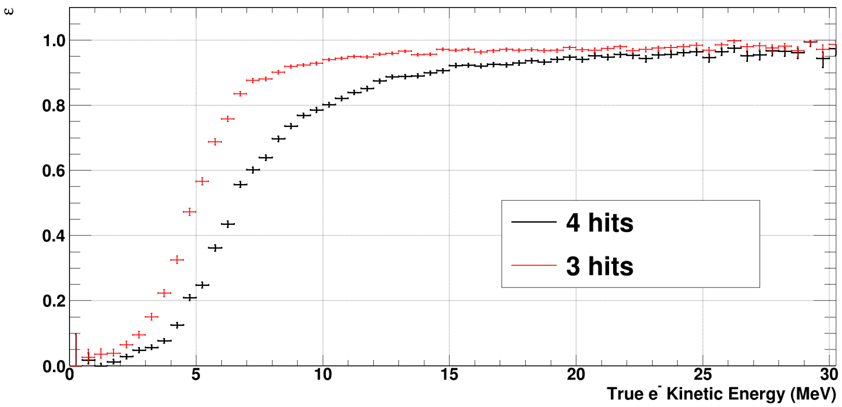
Not including energy profile



Note: we read out ***everything*** for 100 s if we detect a burst---event efficiency does not matter (much) except outside of that window

Supernova Bursts

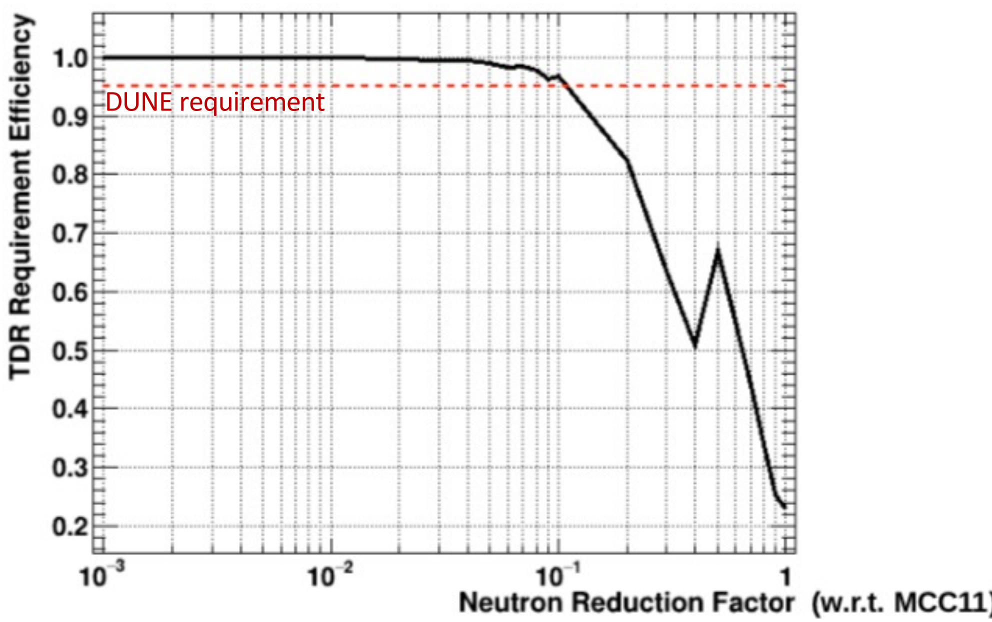
Additional handle: Time and energy profile



T. Bezerra

Can accommodate more aggressive "trigger activity" threshold but lowering single-interaction threshold hurts without energy-weighting

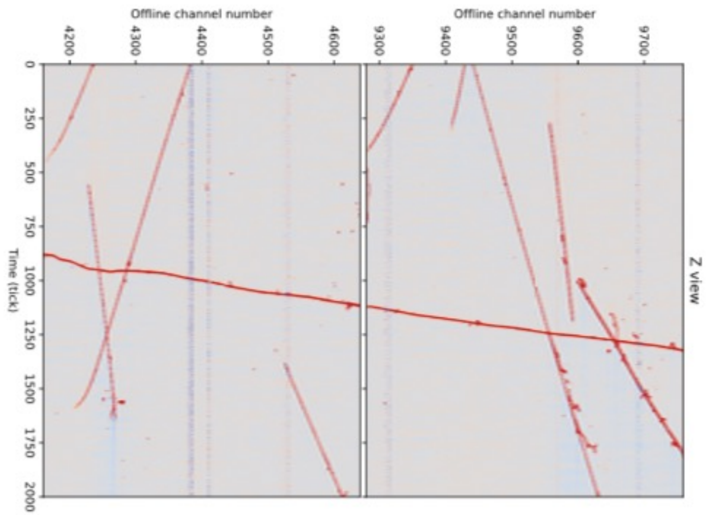
Not including energy profile

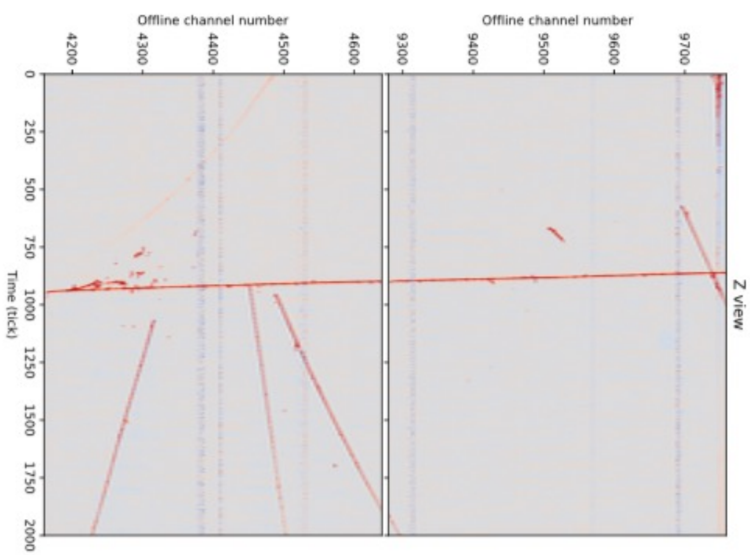


AND: Normal triggering continues in parallel with SN burst trigger---all events above threshold are built, and threshold can be *dynamic*.

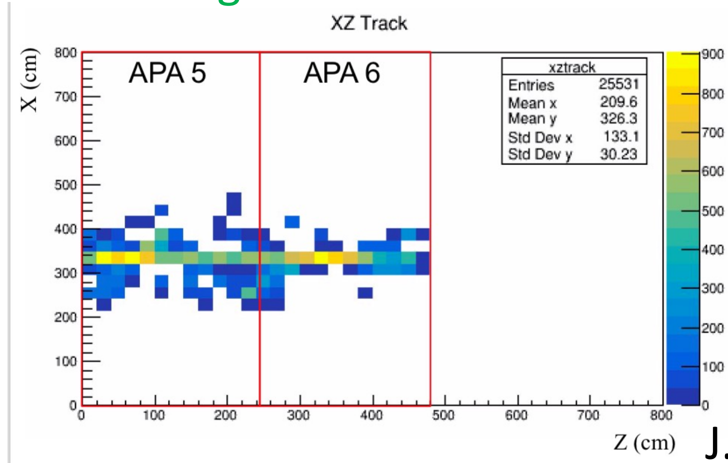
Performance at PD-SP 1

Horizontal Muon Trigger (exclusive)

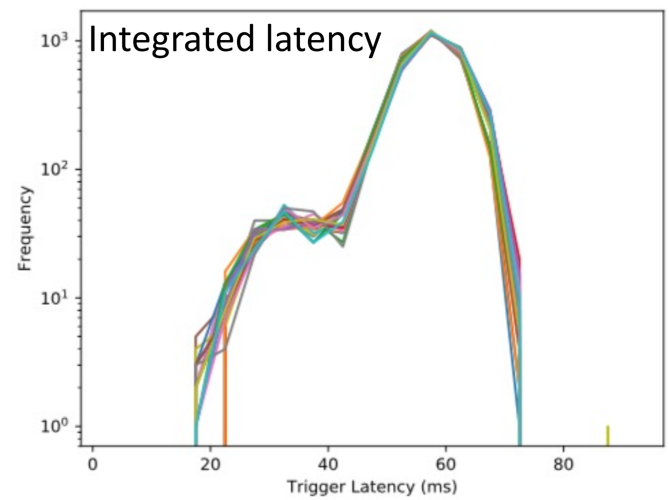




Average over several events



J. Sensenig



Time from data arriving at FELIX to Trigger Decision

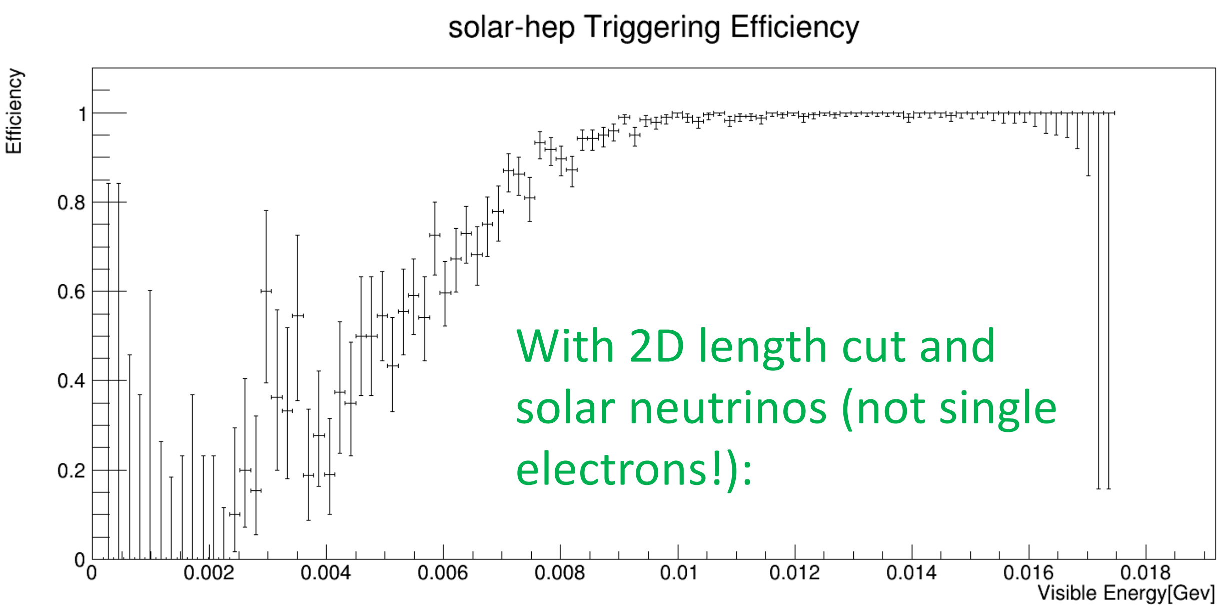
Buffer depth was 1 second

Moving Lower in E

Can exploit topology of TPC:

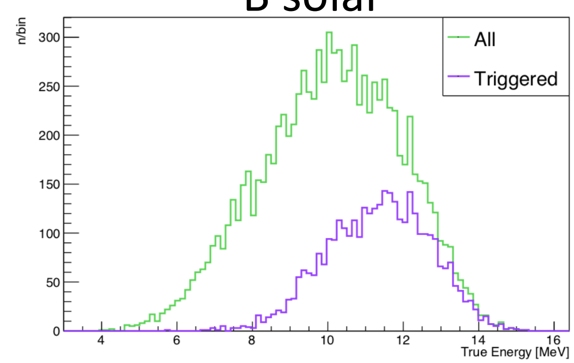
Use (collection-wire) primitives to create a “2D track length cut”

$$\mathcal{L} = \sqrt{(adj_{max} * 5mm)^2 + (TOT_{max} \times v_{drift})^2}$$

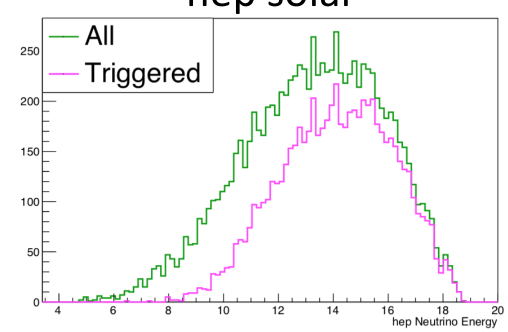


With 2D length cut and solar neutrinos (not single electrons!):

^{8}B solar



hep solar



Trigger rate is ~ 100 Hz from backgrounds

D. Rivera

Triggering is Not the Challenge

Storage Is