Data Collection & Processing at the Pole

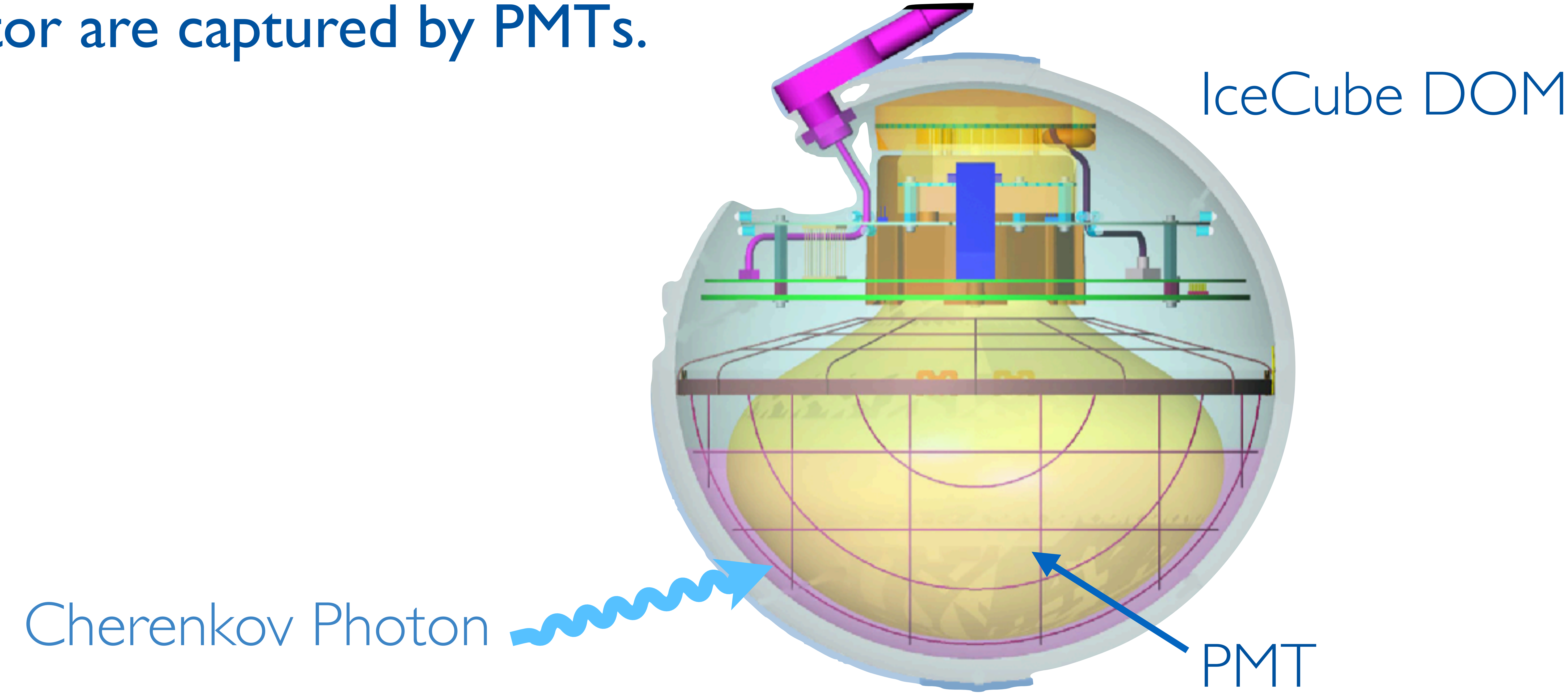
IceCube Summer School
June 5, 2025

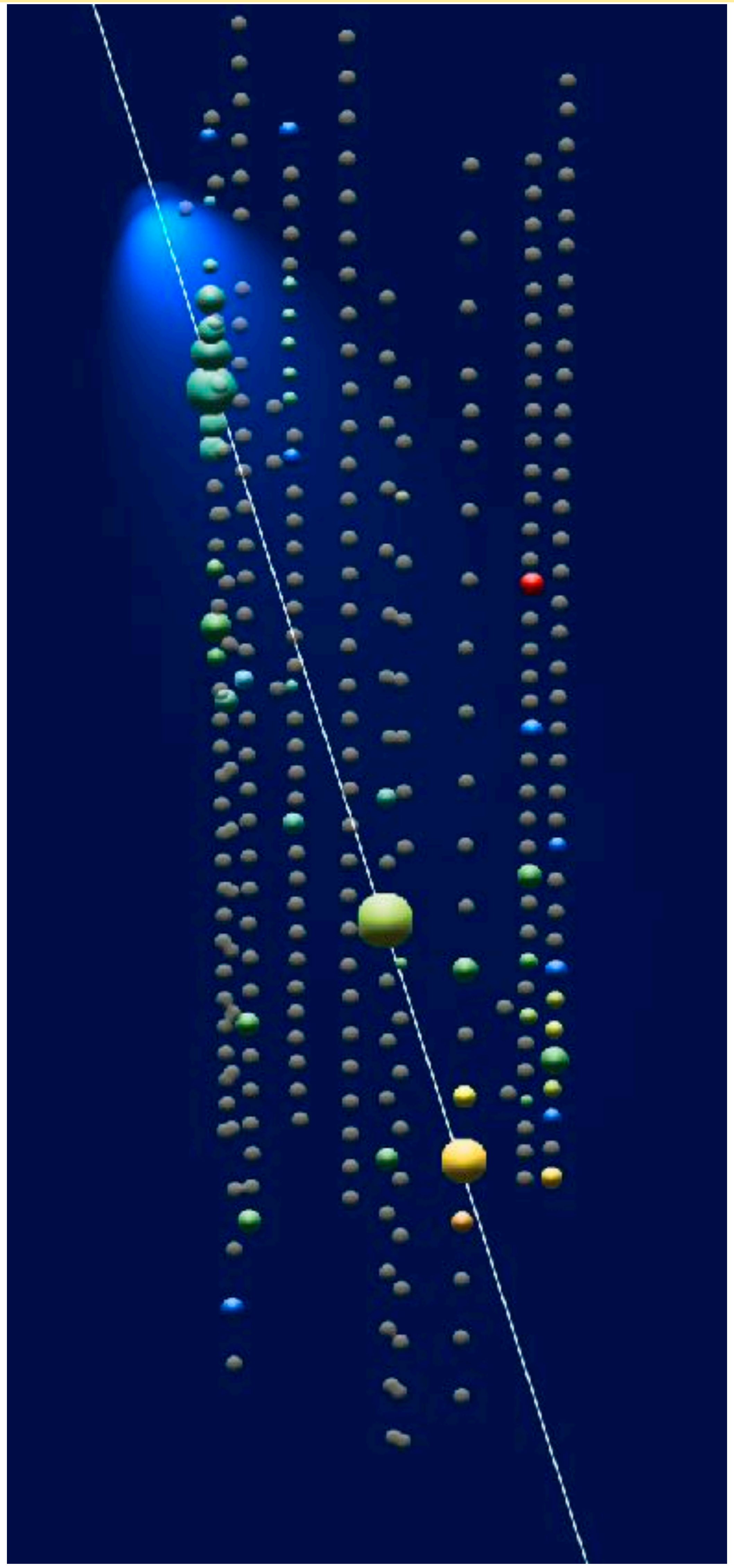
Miles Garcia
milesg@udel.edu

In-Ice Neutrino Detection

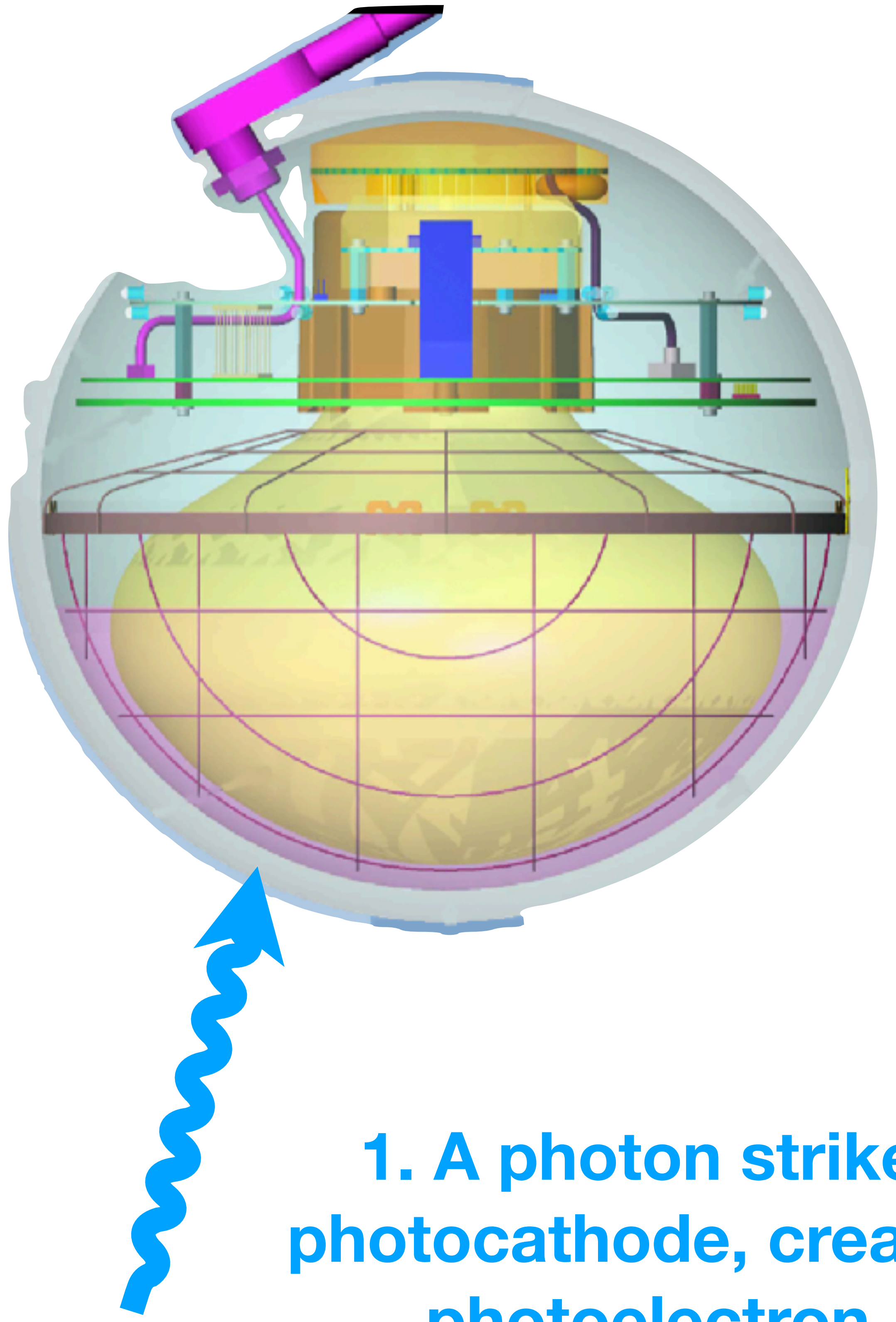
Each IceCube DOM consists of a Photomultiplier Tube (PMT) and electronics to digitize the PMT signal and sent it to the IceCube Lab.

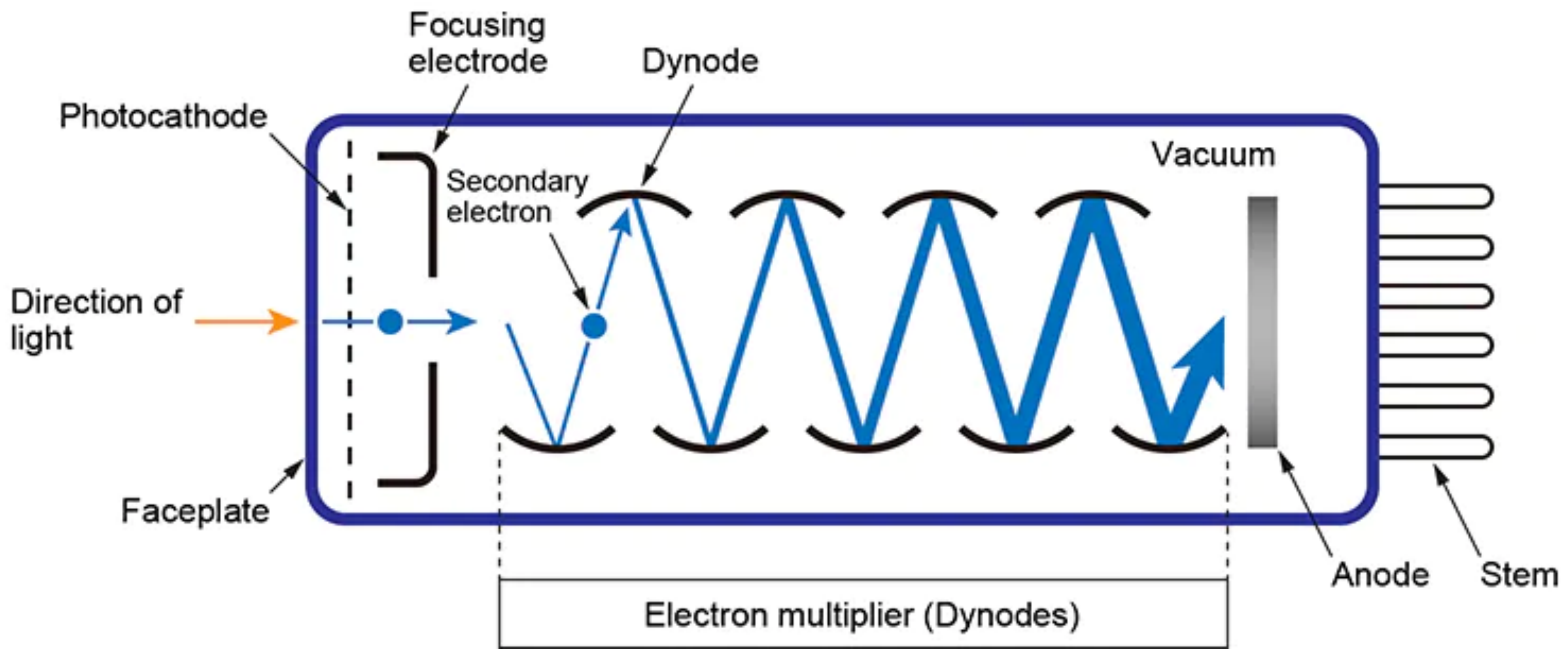
Cherenkov photons from secondary particles emitted when a neutrino interacts with the ice near the detector are captured by PMTs.

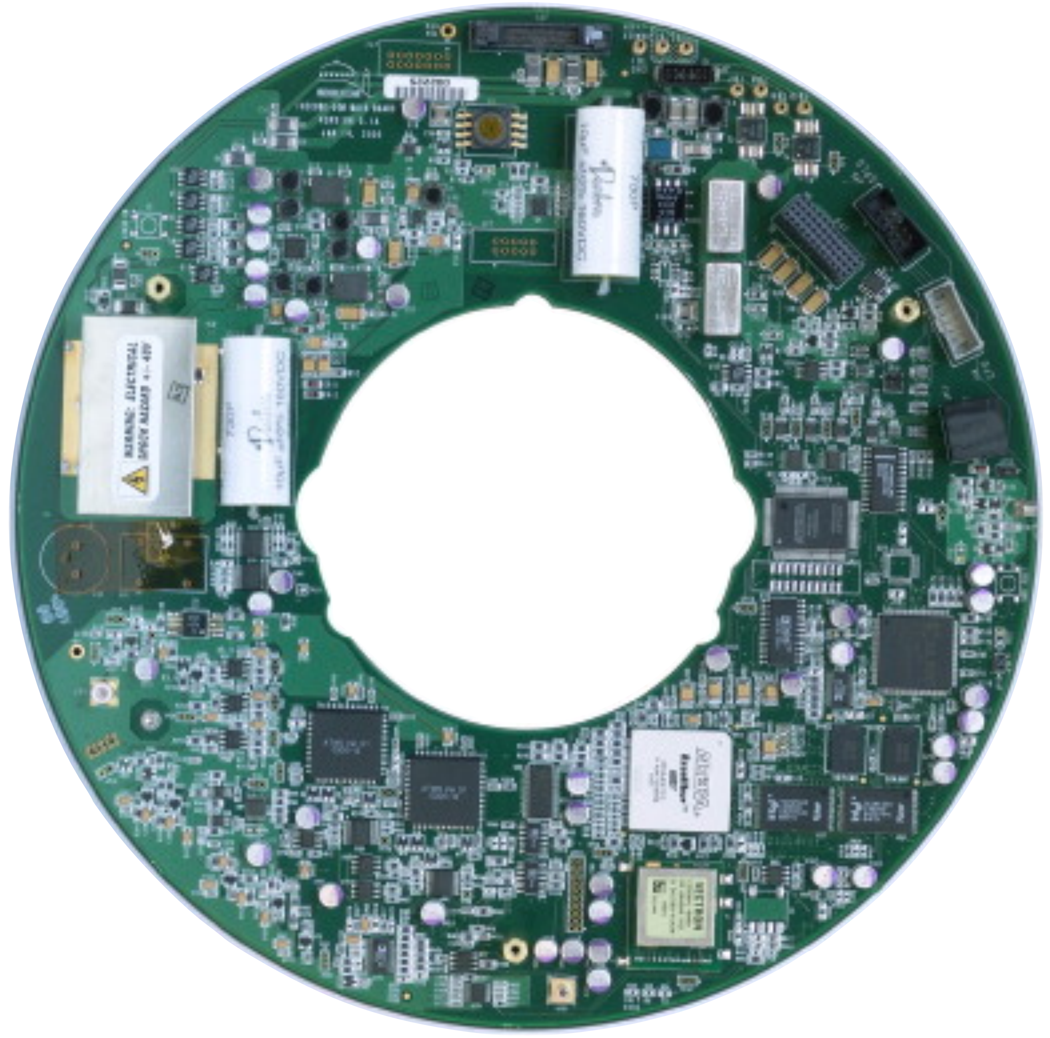




Photomultiplier Tubes: Working Principle







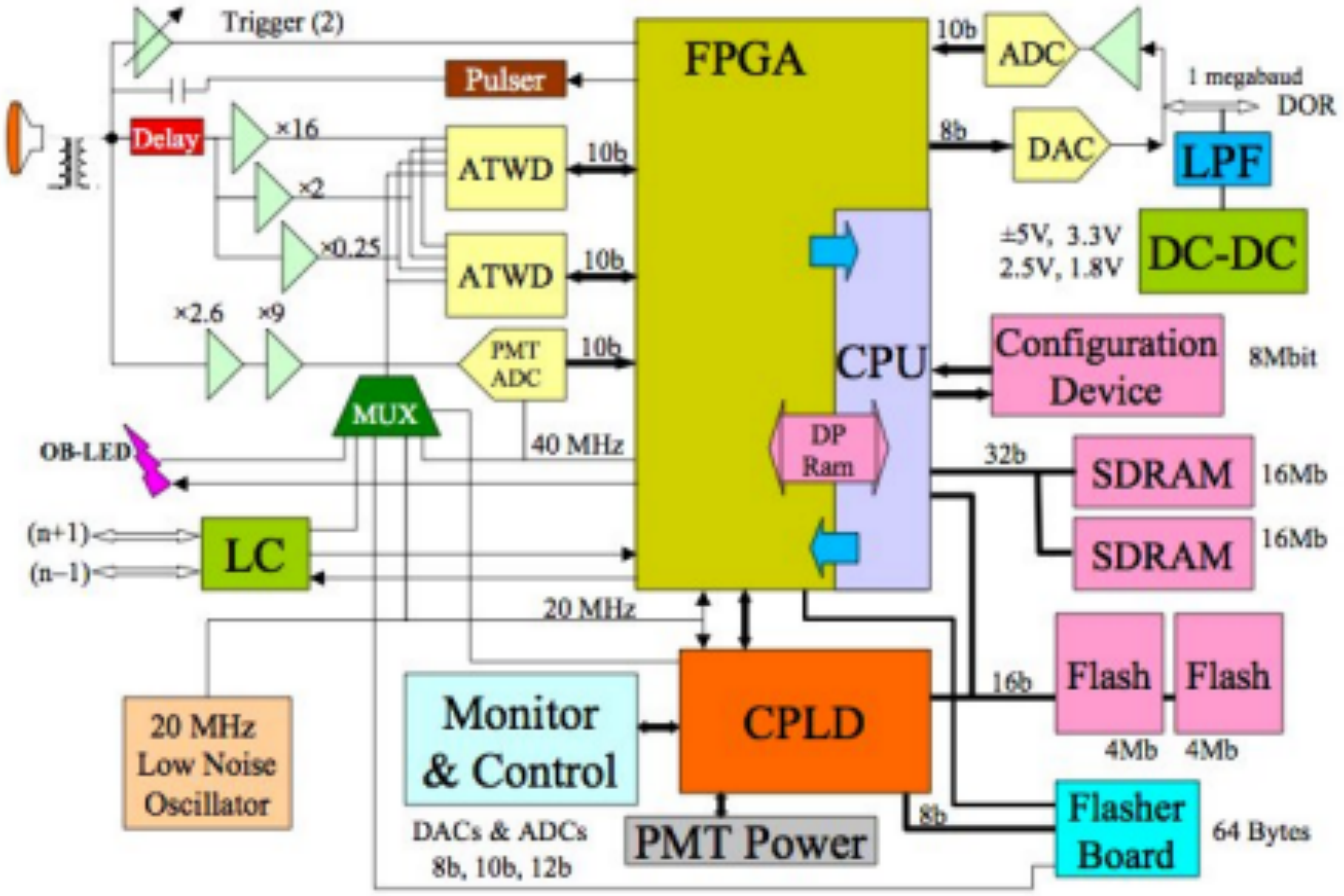
1. A photon strikes photocathode, creating photoelectron.

2. The photoelectron passes through dynode chain, creating more electrons. IceCube PMTs have a gain of 10^7.

3. The PMT charge signal is read out to the DOM electronics to be digitized and sent to the ICL.

DOM Electronics

The DOM mainboard is responsible for triggering data collection, digitizing PMT signals, and packaging information to be sent up to the IceCube Laboratory (ICL).



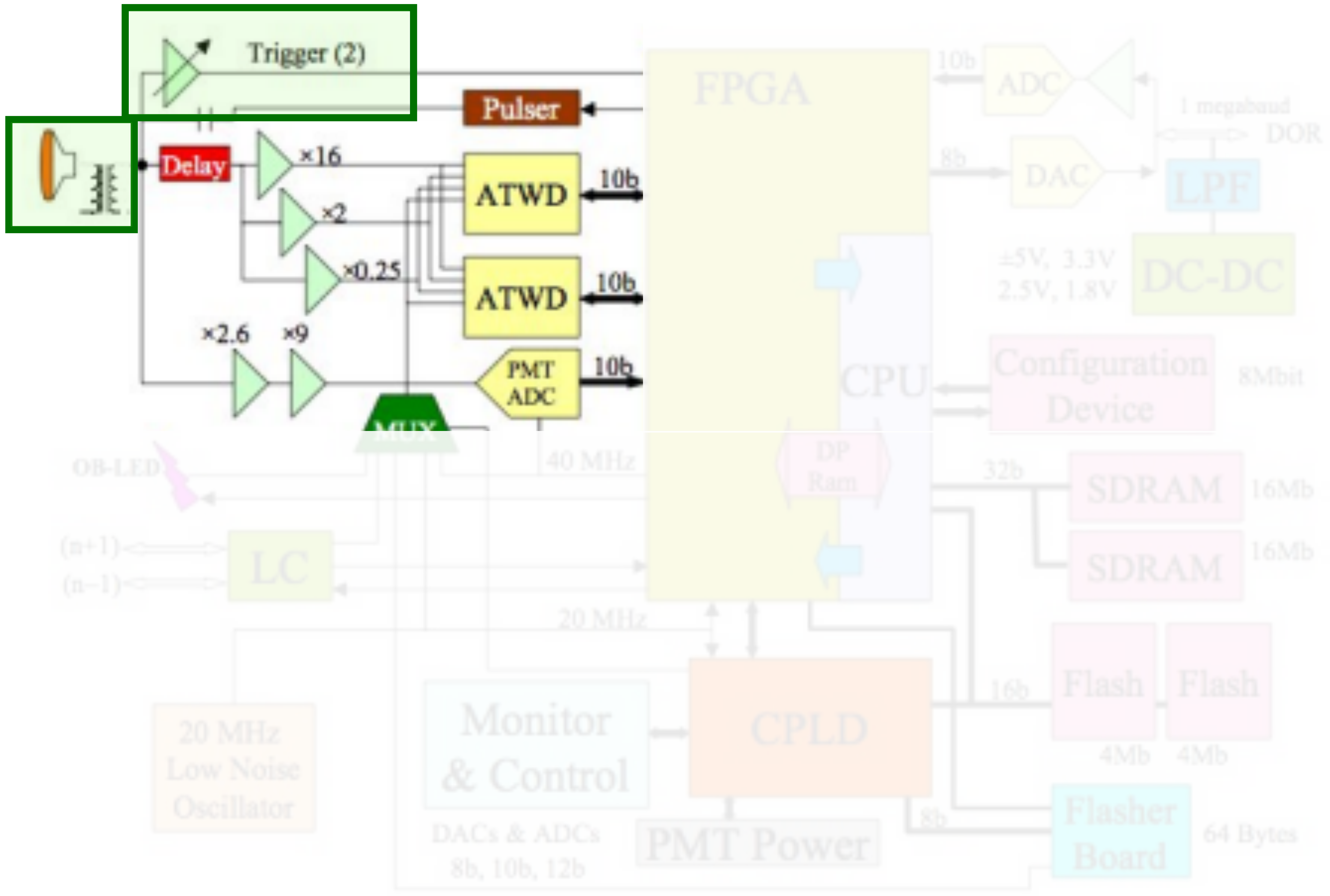
DOM Electronics: Trigger

The trigger system determines whether or not to collect PMT pulses based on their peak size.

Accepting pulses that are too small leads to electronic noise contamination.

The trigger is a comparator circuit with a threshold that can be set. Typically, IceCube uses a threshold of 0.2325 PE, but it can be set as low as 1/6 PE.

The actual PMT waveform is sent through a ~75ns delay board while the trigger decision is made.

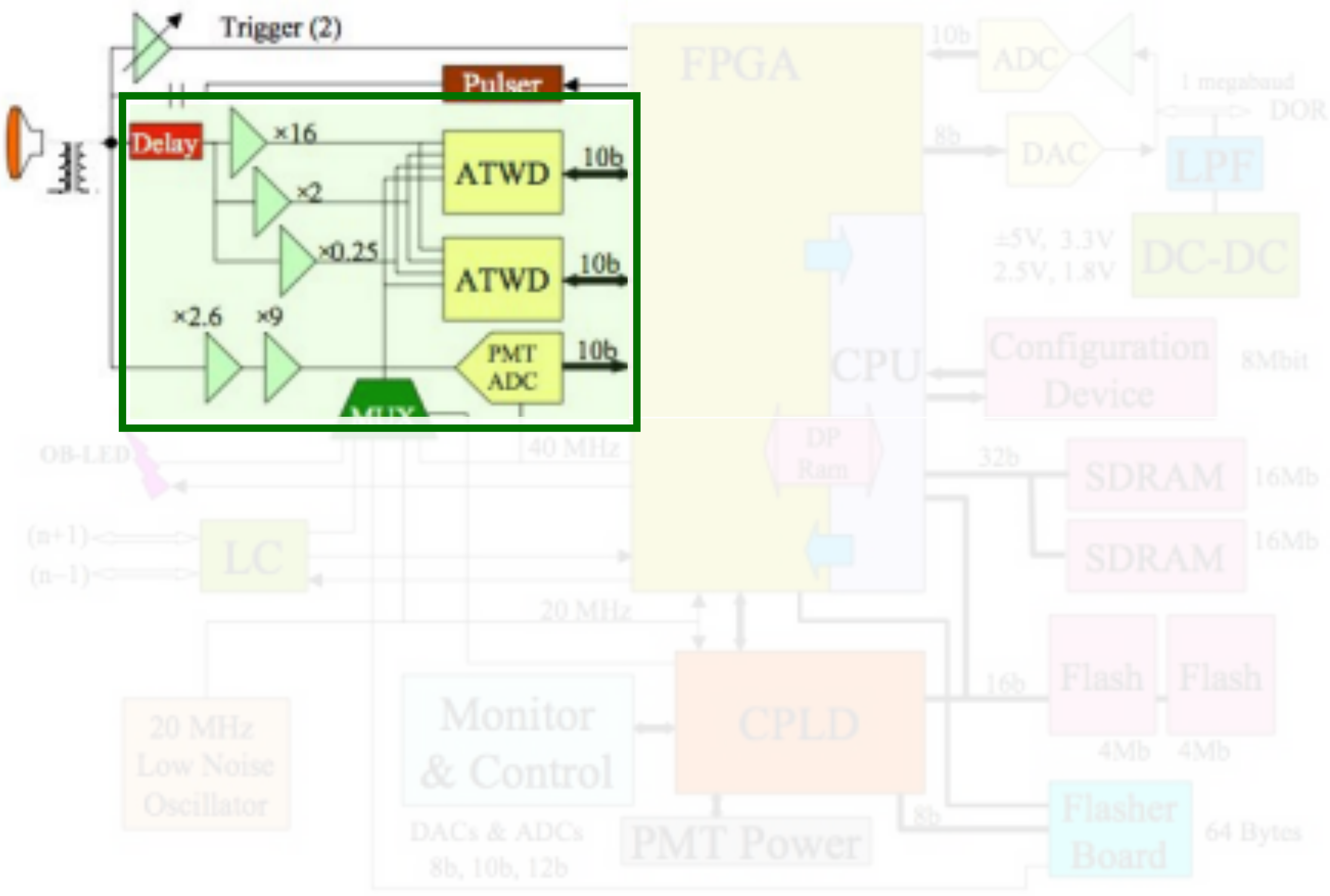


DOM Electronics: Digitization

If the data acquisition is triggered, the DOM takes the raw PMT signal, amplifies it, and converts it into a digitized signal which can be sent to the ICL.

DOMs have two digitizer types:

- **Analog Transient Waveform Digitizer (ATWD)**
- **Fast Analog to Digital Converter (fADC)**



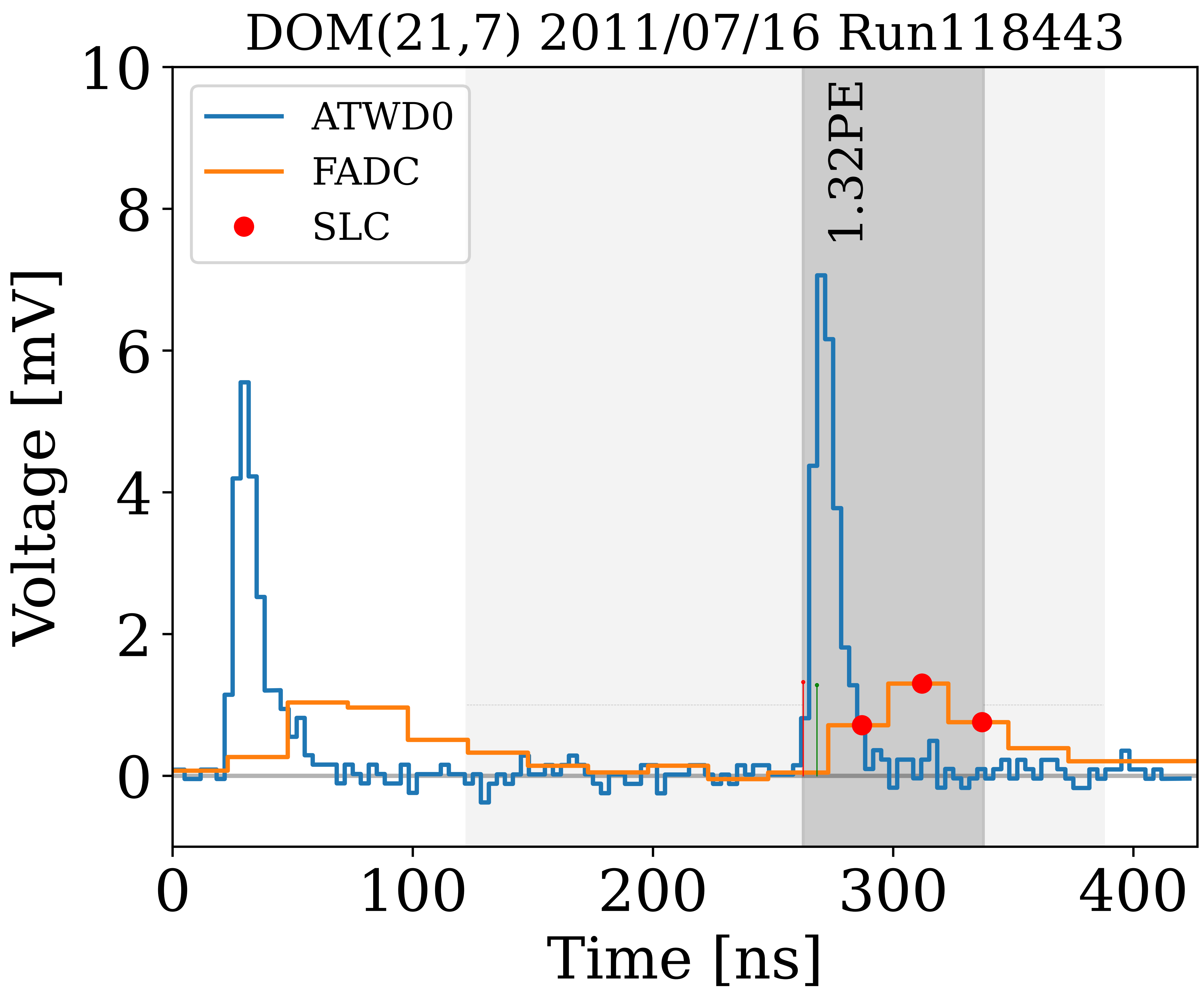
DOM Electronics: ATWD

The ATWD is a high-resolution ADC that gives good detail on pulses over a relatively short period of time.

- 128 sample depth
- 300 MSPS (427 ns of data collection)
- 10-bit ADC

ATWD pulses are amplified by 0.25x, 2x, and 16x. The highest non-saturated amplification channel is digitized.

The ATWD takes 29 µs to digitize a waveform, so there are 2 ATWDs per DOM to reduce deadtime.



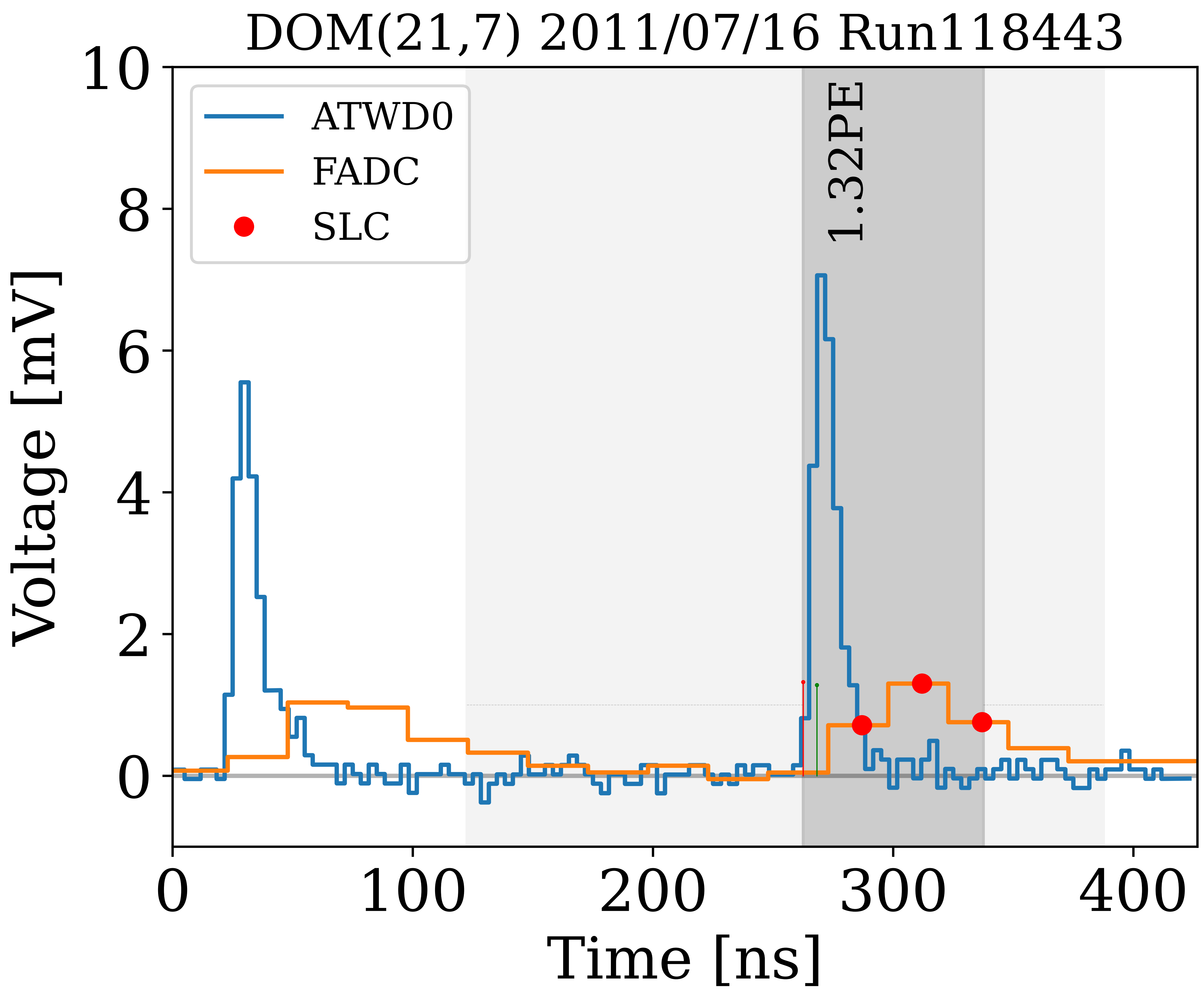
DOM Electronics: fADC

The fADC is a lower-resolution digitizer that takes data over a longer period of time than the ATWD.

- 256 sample depth
- 40 MSPS (6.4 μs of data collection)
- 10-bit ADC

The fADC is amplified by ~23x

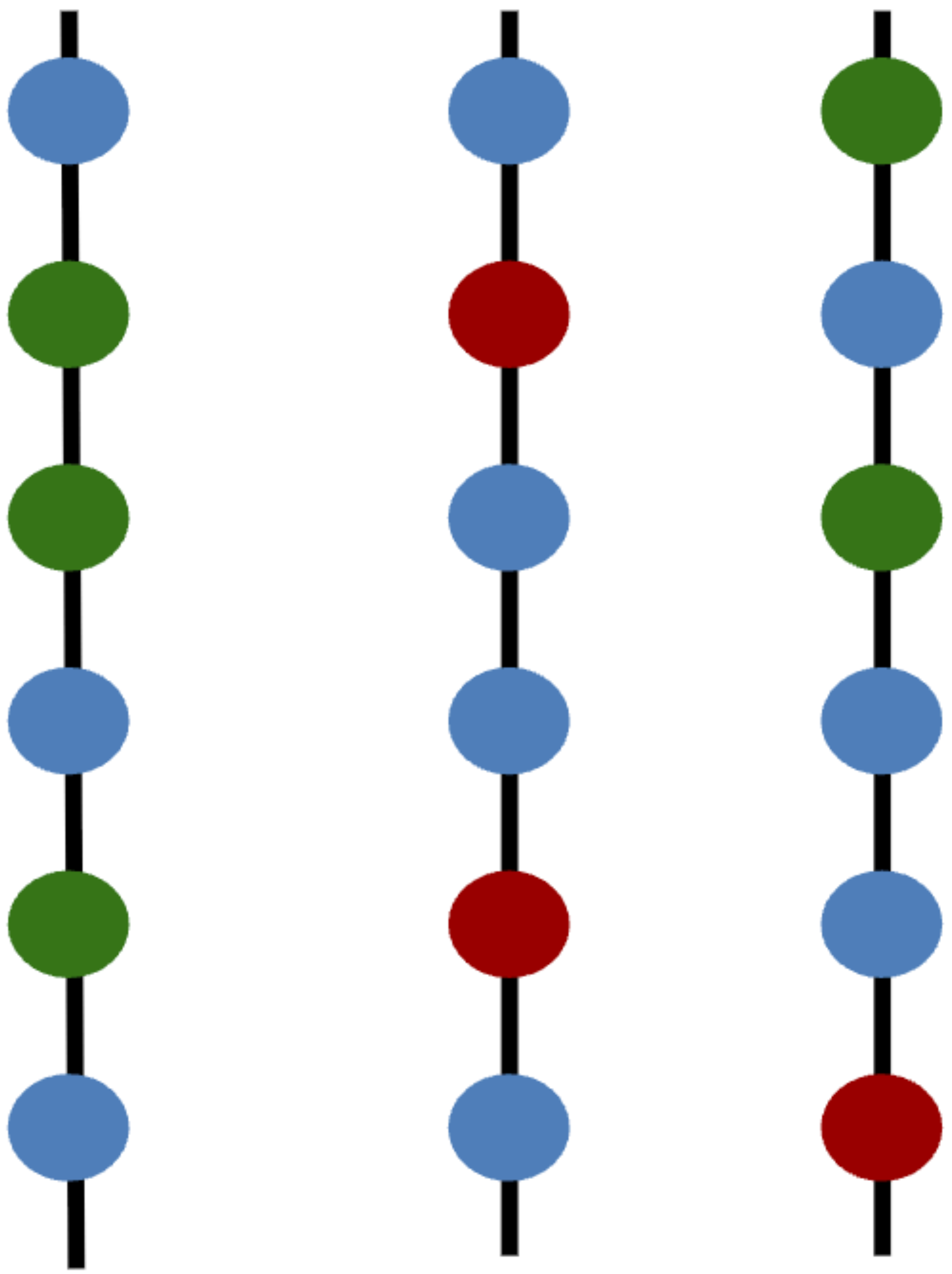
The fADC continuously samples, but the 256 samples after a trigger are saved.



DOM Electronics: Local Coincidence

DOMs in each string can communicate between each other to determine if multiple DOMs saw hits in a small time window, which would suggest a more interesting event. There are two types of coincidence flags:

- **Hard Local Coincidence (HLC):** A DOM and one of its neighbors or next-to-nearest neighbors on a string are both hit within 1 μs
- **Soft Local Coincidence (SLC):** A DOM is hit, but the HLC flag conditions are not met



IceCube Strings. Blue DOMs did **not** record a hit, red and green DOMs did (in a 1 μs period). Green DOMs recorded **HLC** hits, red DOMs recorded **SLC** hits.

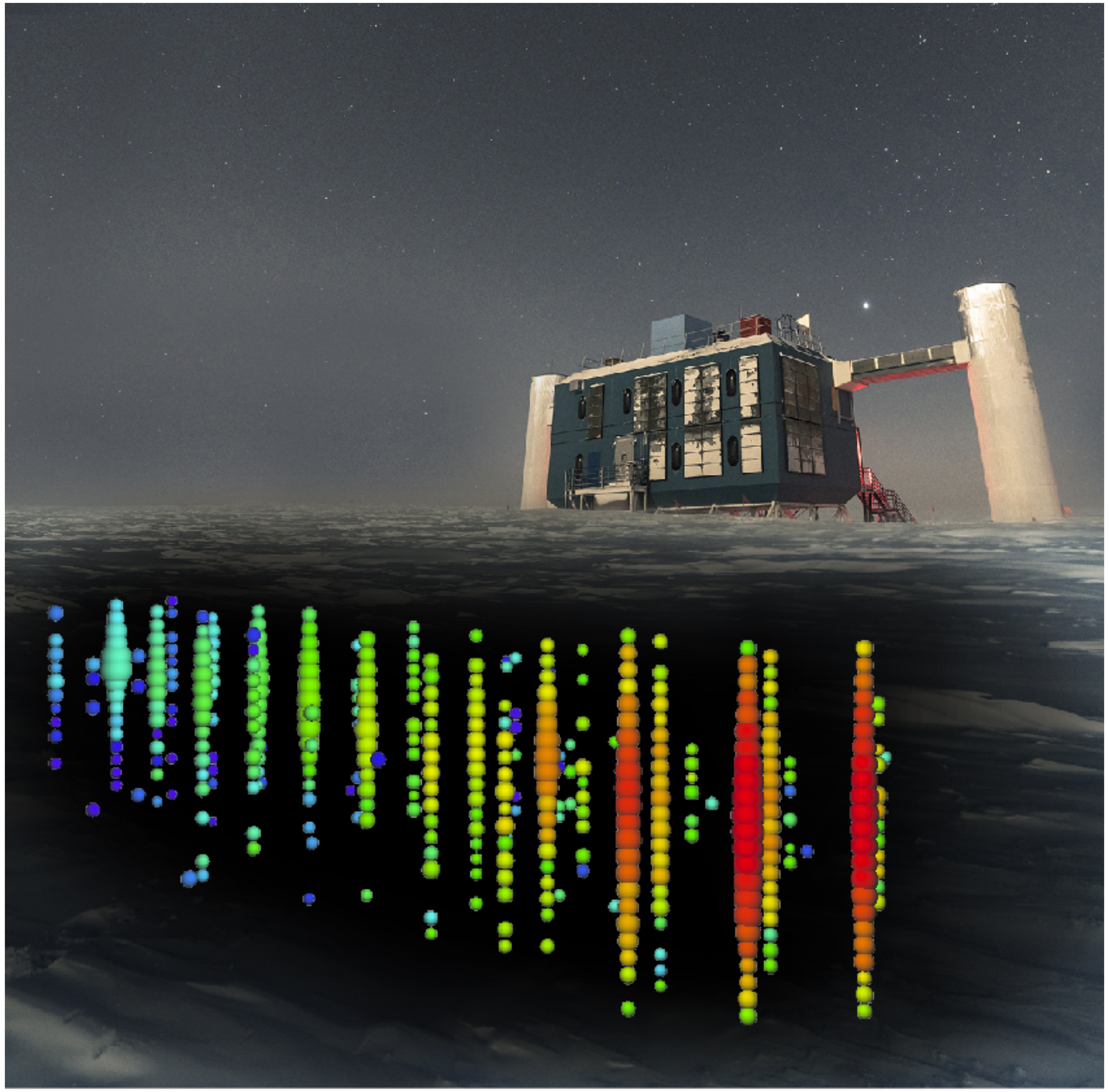
Compression and Transmission to ICL

Based on the LC flags associated with a waveform, data is sent to the ICL for further processing.

For every waveform, a header is sent to the ICL, containing the timestamp, coarse charge information (3 fADC charge values centered at the peak), trigger information, and hit information.

For **HLC**, the full ATWD and fADC waveforms are sent to the ICL; for **SLC**, just the header is sent.

The waveform information is compressed by ~3.8x losslessly by saving the sample-to-sample differences in ADC counts.

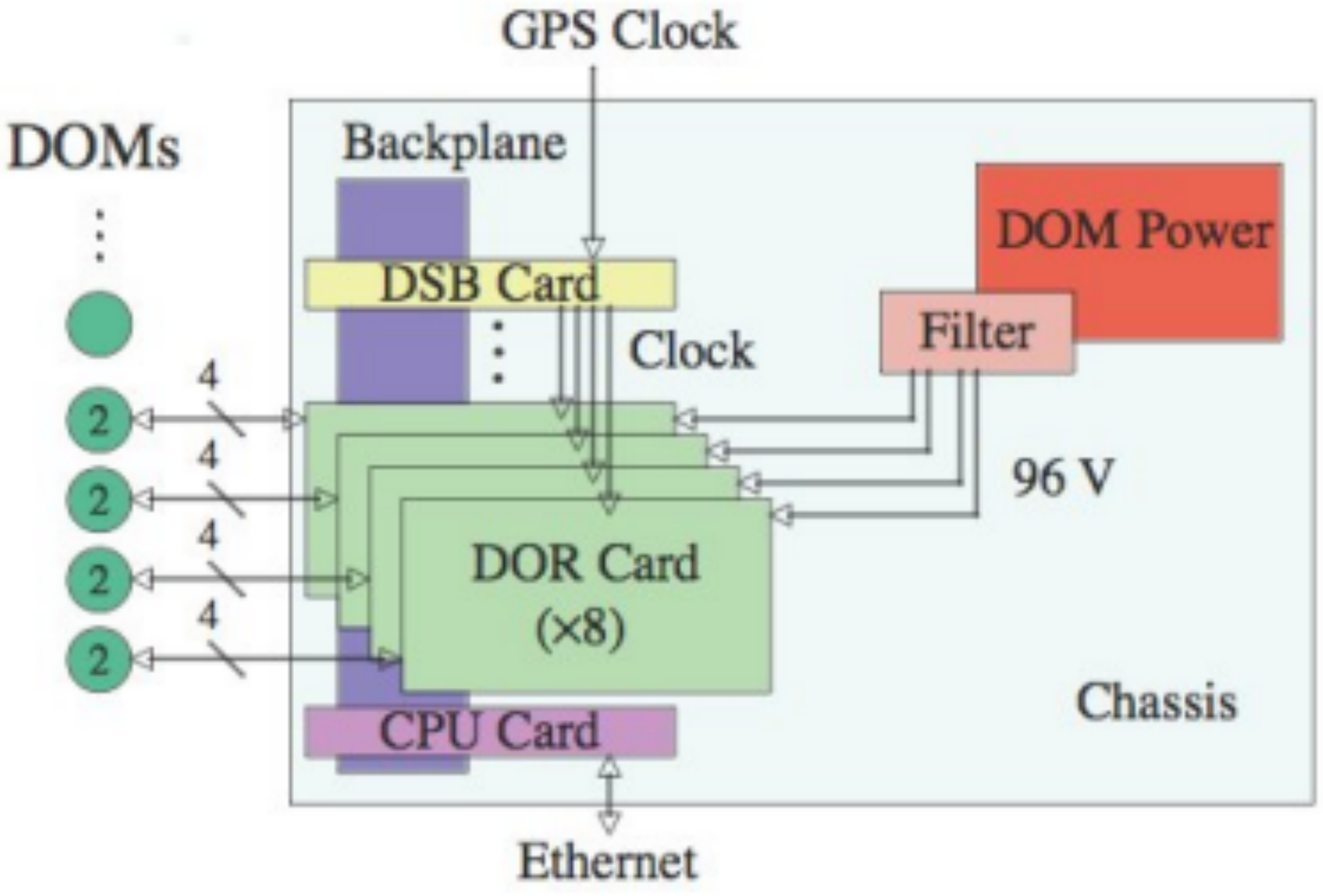


ICL Computer Systems

At the ICL, each string has a computer, known as a DOMHub, tasked with collecting the data from all 60 DOMs on the string.

Each DOMHub has 8 PCI DOM Readout (DOR) cards to collect the data, with each DOR card collecting from up to 8 DOMs.

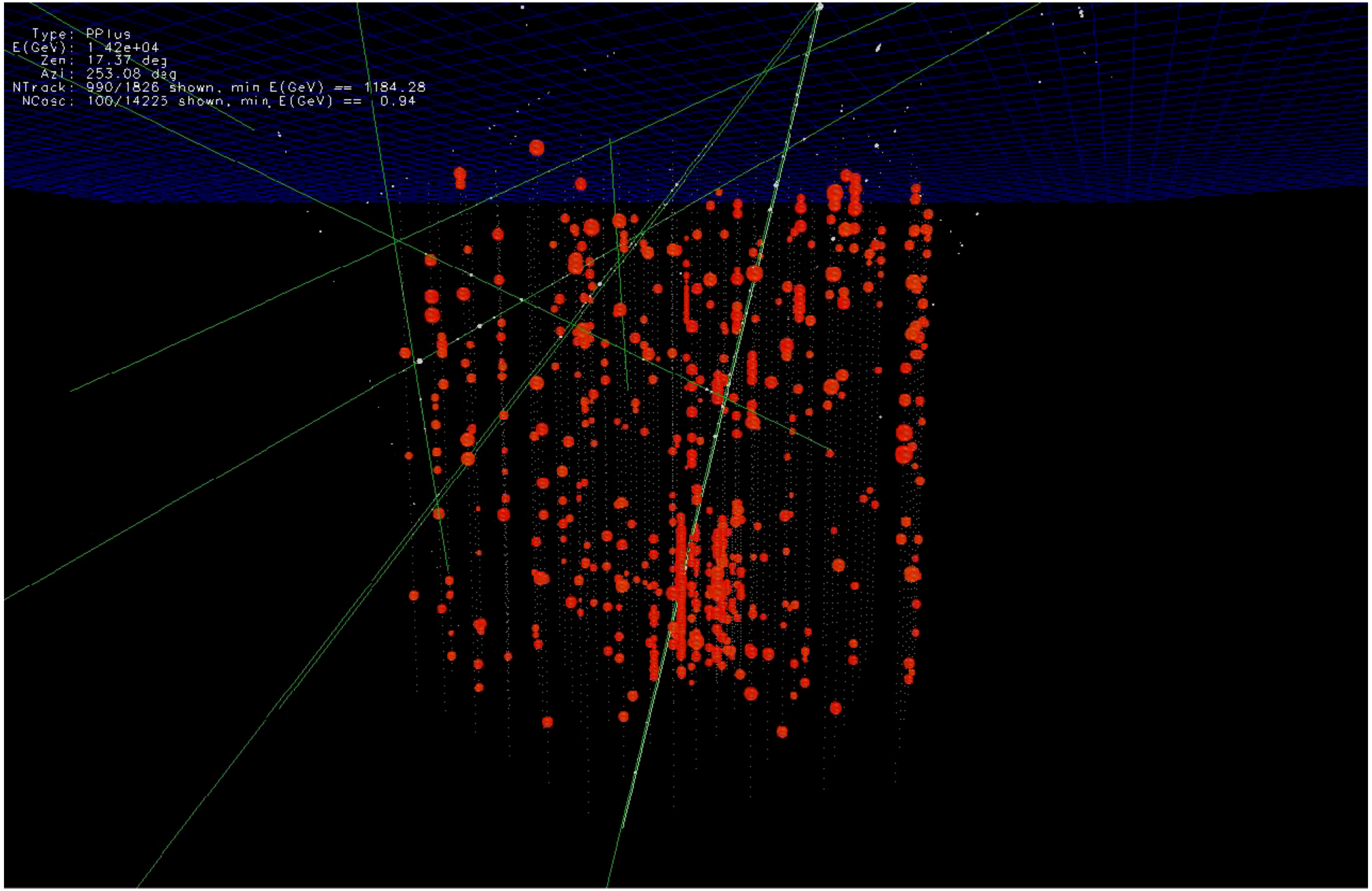
The read-out data from each DOMHub can be used to look at detector-wide information to create events.



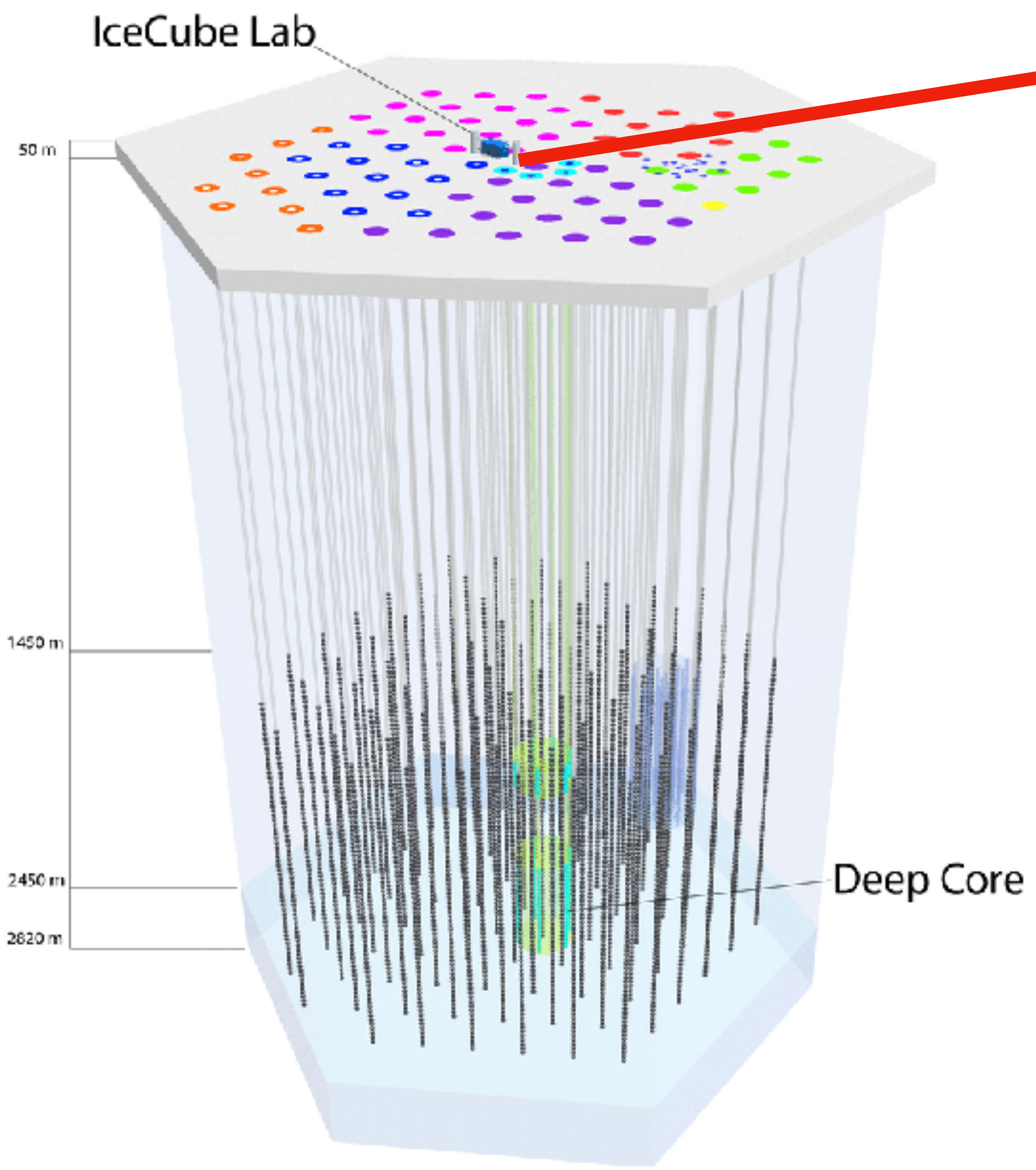
IceCube Has a Data Collection Challenge

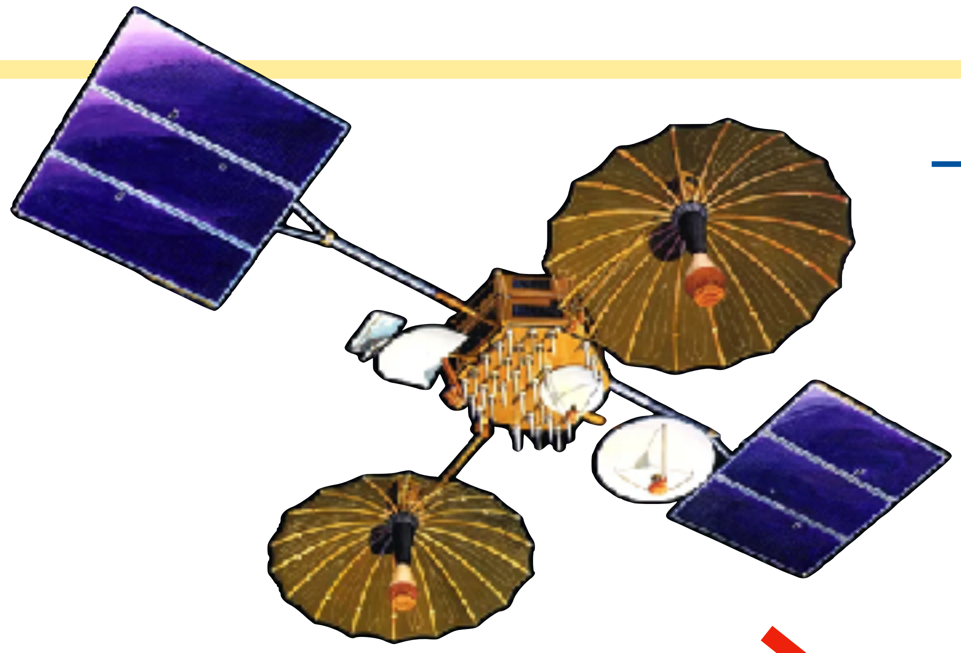
IceCube consists of 5160 DOMs which can detect single photons. This means IceCube collects an enormous amount of data!

This movie shows 10ms of IceCube data collection:



IceCube Has a Data Collection Challenge





TRDSS Satellite

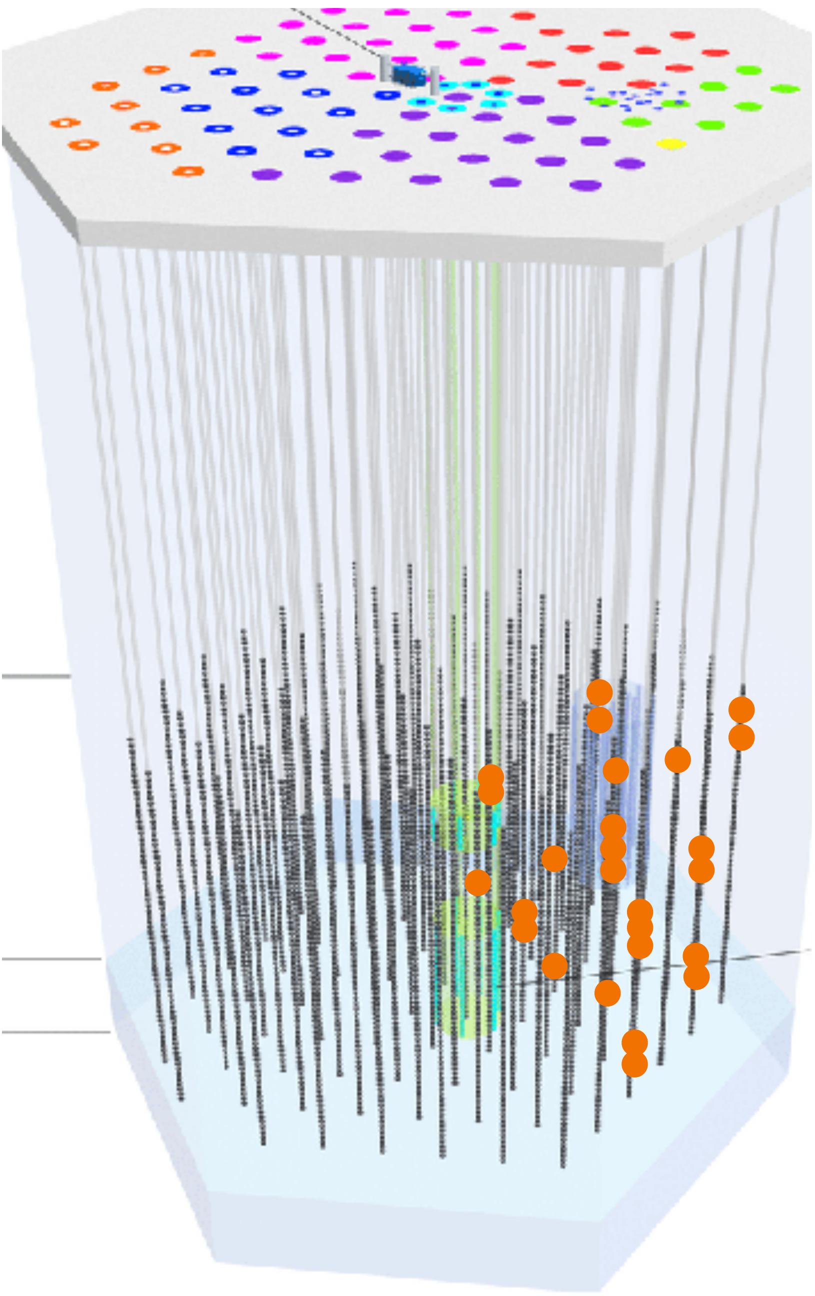
~100 GB/day

(Raw DAQ Output:
1 TB/day!)

We have to intentionally process data to maximize how much can quickly be sent North.

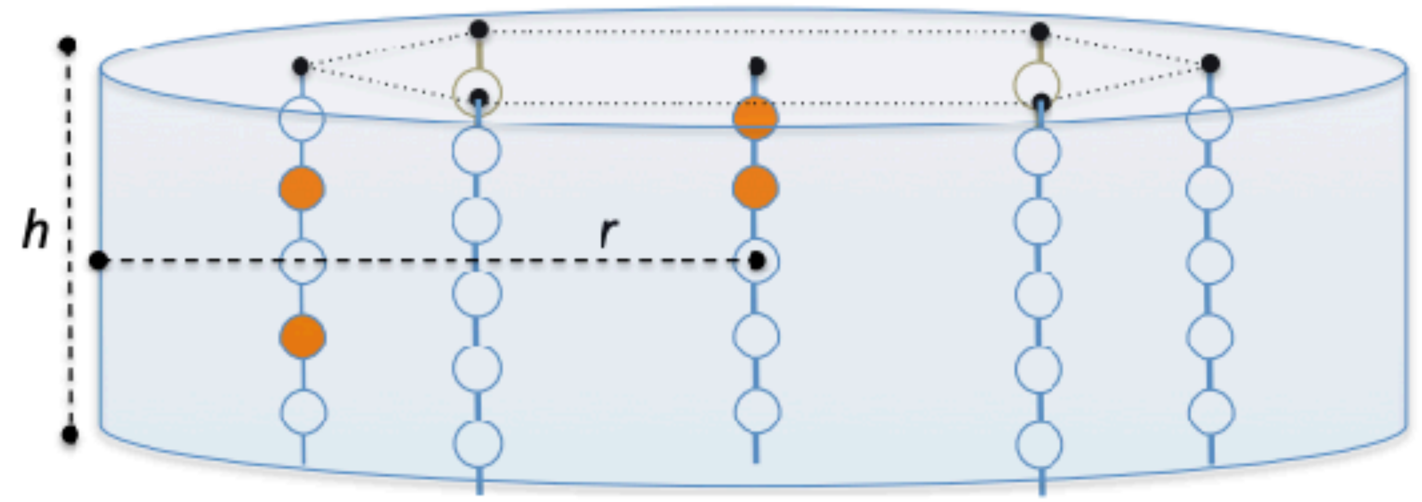
Data Processing at the ICL: Events

Waveforms from many DOMs are grouped into **events** based on their relative timing and location in the detector to separate interactions in the detector from noise. **Triggers** determine what conditions must be met to group events. Separate triggers can be used for separate detector regions, such as DeepCore or IceTop.



Simple Multiplicity Trigger (SMT)

String Trigger



Volume Trigger

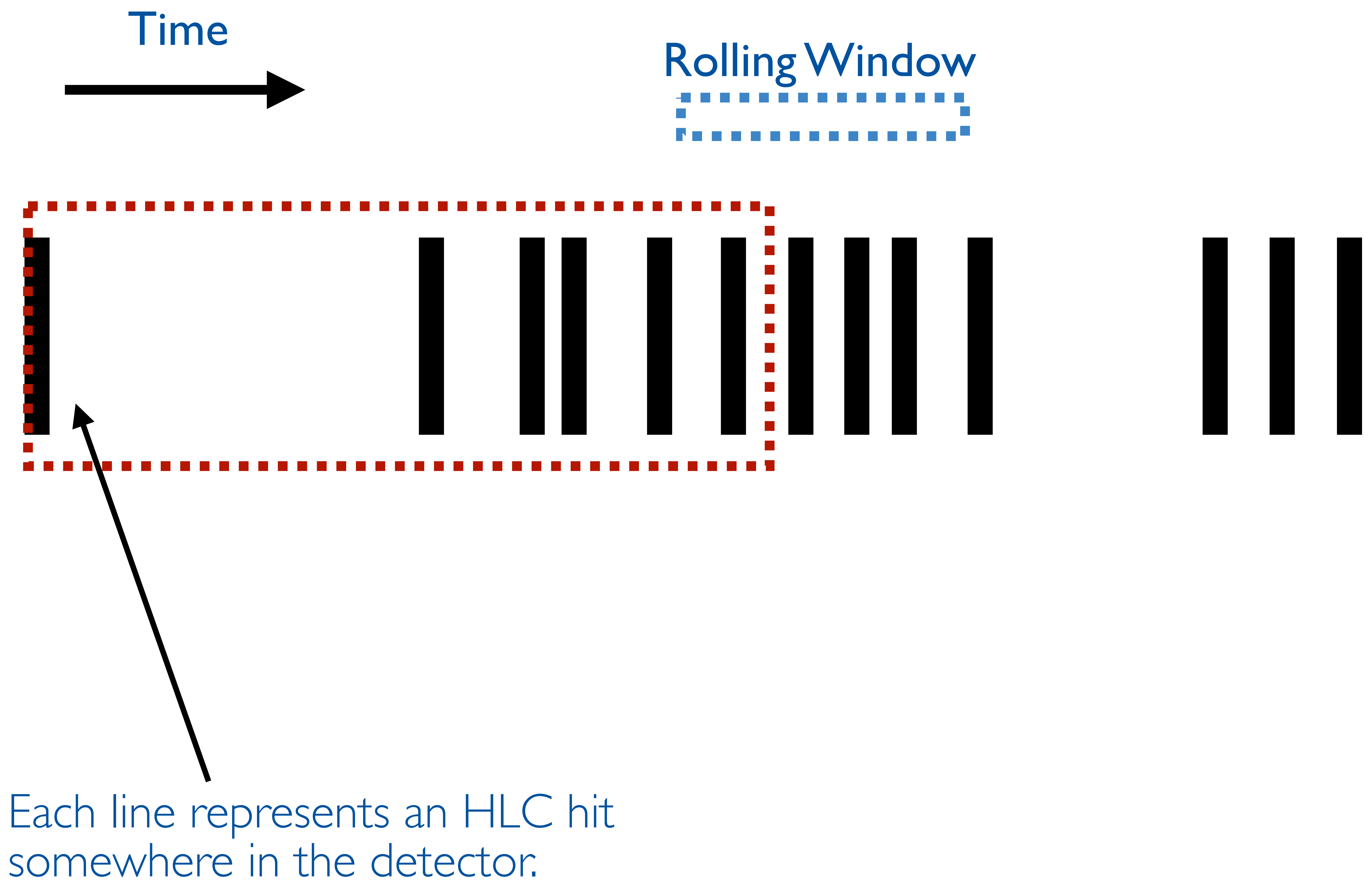
Simple Multiplicity Trigger

- *N* HLC hits in a rolling time window over the entire detector

- As long as *N* or more HLCs are present in the time window, the trigger is extended

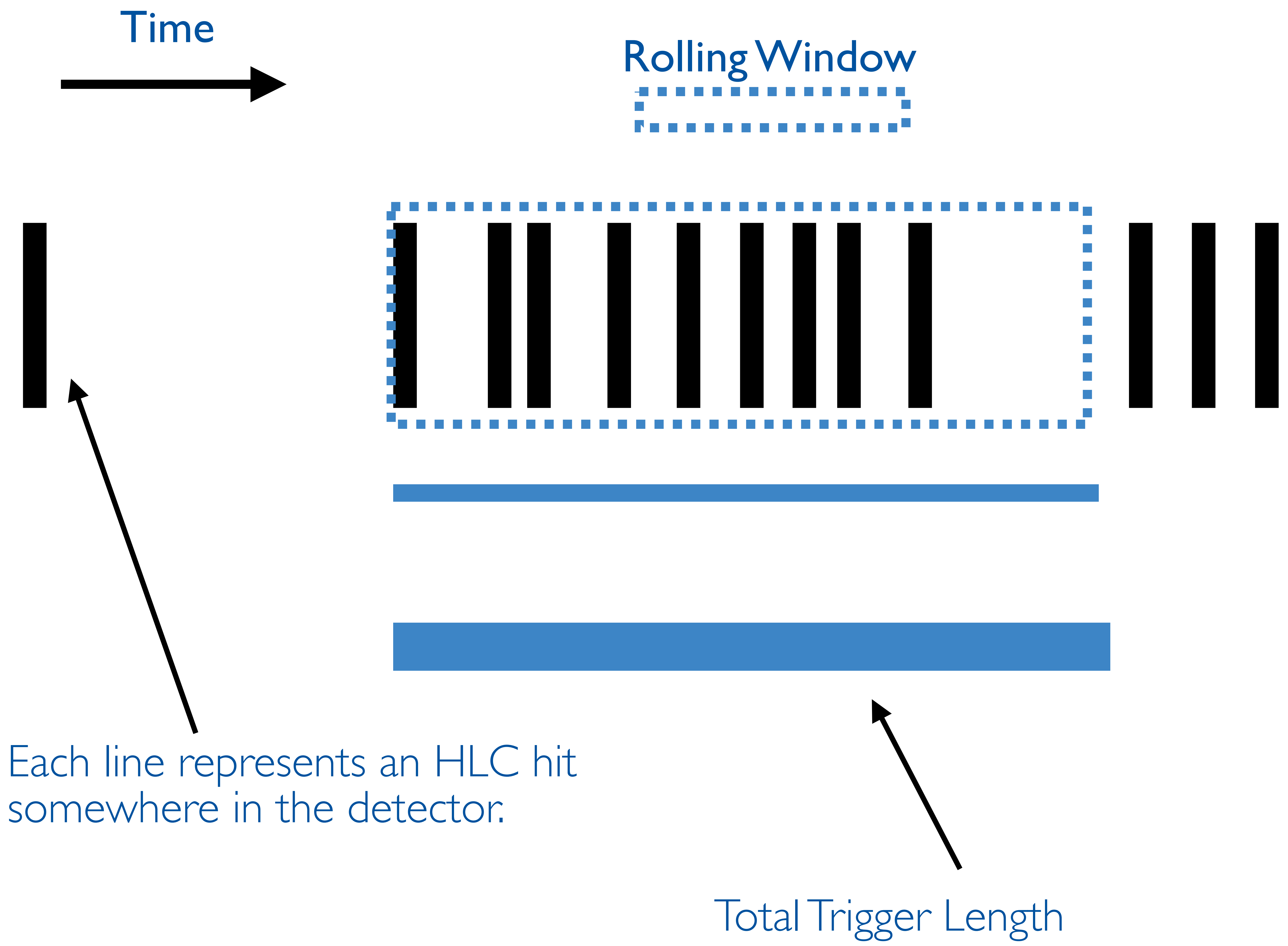
- Readout window is extended to before and after trigger (-4, +6 μs) to capture early and late pulses

- Commonly used In-Ice: SMT-8 (8 HLC hits in 5 μs), 2100 Hz rate



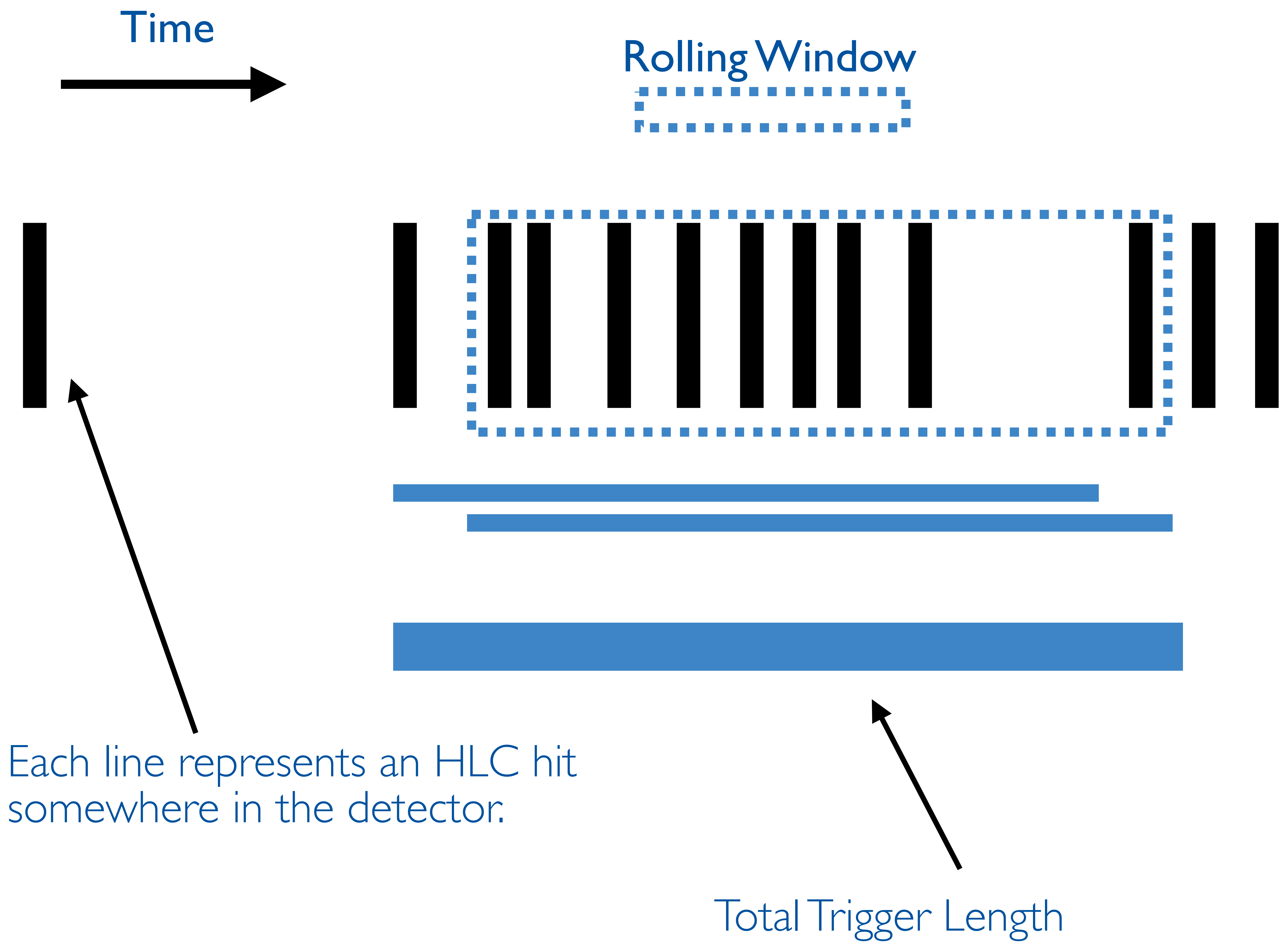
Simple Multiplicity Trigger

- *N* HLC hits in a rolling time window over the entire detector
- As long as *N* or more HLCs are present in the time window, the trigger is extended
- Readout window is extended to before and after trigger (-4, +6 μs) to capture early and late pulses
- Commonly used In-Ice: SMT-8 (8 HLC hits in 5 μs), 2100 Hz rate



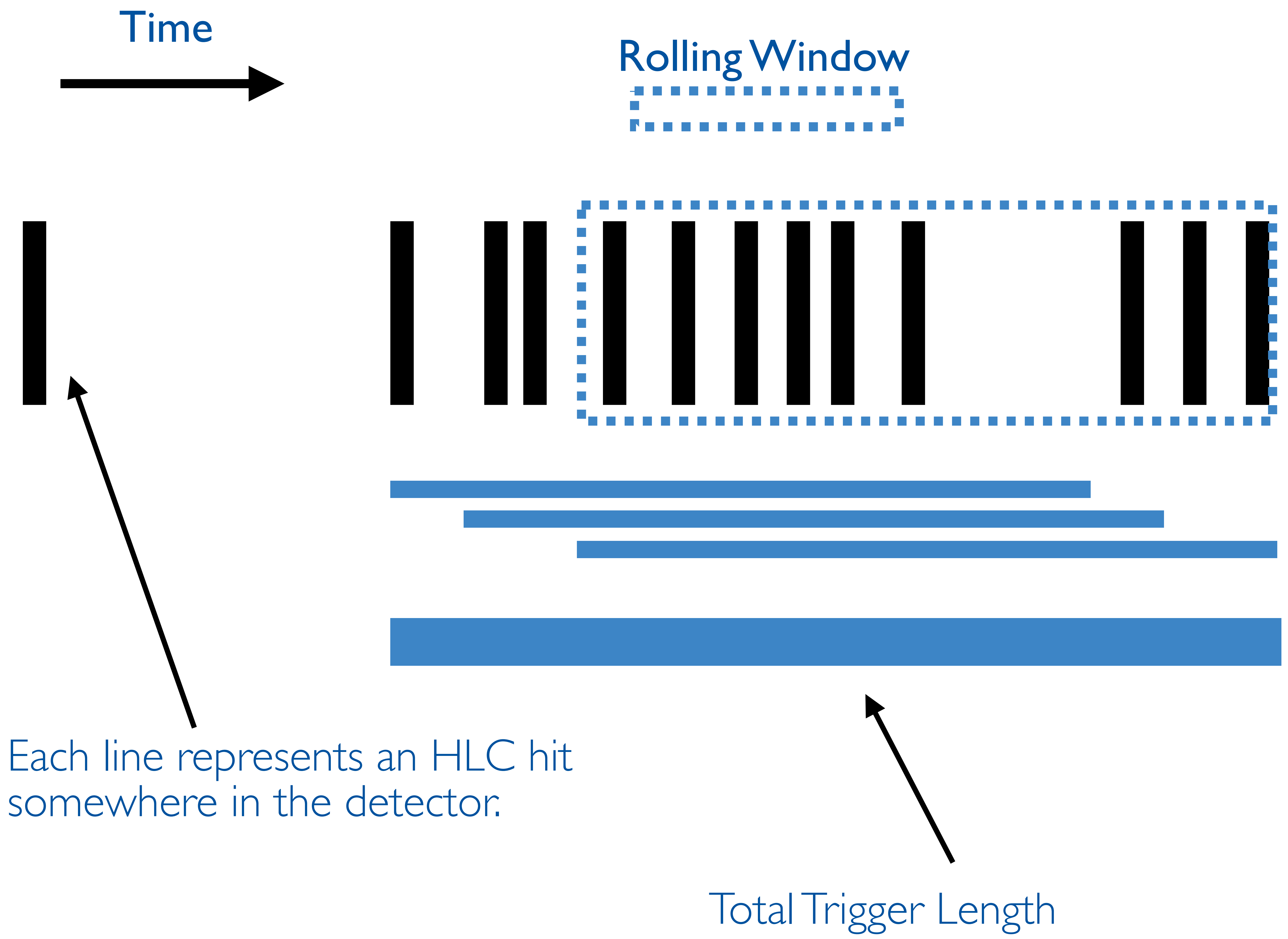
Simple Multiplicity Trigger

- *N* HLC hits in a rolling time window over the entire detector
- As long as *N* or more HLCs are present in the time window, the trigger is extended
- Readout window is extended to before and after trigger (-4, +6 μs) to capture early and late pulses
- Commonly used In-Ice: SMT-8 (8 HLC hits in 5 μs), 2100 Hz rate



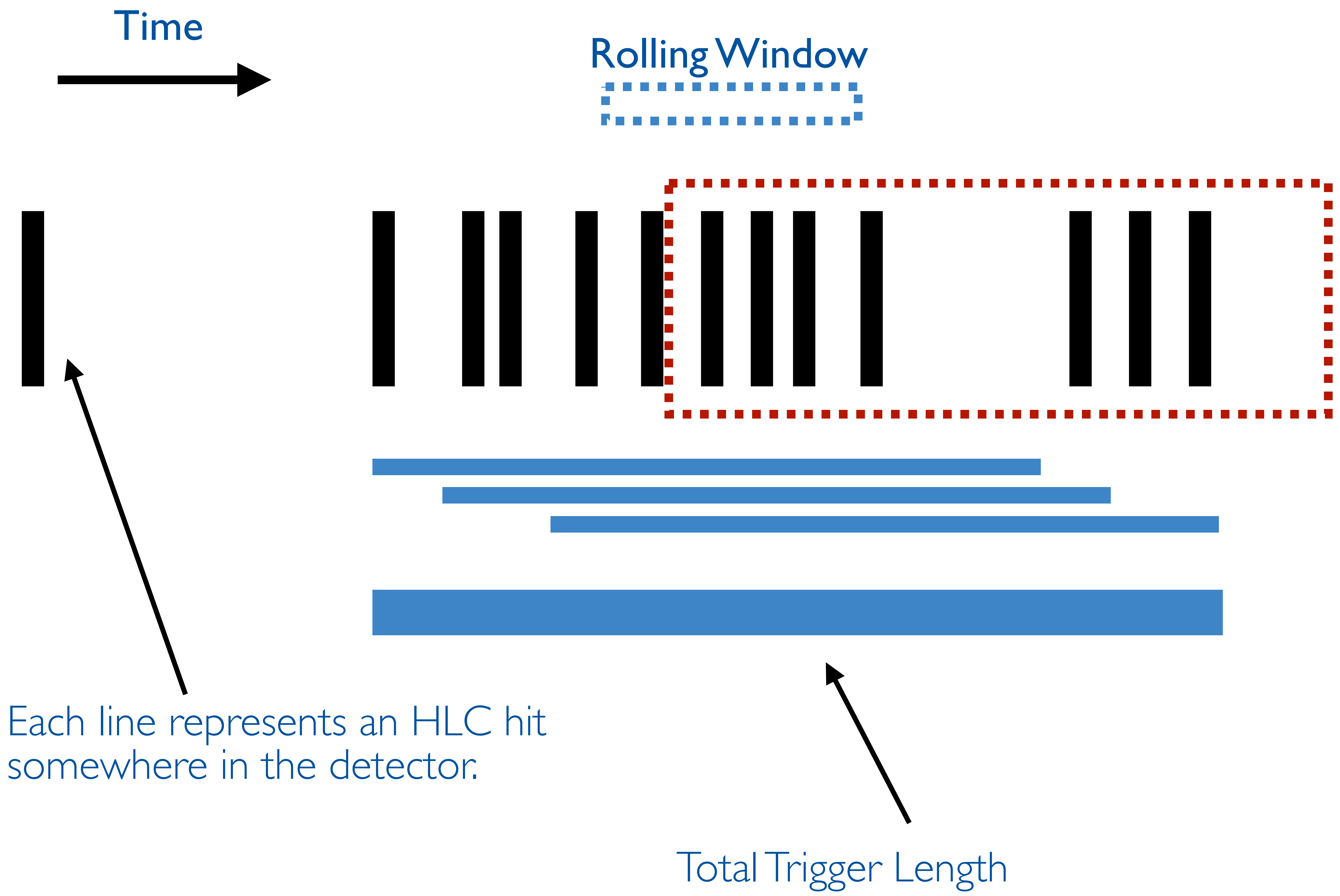
Simple Multiplicity Trigger

- *N* HLC hits in a rolling time window over the entire detector
- As long as *N* or more HLCs are present in the time window, the trigger is extended
- Readout window is extended to before and after trigger (-4, +6 μs) to capture early and late pulses
- Commonly used In-Ice: SMT-8 (8 HLC hits in 5 μs), 2100 Hz rate



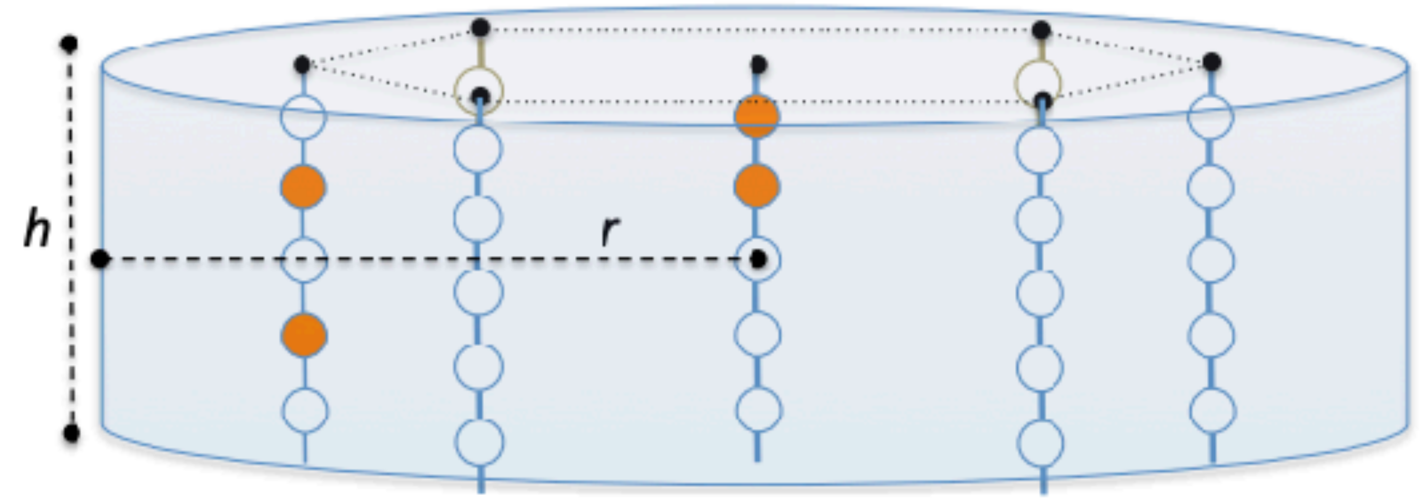
Simple Multiplicity Trigger

- *N* HLC hits in a rolling time window over the entire detector
- As long as *N* or more HLCs are present in the time window, the trigger is extended
- Readout window is extended to before and after trigger (-4, +6 μs) to capture early and late pulses
- Commonly used In-Ice: SMT-8 (8 HLC hits in 5 μs), 2100 Hz rate



Topological Triggers

- *N* HLC hits in some defined region of the detector surrounding any DOM in a given time window.



String Trigger: *N* HLC hits in any *X* adjacent DOMs on a string.

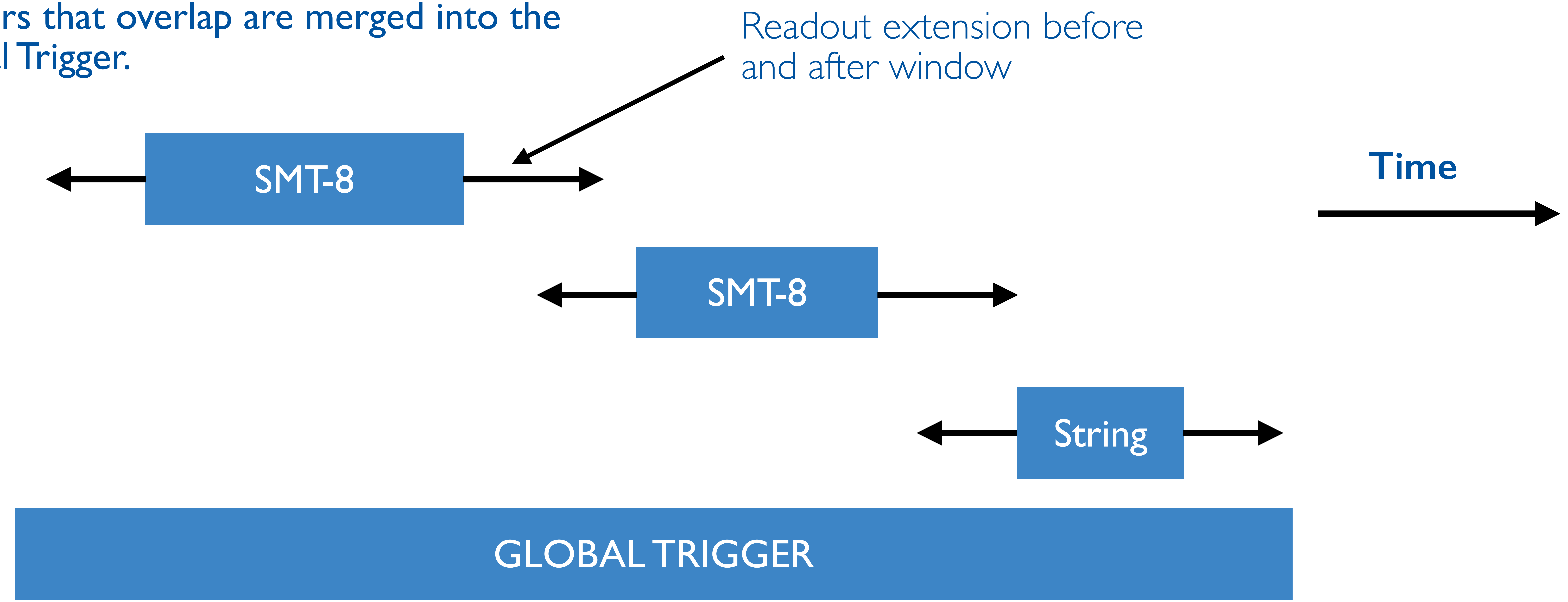
Typically used In-Ice: 5 HLC hits from 7 adjacent DOMs in 1.5 μs; 2200 Hz

Volume Trigger: *N* HLC hits in a cylinder of radius *r* and height *h* surrounding any DOM

Typically used In-Ice: 4 HLC hits in cylinder of *r* =175m, *h* = 75m in 1.5 μs; 3700 Hz

Global Trigger

Triggers that overlap are merged into the Global Trigger.



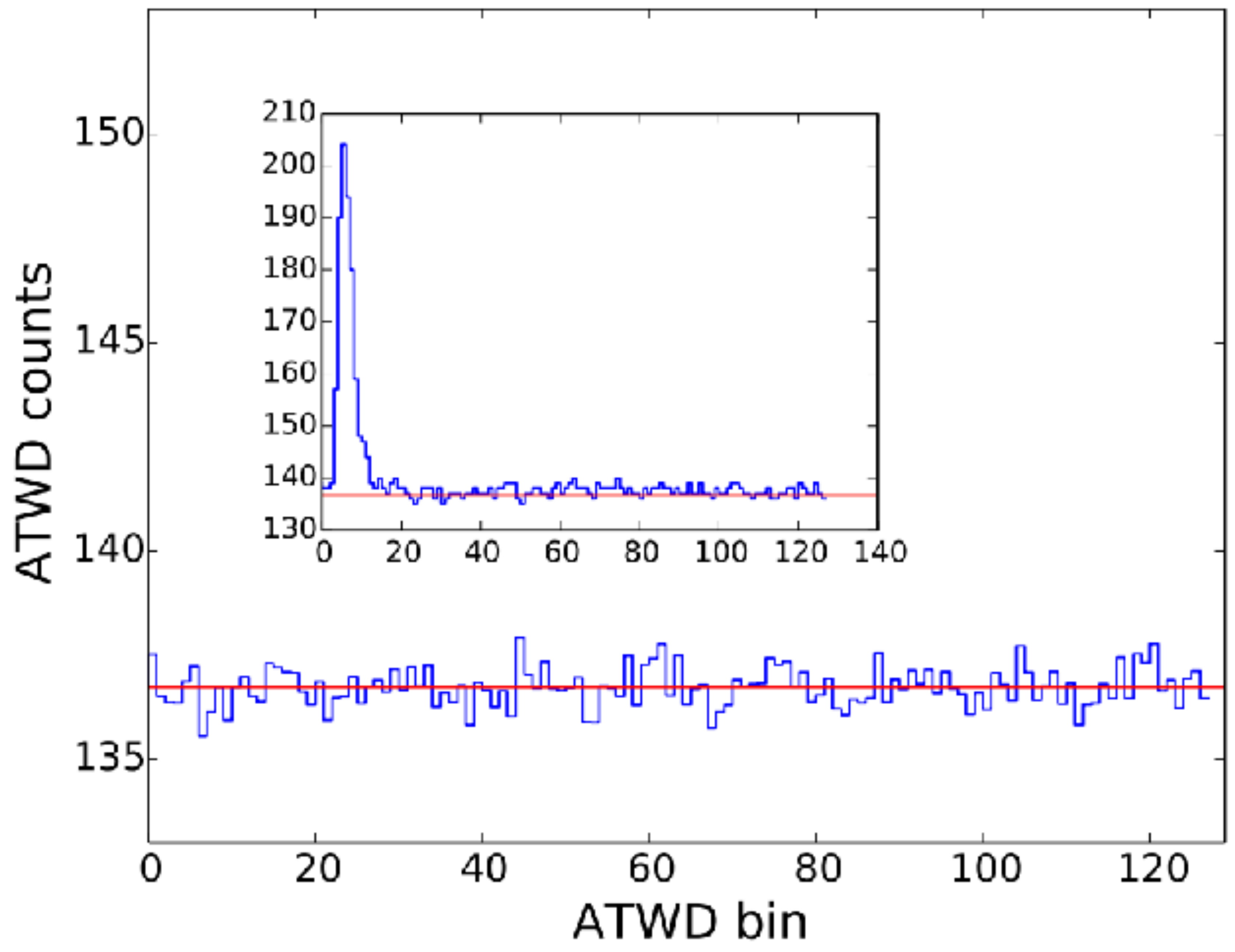
Information about the individual triggers are maintained, but repeated DOM hits from multiple triggers are avoided.

Waveform Calibration

At the Pole, simple steps are taken to extract useful information from the waveforms in events.

Baseline & Pedestal Removal: 200 forced-trigger waveforms are taken at the start of each run to determine the pedestal pattern for each digitizer in each DOM.

Gain Calculation: In each DOM, DOMCal uses darknoise photons to calibrate HV setting to tune the DOM gain to 10^7.



Red line: baseline (set value)

Blue line: pedestal pattern averaged over 1000 waveforms

Inset: typical ATWD SPE waveform

Waveform Calibration

With known pedestal pattern, baseline, and gain, waveforms can be converted into voltages:

$$V = (\text{counts} - \text{baseline}) \times \text{gain}$$

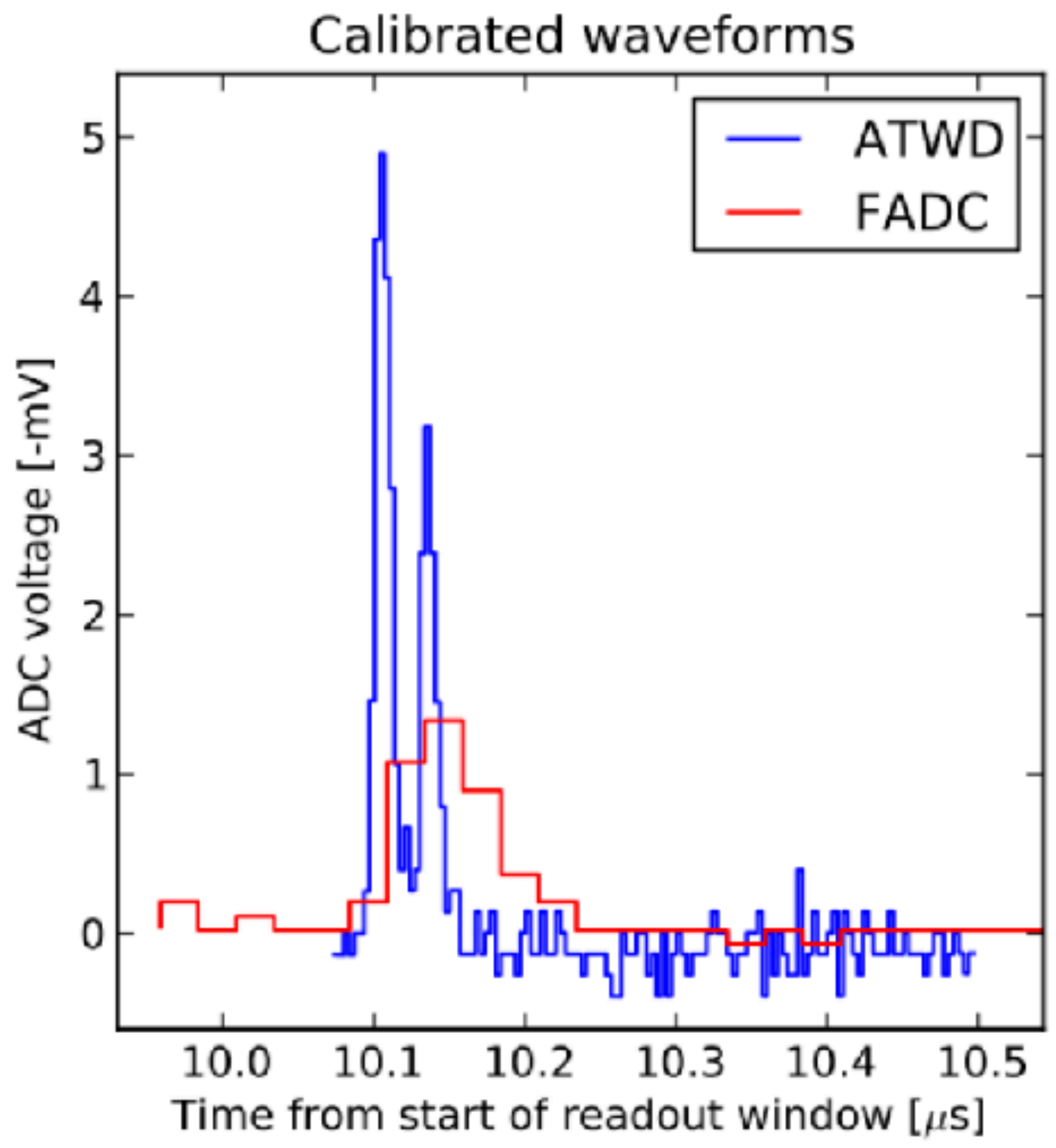
The hit time of ATWD waveforms are calibrated by accounting for the travel time through the PMT and delay board:

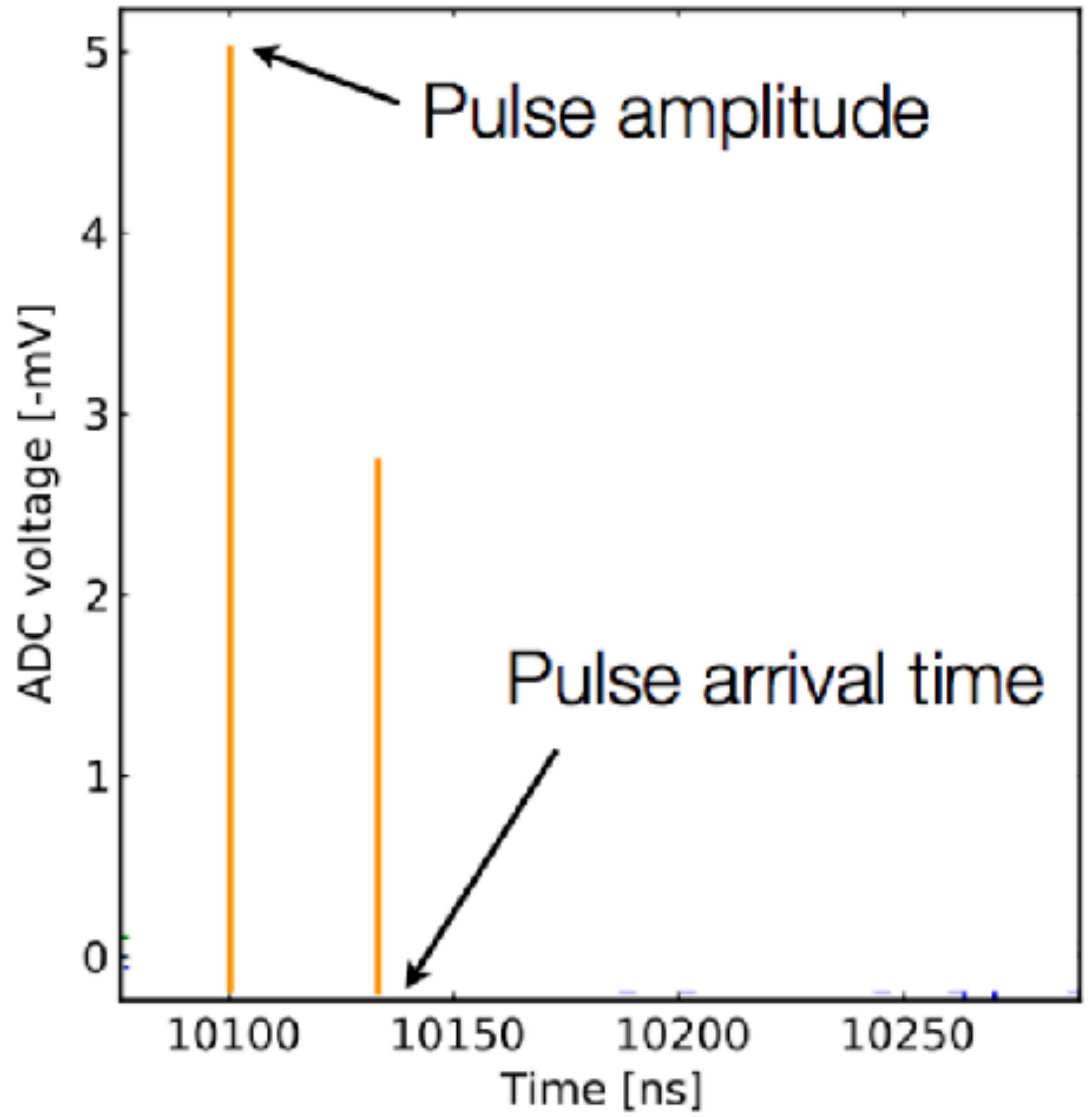
$$t_{\text{transit}} = \frac{m}{\sqrt{V}} + b$$

m and b are calculated by DOMCal, and the average t_{transit} time is ~130 ns.

Charge and Time Extraction: WaveDeform

- **WaveDeform** takes a calibrated "raw" waveform and fits for the arrival time and charge of pulses in the waveform
- Uses the known DOM response to single photons to fit pulses to the waveform
- The result is a "Pulse Series" of time and charge information which is much smaller in size than a raw waveform, but is enough information to use for reconstructions and physics for most events.

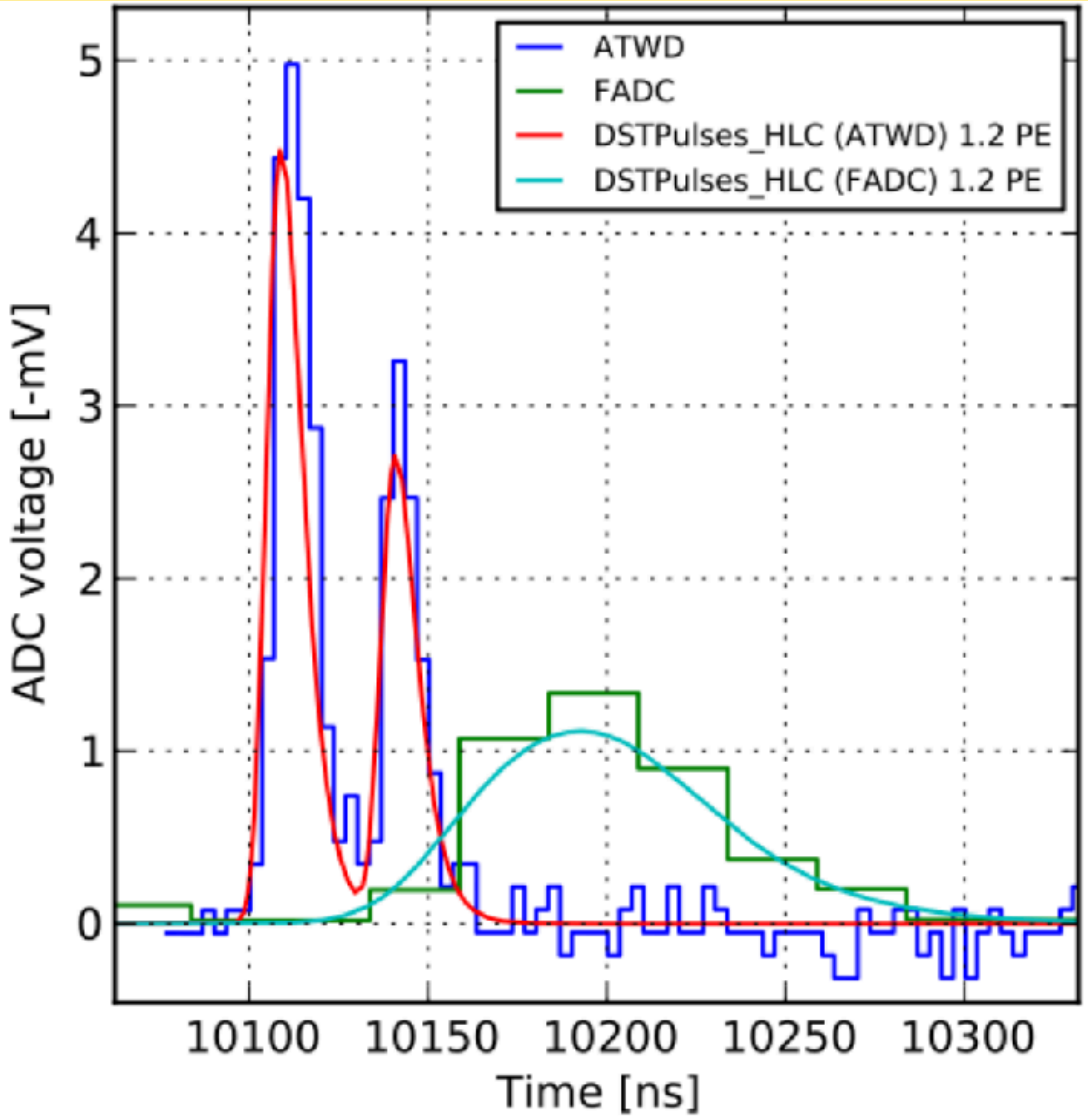




WaveDeform takes a raw waveform and gives time and charge information

Data Compression: SuperDST

- Reconstructed waveforms are compressed using a format known as **SuperDST** (Super Data Storage and Transfer)
- Using ~9% of the space, entire waveform information can be stored
- Slight errors on time and charge values due to compression are negligible compared to uncertainties on those values
- Raw waveforms are still sent in cases where WaveDeform/SuperDST performs poorly, or the waveform has high charge.

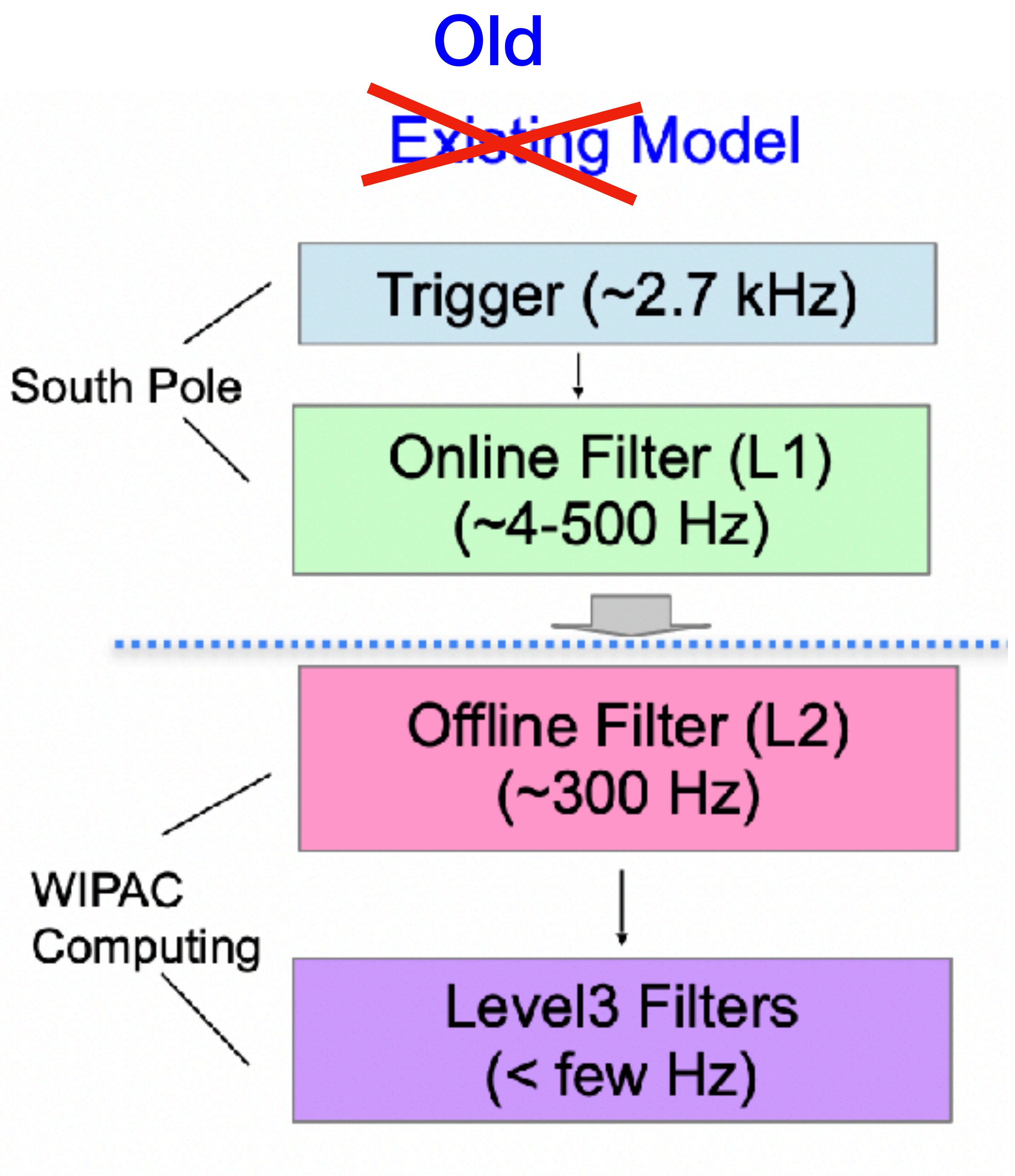


SuperDST reconstructed waveforms compared to "raw" waveforms

Pass2 Triggering and Filtering Model

Prior T&F Model (Pre-SuperDST):

- Online filters at the Pole are used to filter events for specific analyses
- These reduce the rate significantly, so less-well-compressed data can still be sent via satellite to the North



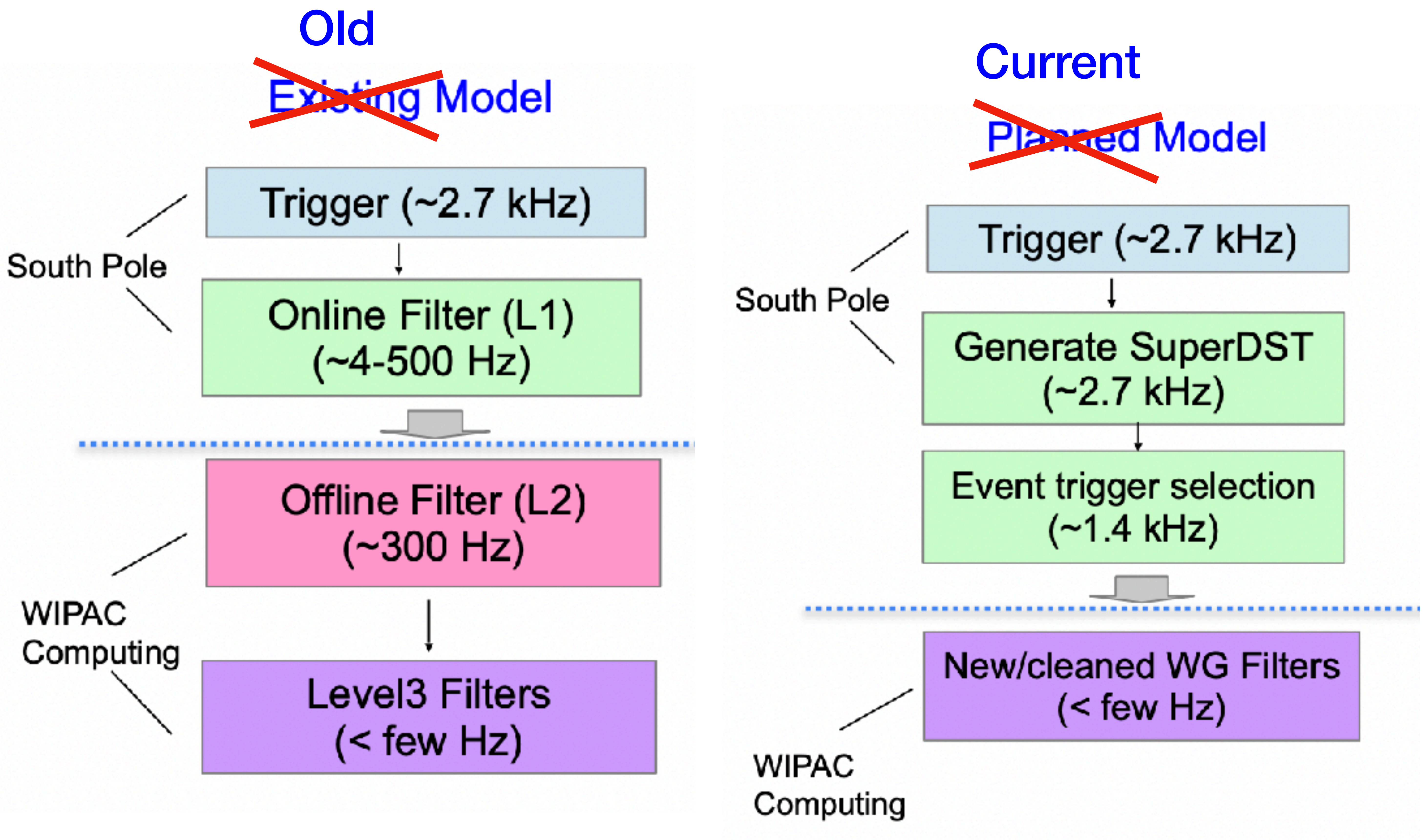
Problems:

- Online filters poorly maintained, use old code, and thus require more specific knowledge to update
- SuperDST means we can send almost every event North, so little reason to cut events with poorly maintained filters.

Pass3 Triggering and Filtering Model

New Model Changes:

- Apply new calibration and feature extraction to all events
- Only apply online reconstructions & selections needed for realtime events
- Can't quite send everything North… so send events passing through one of a couple re-trigger options (SMT12+, DC, IceTop, etc…)
- Once in the North, offline processing filters can be applied

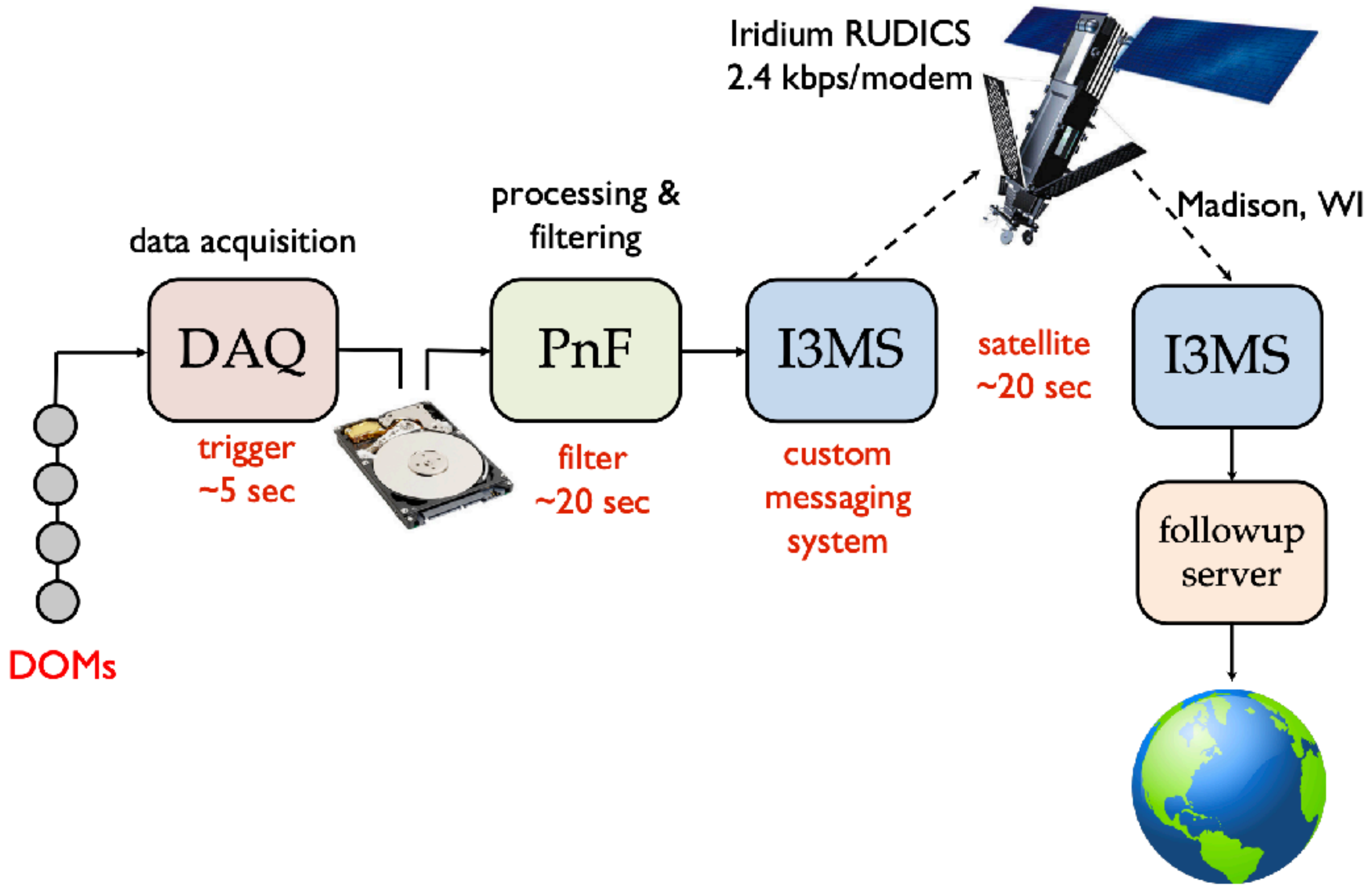


* New Model in use since 2023 season!

Real-Time Alerts

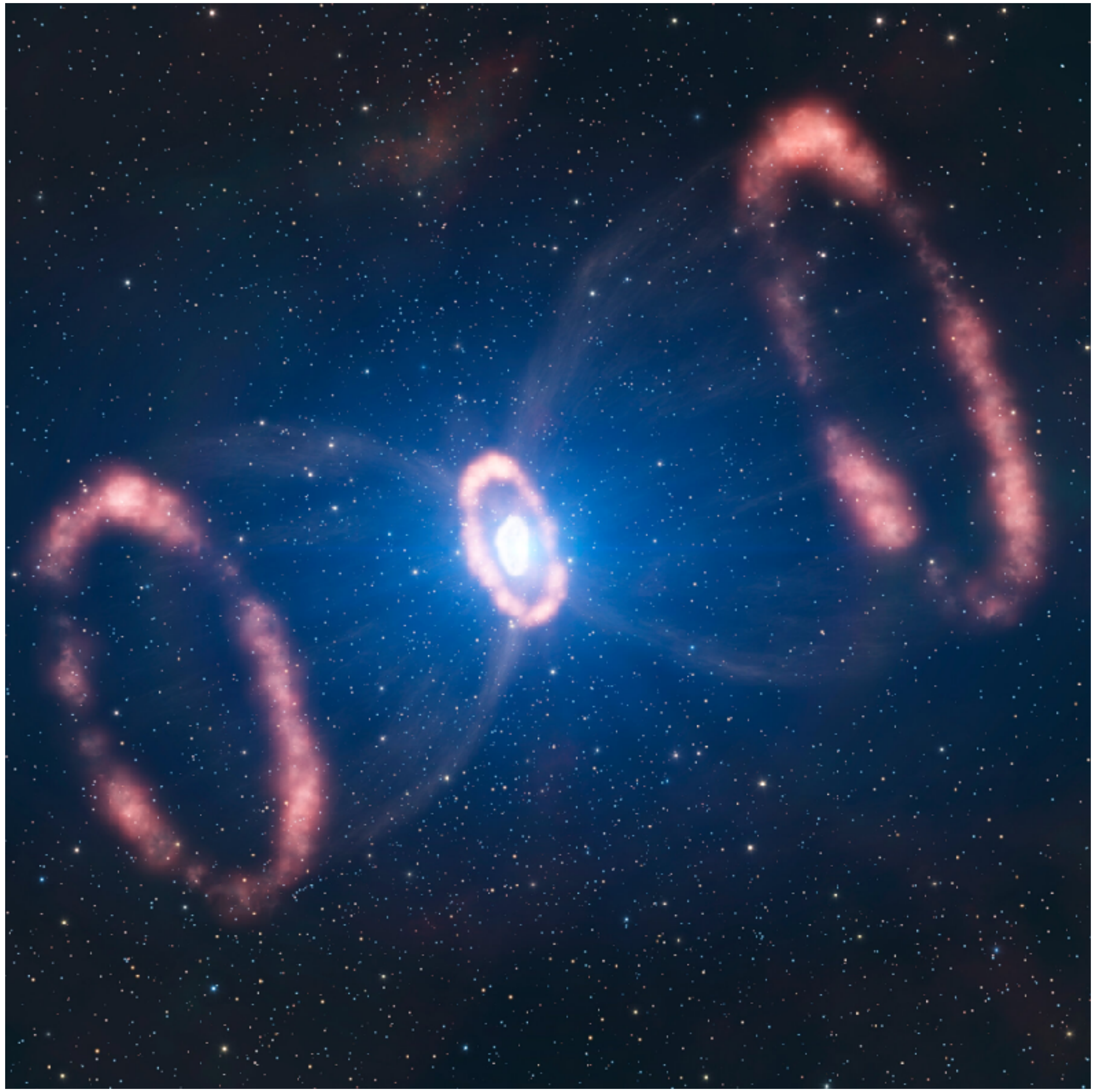
IceCube can also send alerts to the community about interesting Astrophysical Events immediately from the Pole to maximize follow-up time.

- Fast reconstructions for direction and energy are completed at the Pole
- Messages sent within ~1 minute over a separate satellite system



Supernova DAQ (SNDAQ)

- IceCube looks for SN by looking at the noise rate over time for all in-ice DOMs
- Look for a rise in the noise rates in many DOMs, possibly indicating a SN
- Separate SN Alert system (SNEWS)

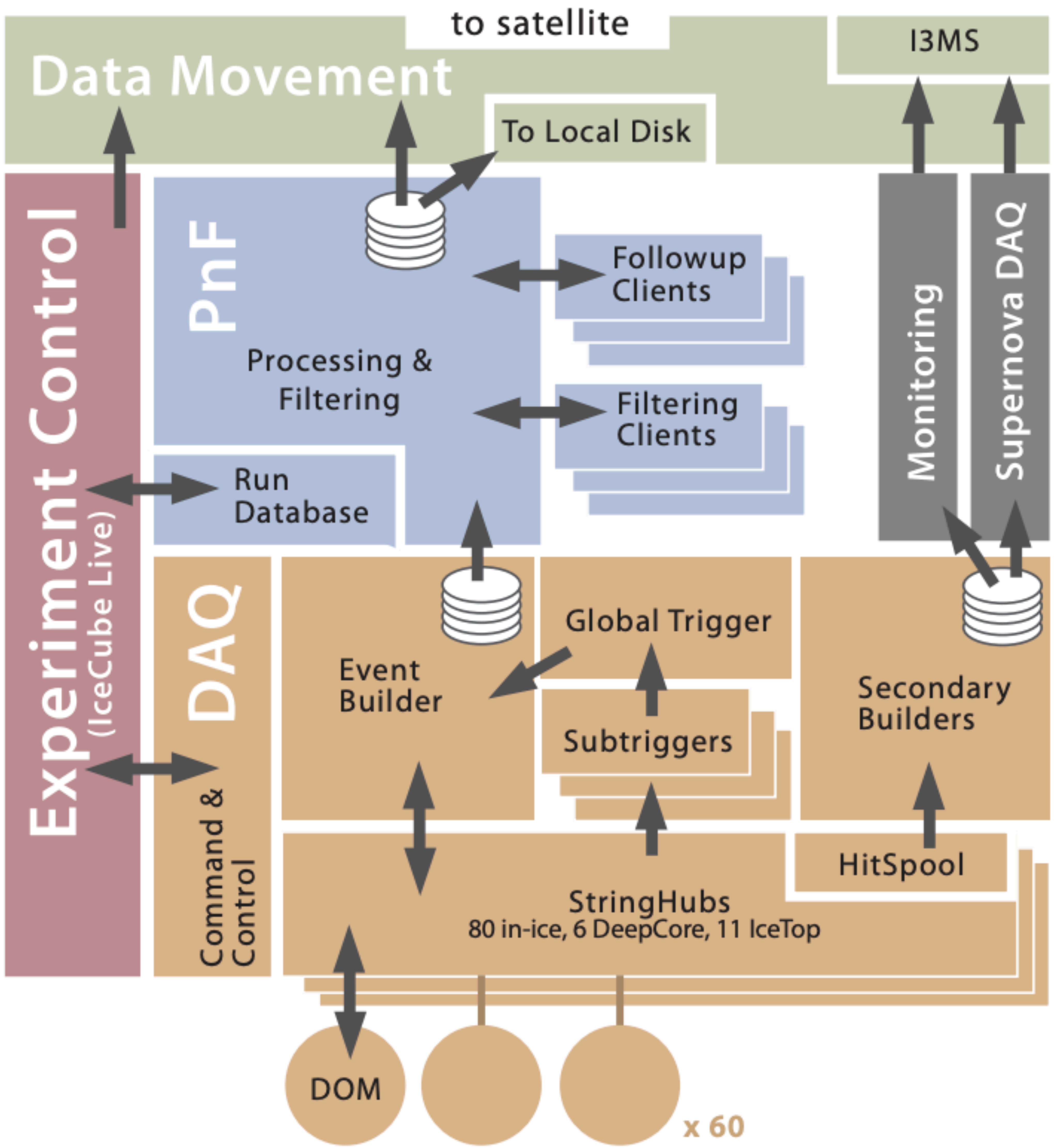


Hitspool and Data Storage

The **Hitspool** system saves all hits (time + waveform) to a local disk that can be used to recover full waveform information.

These are saved for ~12 days before being overwritten on the hitspool cache.

All events are also stored on local disks which can be manually transported to the North.



More Information

- Previous years' bootcamp/summer school presentations: http://wiki.icecube.wisc.edu/index.php/Bootcamp (some slides especially taken from John Kelley's detector presentation from last year)
- IceCube PMT Paper https://docushare.icecube.wisc.edu/dsweb/Get/Document-53922/
- IceCube DOM-DAQ Paper "The IceCube Data Acquisition Subsystem: Signal Capture, Digitization, and Time-Stamping" Nuclear Instruments and Methods in Physics Research A 601 (2009) 294–316 https://arxiv.org/pdf/0810.4930
- IceCube Detector Paper "The IceCube Neutrino Observatory: Instrumentation and Online systems" Journal of Instrumentation 12 (2017) P03012 https://arxiv.org/pdf/1612.05093.pdf