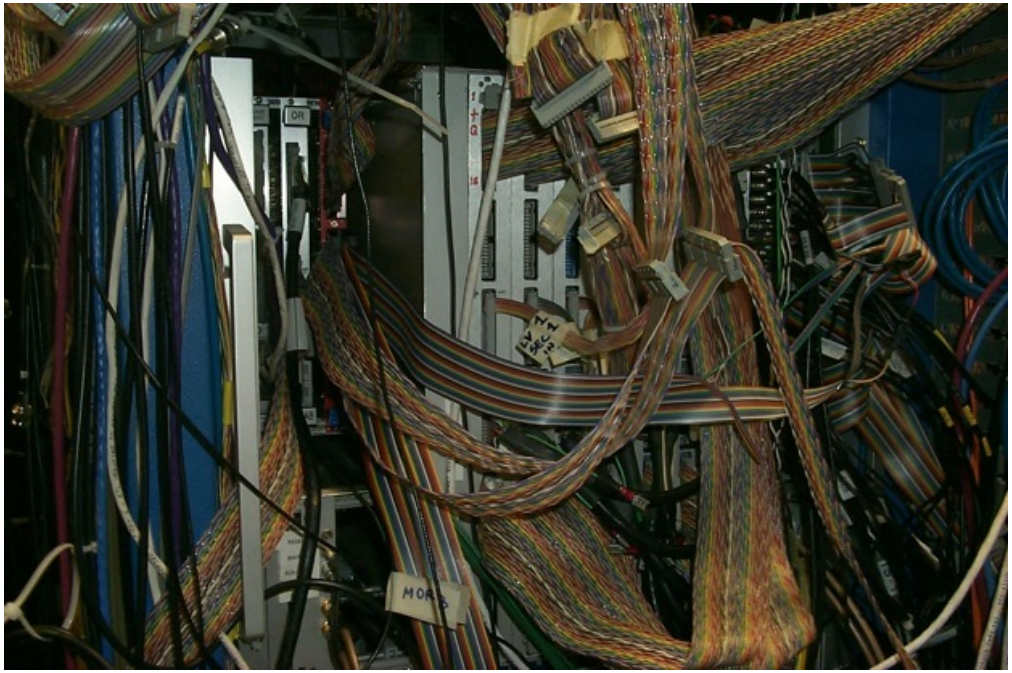
Data Acquisition at JLab



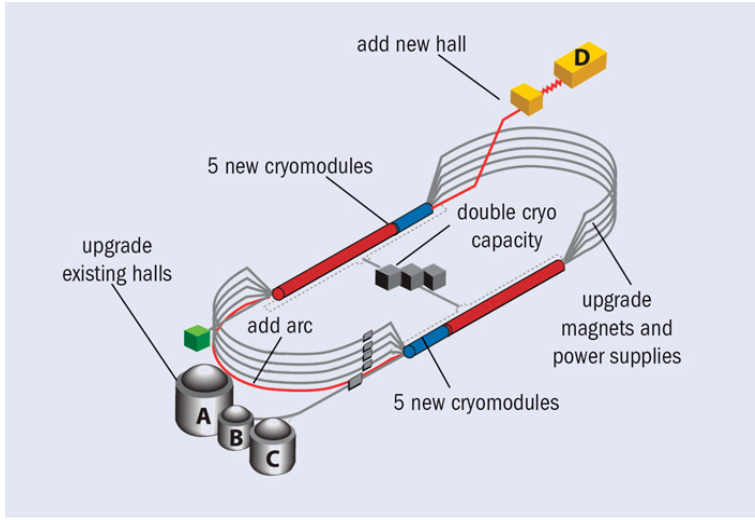
Graham Heyes

Data Acquisition Support

Experimental Nuclear Physics

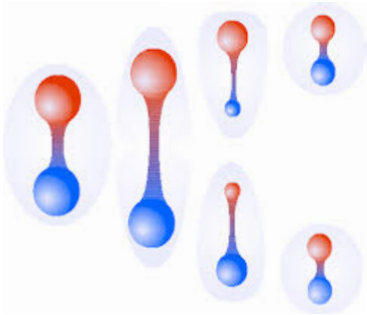
Jefferson Lab

- The principle goal of Jefferson lab is nuclear physics research using the CEBAF electron accelerator.
  - Two superconducting LINACs with recirculating arcs.
  - Simultaneous beam to multiple halls.
- Each hall has equipment designed to study different, but complementary, aspects of matter in the nucleus.

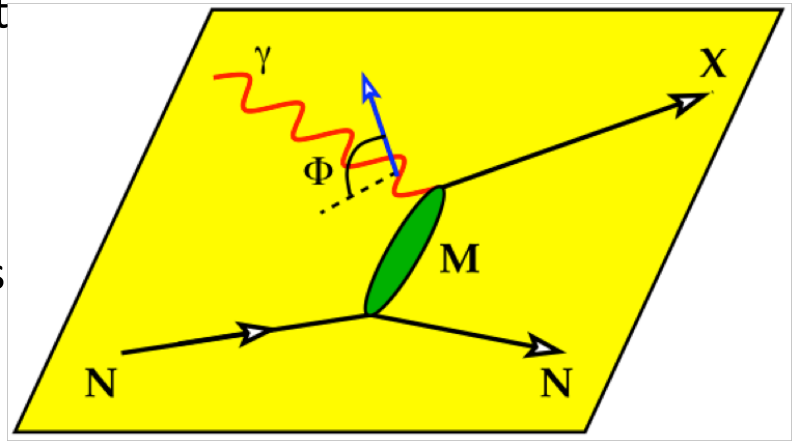


Physics

- In this talk I will use the GLUEX experiment at JLab as an example.
- Both theory and experiment tell us that quarks cannot be found individually they are always found in at least pairs (mesons) or threes (baryons, like protons and neutrons). (Or fives, according to LHCb)
  - This is called confinement.



- The specific goal of the GlueX Collaboration at JLab is to better understand confinement.
- The experimental setup is to hit baryons with high energy photons to stretch the glue between the quarks until it snaps and mesons are produced
- Produce many photon interactions and study the interaction using statistics.



Detection and data acquisition

- When a particle interacts with a target it, and any particles produced by the interaction, and their byproducts (for example mesons decay quickly into other particles) spread out from the interaction point.

- An array of detectors collect energy deposited by these particles and convert it into an electrical signals.

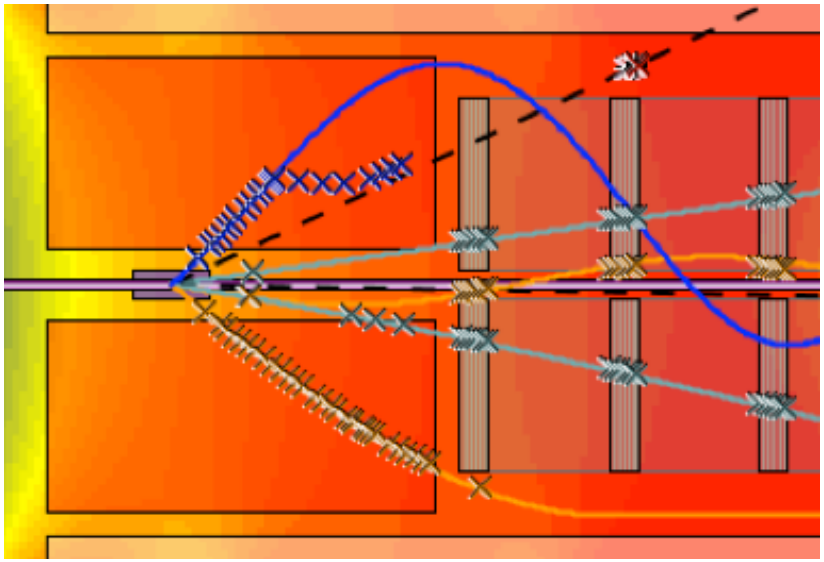
- Three basic types of measurements are charge, time and count.

- A combination of electronics and software convert the electrical signals into digital data in a format that can be stored and later analyzed.
  - ADC = charge, TDC = time, Scaler = count

- All of the data generated from one interaction is called an event.

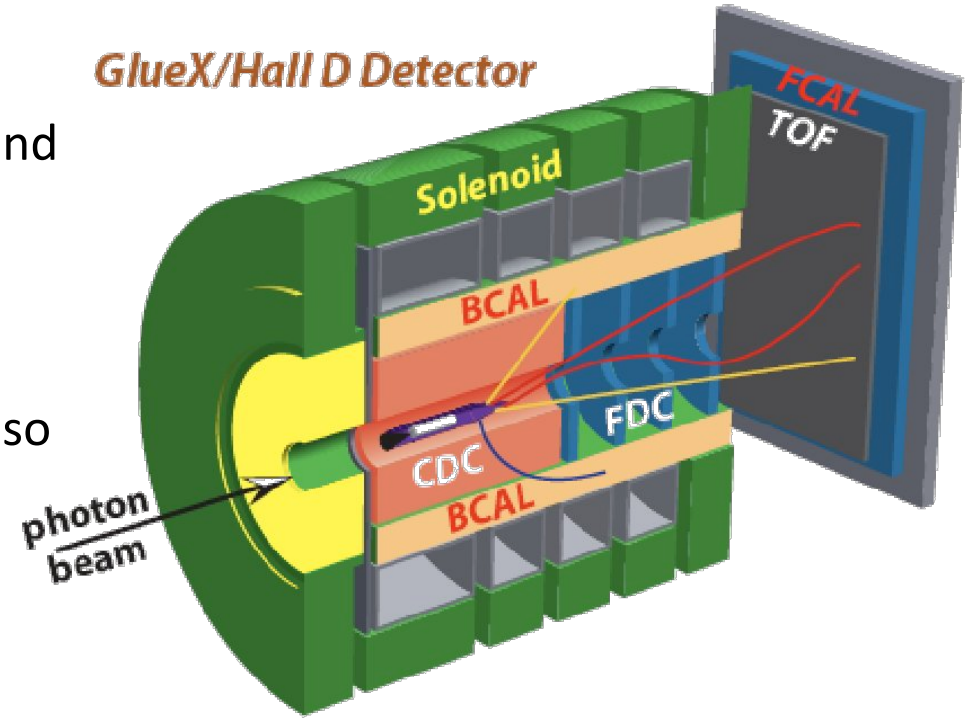
Properties of nuclear physics data

- Data from one event has no history. It doesn't depend upon events that went before and doesn't influence later events.
- Events occur with random timing.
  - Hardware may not be ready for new data.
    - Dead time when data is lost.
  - Events may overlap in time, event pileup.
  - Peak event rate can be more than the average.
- Event size depends upon the physics.
  - Accidental hits unconnected with event.
  - Electronic noise.
  - Distribution of event sizes.
  - Some very large events.

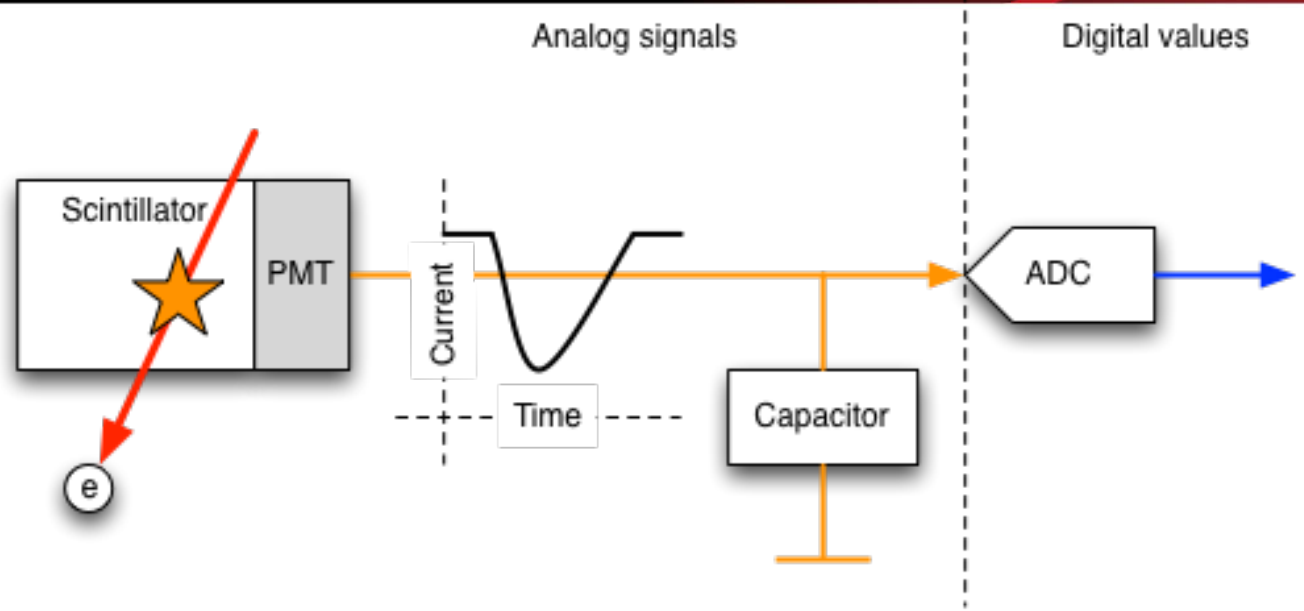


Other challenges for data acquisition

- Detecting hardware is physically distributed within the detector so we need to:
  - Log where the data came from (and when).
  - Gather the data from one event together.
- Experiments run for months or years so need to:
  - Control the whole system.
  - Monitor conditions under which data was taken.



Detector example, a scintillator

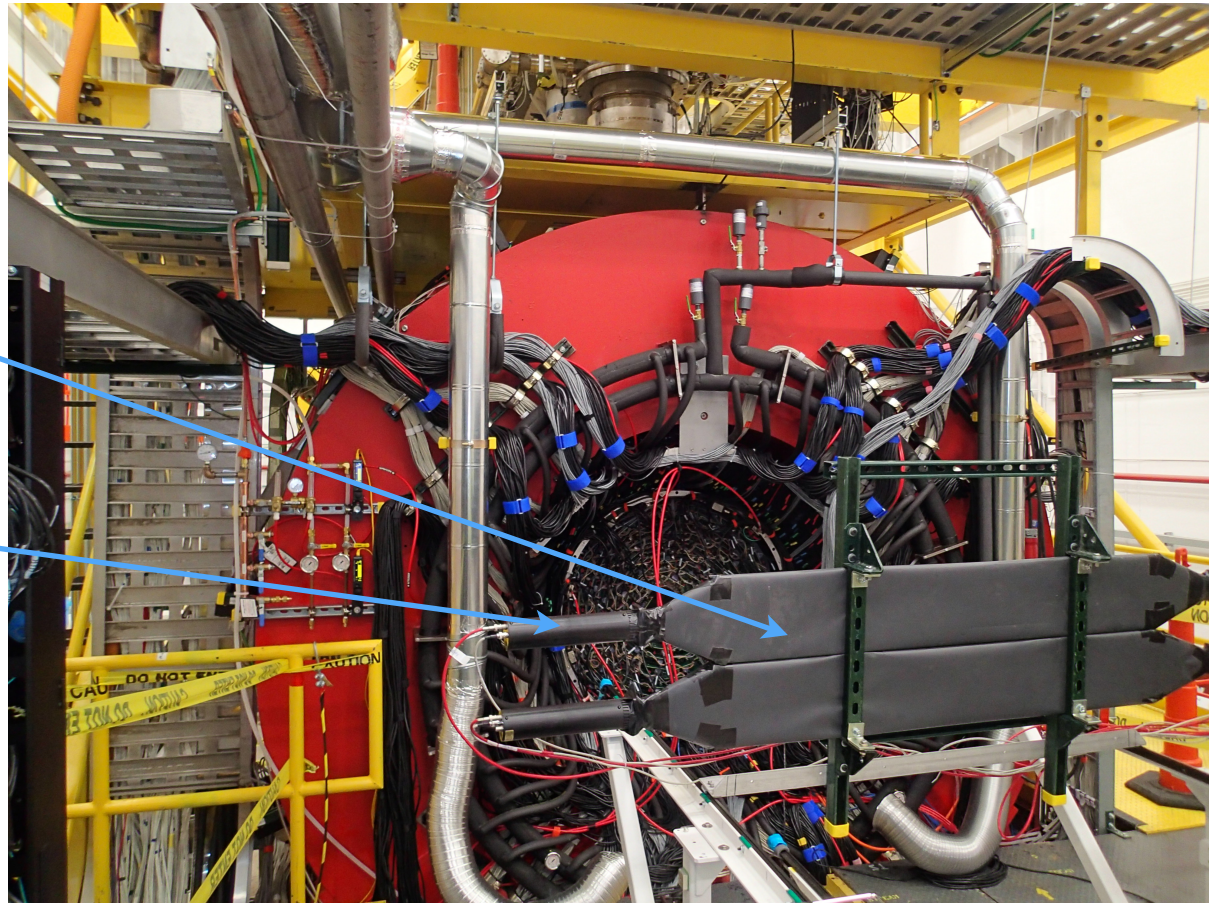


- A particle deposits energy in a scintillating material that converts it into light.
- A Photo Multiplier converts the light into a pulse of electricity.
- The pulse charges a capacitor.
- An ADC converts the voltage into a digital value.

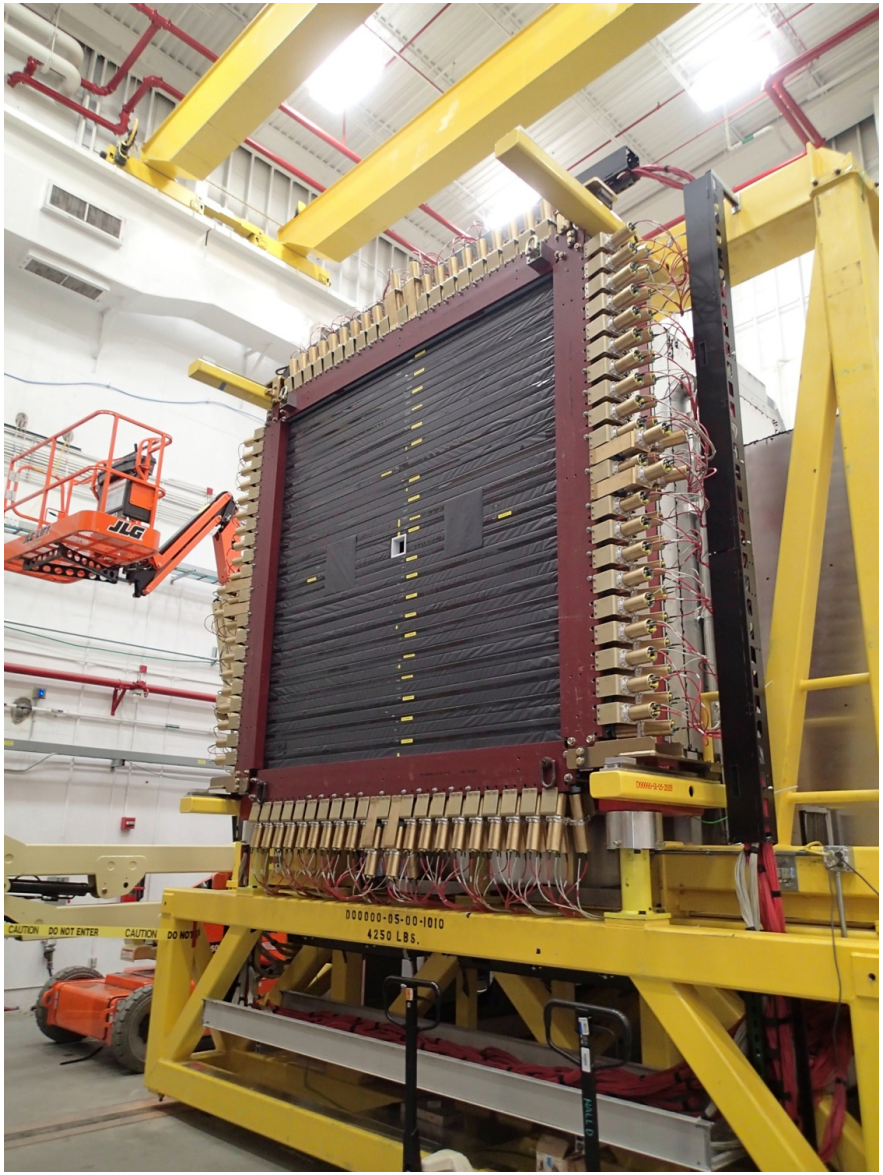
Picture of test scintillators in hall-D

Scintillator

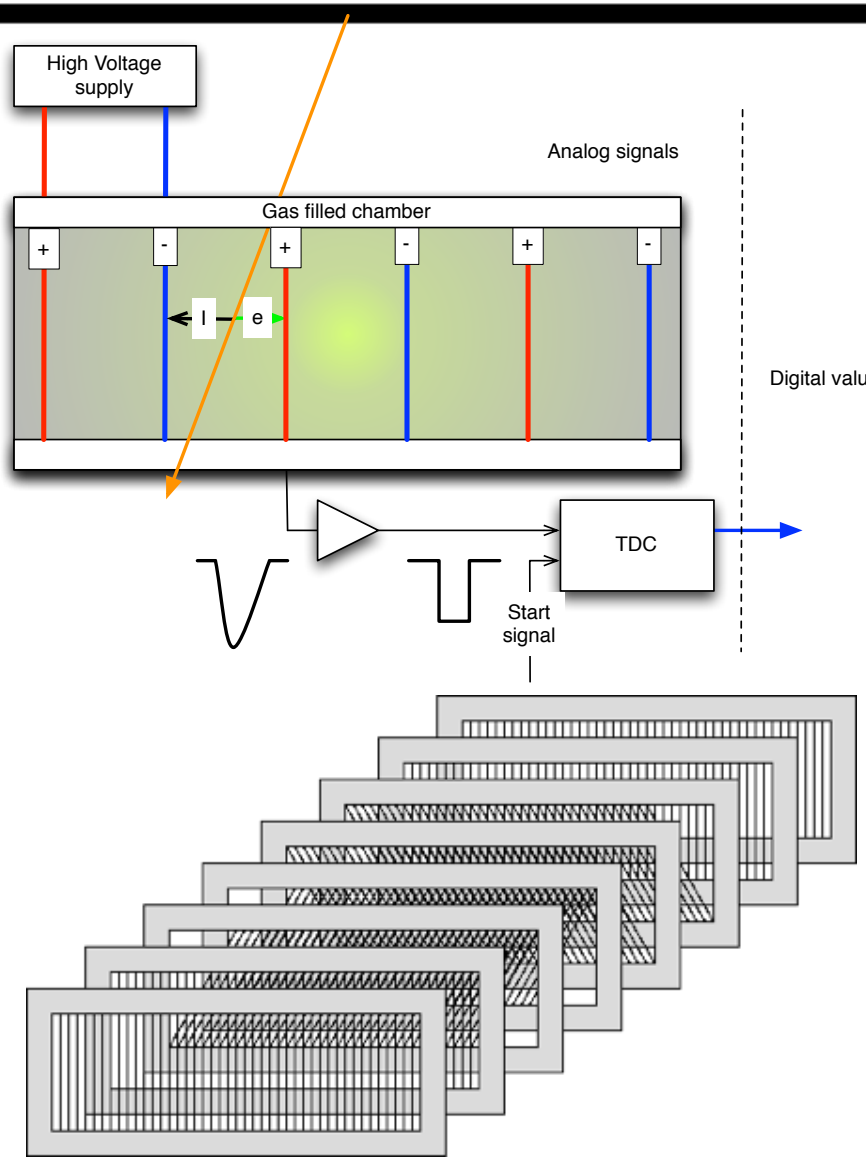
Photomultiplier



They come in big sizes too



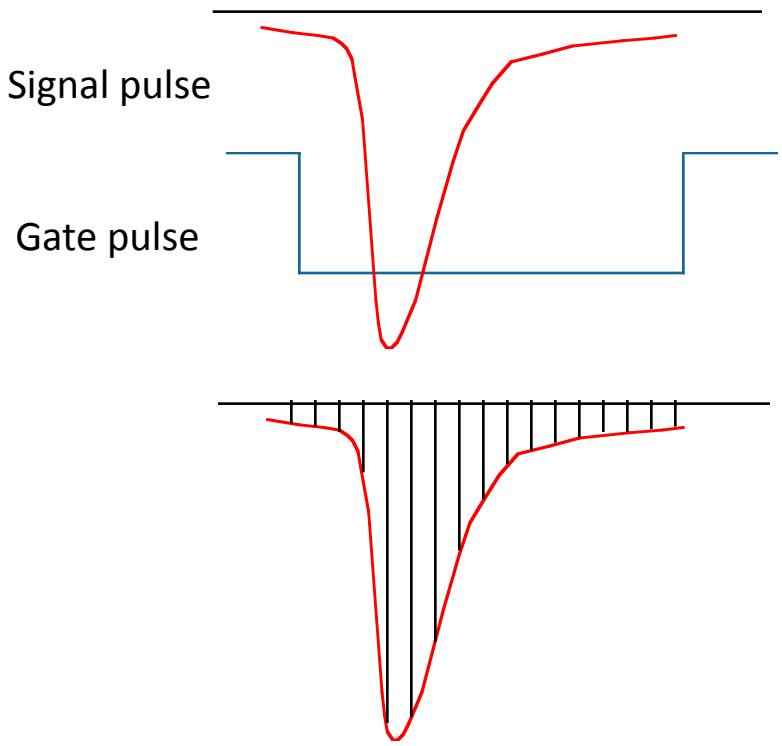
Detector example, a wire (or drift) chamber



- A chamber contains planes of parallel wires and a special gas mixture. A high voltage is applied to the wires so that alternating wires are charged positive and negative.
- A particle ionizes the gas. Ions drift to negatively charged wires and electrons to positive.
- A signal from another detector is used to start a Time to Digital Converter (TDC).
- The electrons produce a pulse which is used to stop the TDC.
- The time value measures the distance the electrons drifted. In combination with the position of the sense wire this tells us where the particle crossed the plane.
- Several planes are used to reconstruct the particle track in three dimensions.

Sampling vs Integration

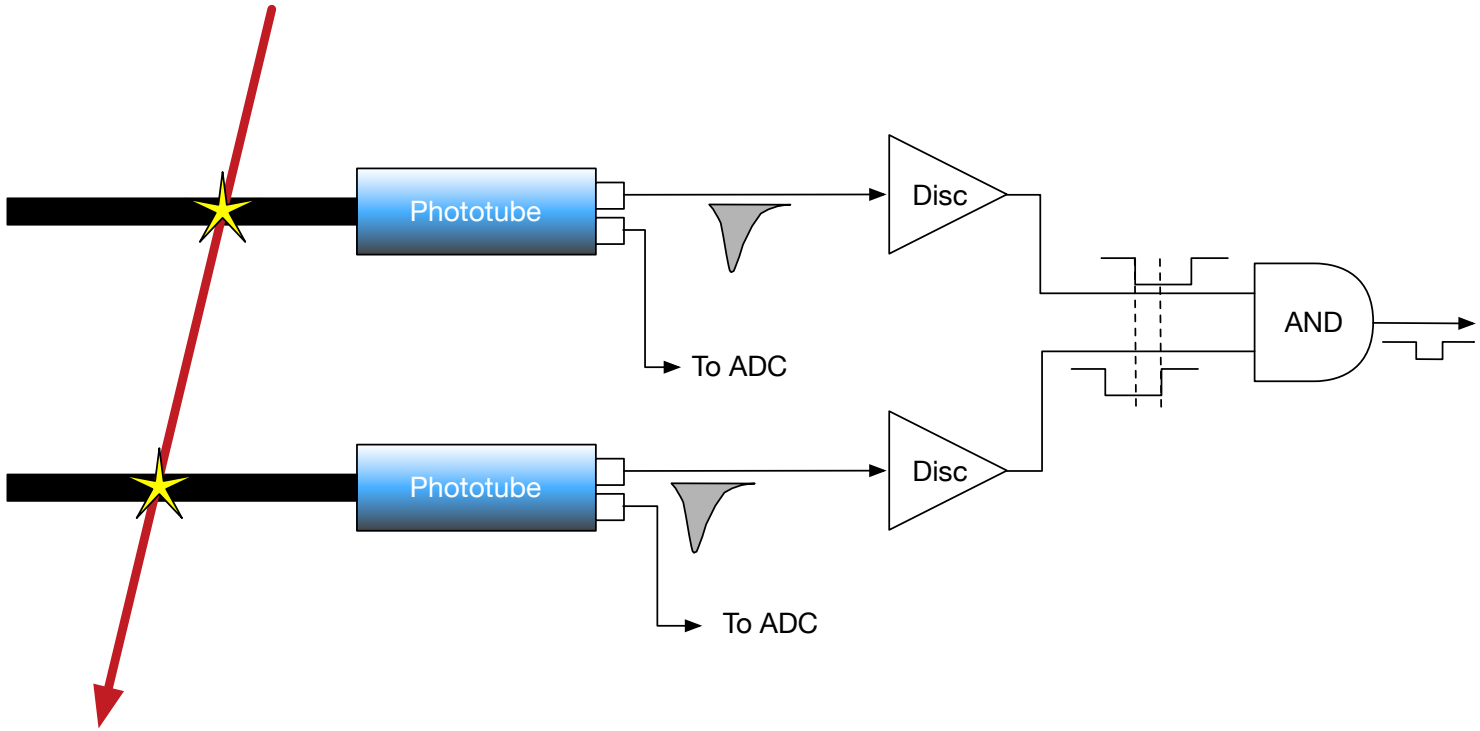
- A traditional “integrating” ADC takes 6 to 10 µsec to digitize a pulse. A gate pulse, generated by other detectors marks the region of interest to electronics electronics that integrates the charge from the signal pulse.
- This type of ADC generates a single measurement representing the charge sum during the gate.



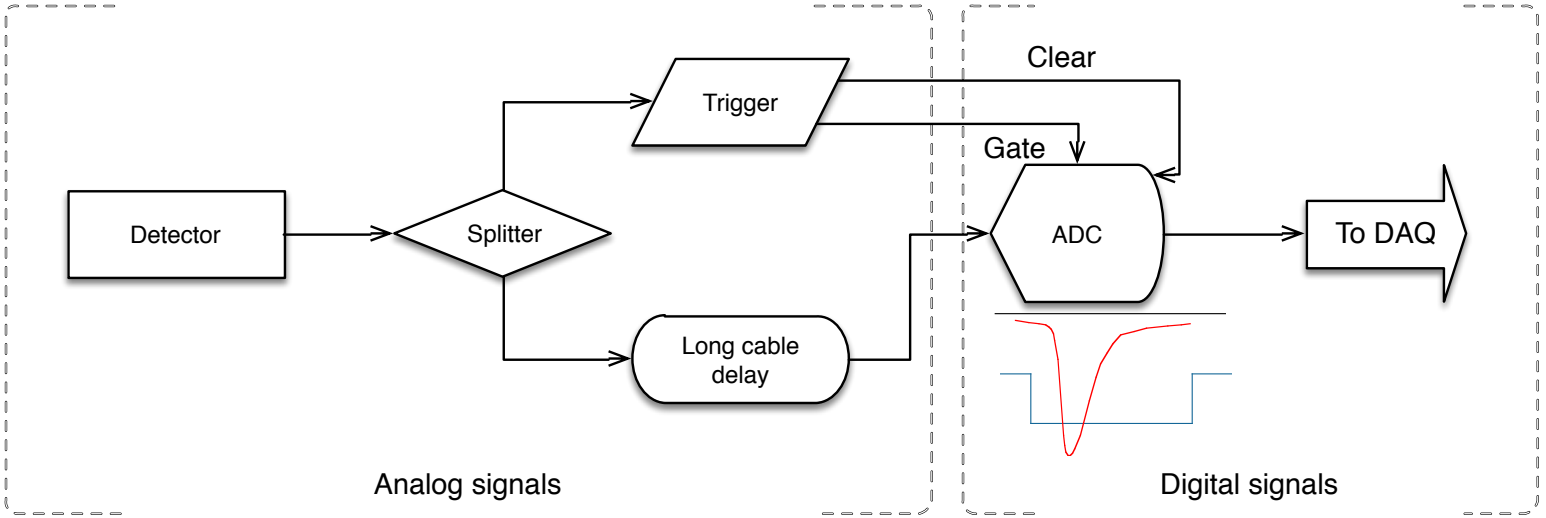
- A Flash ADC samples continuously at a fixed rate.
- In this example a 250 MHz ADC samples every 4 nsec and generates ~10-15 measurements during the gate. These describe the pulse shape as well as charge.

A Simple Trigger

- That's great but how do we know the signal came from an event and not a random fluctuation?
  - Fortunately we have more than one detector.
  - Combine data from different detectors to characterize events.
  - Determine which events are interesting.

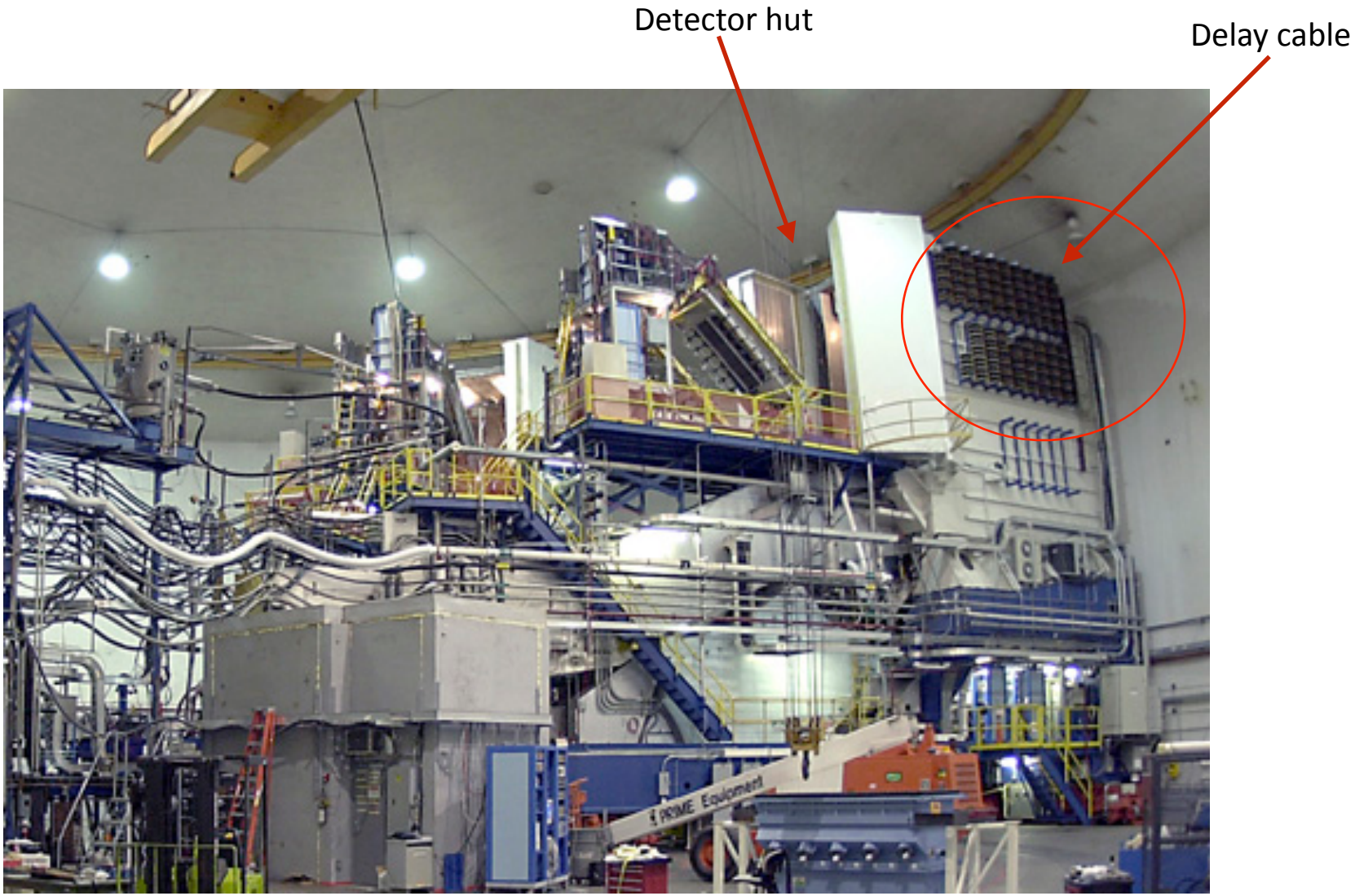


An analog trigger

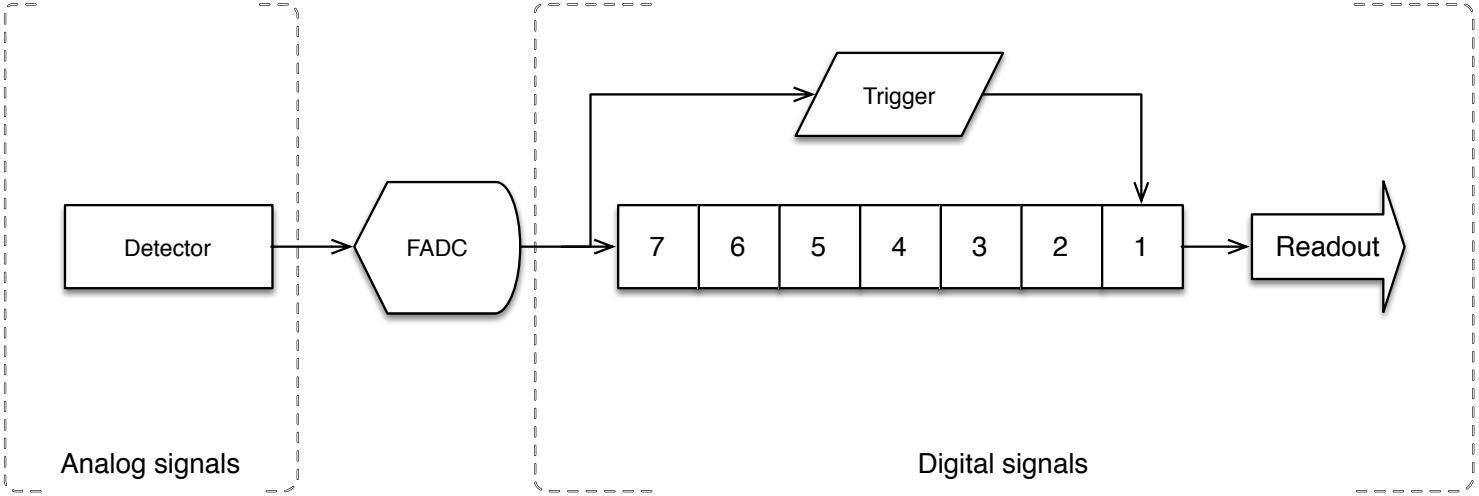


- It takes some time for the trigger logic to decide if a signal should be digitized.
- The analog signal must be delayed so that the gate and signal arrive at the ADC at the same time. Typical coax cables ~1 ns/ft so you could simply delay the signals using long cables.
  - Matching cable lengths is very important.
  - The ADC cannot process a new signal until it is read or cleared.
  - This limits the trigger rate.

Here's the long cable in Hall-A.



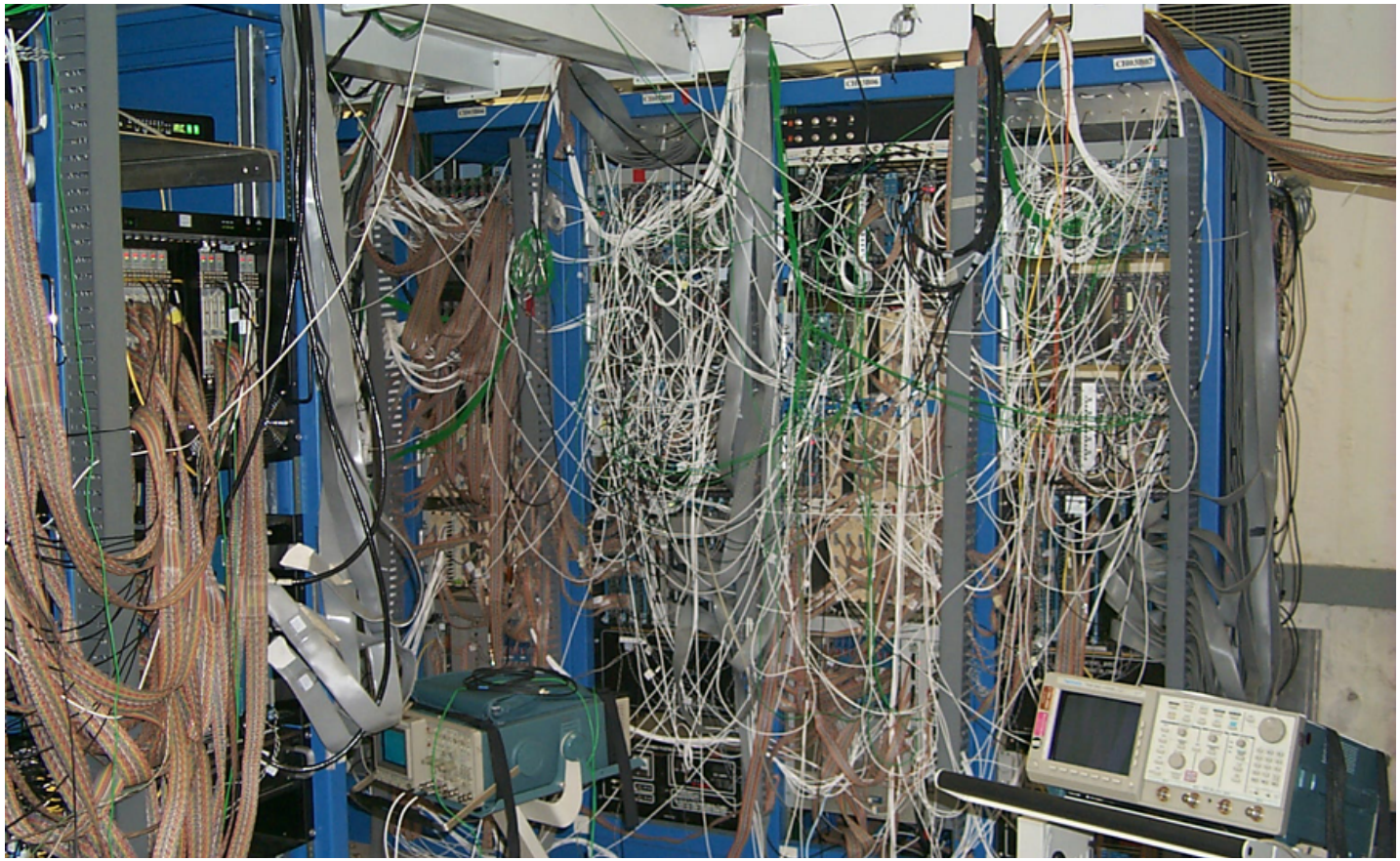
Pipeline trigger



- Replace all that cable with digital memory.
- In a pipelined system a Flash ADC digitizes at a constant rate and stores the values in a memory. Values are clocked into memory at the same rate as the Flash ADC clock which, in the case of GLUEX, is 250MHz (4 nS).
- If, for example, the trigger logic takes 28 nS we know that trigger corresponds to measurements 7 cells down the memory pipeline.
- The readout software can read and clear the entire memory at once.

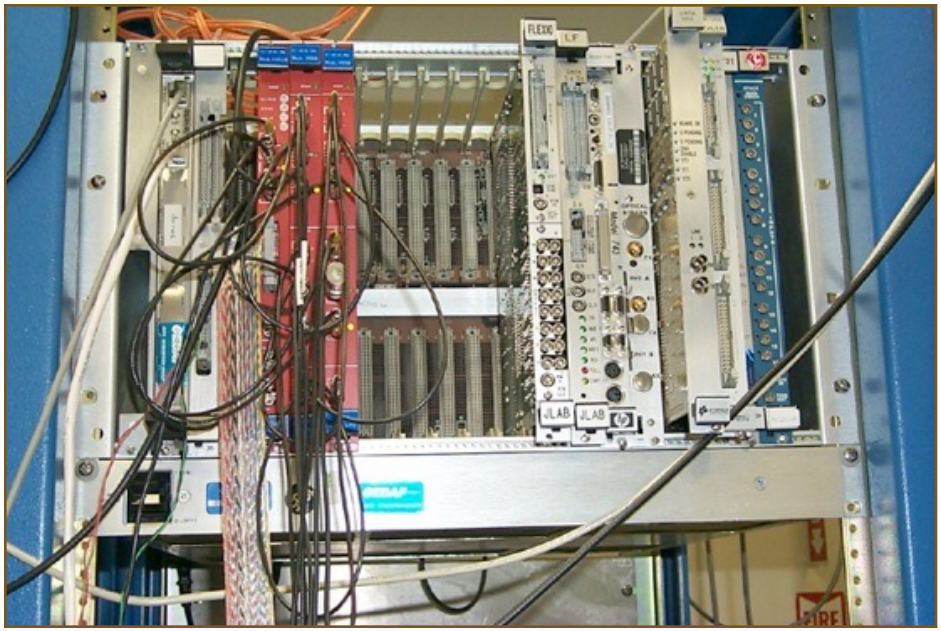
Trigger logic

- Triggers used to use a lot of electronics wired together. We can't do that now:
  - Propagation times down cables limit trigger rates.
  - Modern experiments require very complex trigger decisions.

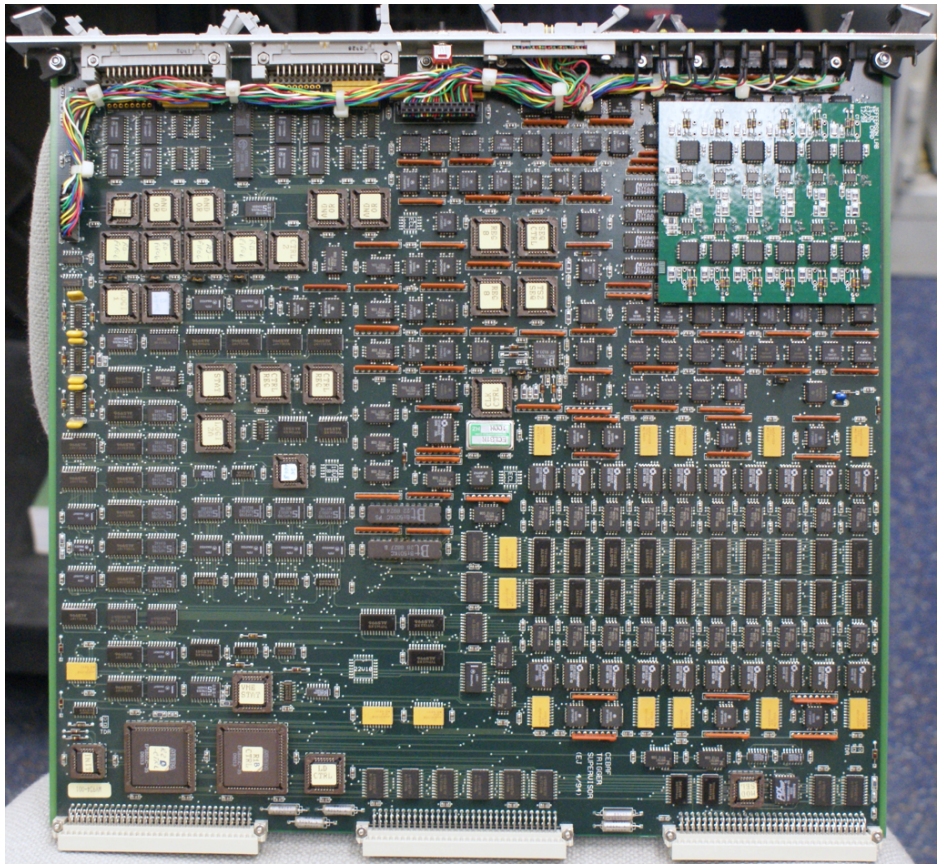


Putting together a system

- In reality there are many data sources in a detector like GLUEX so we need many ADCs, TDCs and other electronics.
- Devices are connected together using a bus.
  - GLUE uses VXS, which is a variant of the VME standard for interconnecting electronics.
- Boards slide into slots in a “crate” and plug into one or more backplanes that provide power and interconnect the boards.
- Usually there is a single board computer in the left most slot to configure and read out the boards.
- VXS has a third “backplane” that provides high speed serial data links between boards that used by the GLUEX trigger.



Complex electronics

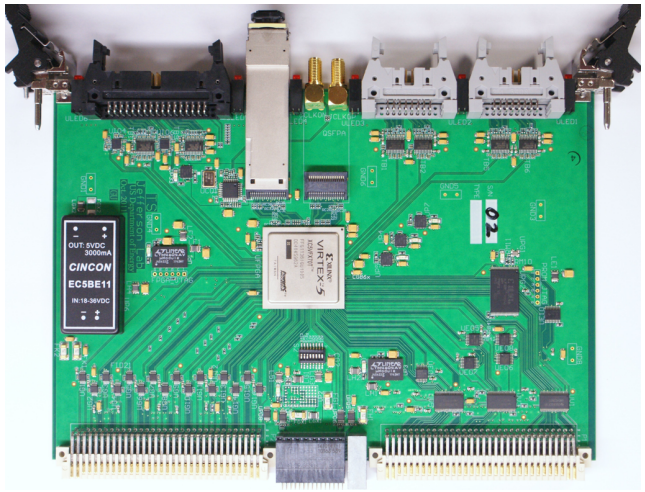


We can now buy programmable logic arrays that allow us to implement complex algorithms in the firmware on a single chip.

The two pictures are of boards with similar functionality.

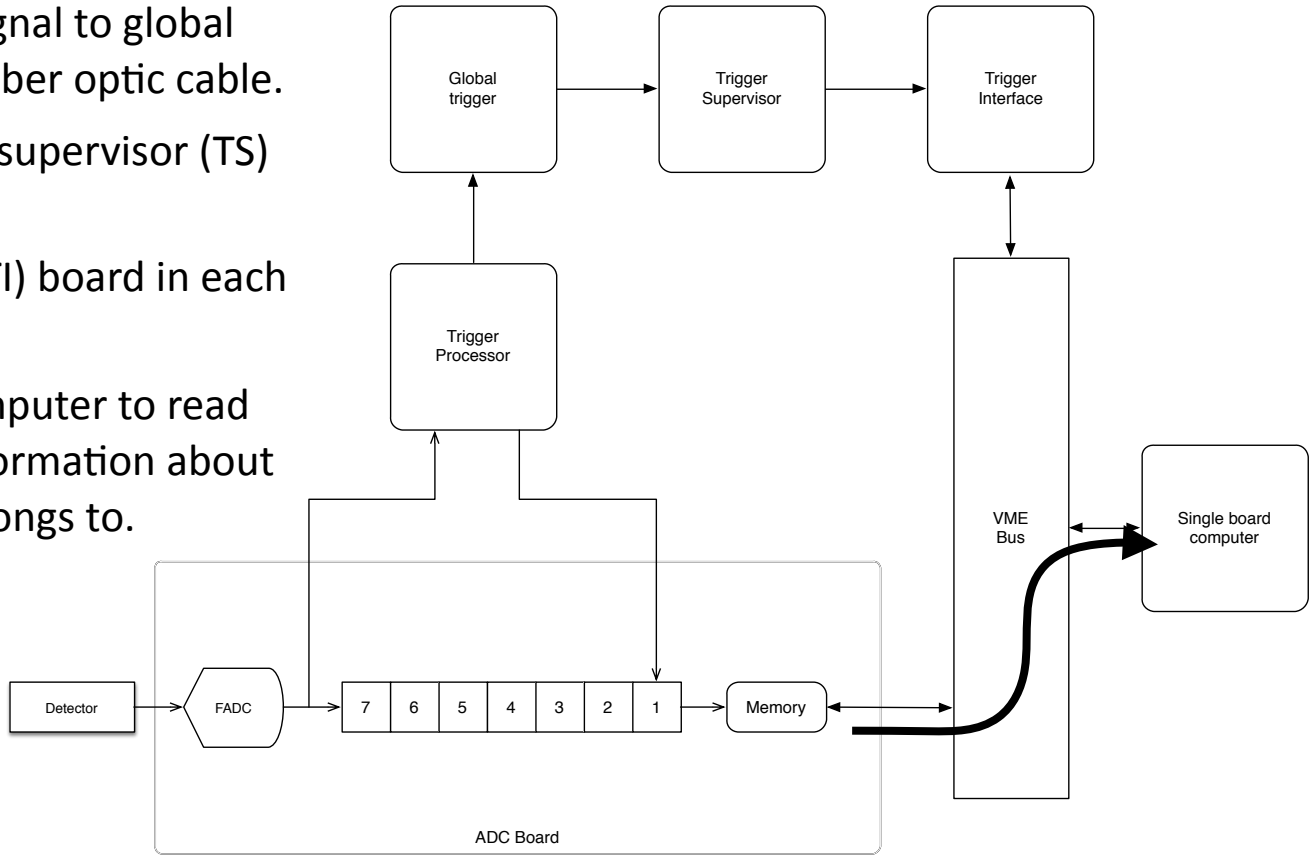
The one on the left designed in the early 1990's.

The one below designed in 2011.



GLUEX trigger

- Each ADC sends signals to a trigger processor over VXS serial bus.
- Trigger processor sends signal to global trigger for all crates over fiber optic cable.
- Global trigger tells trigger supervisor (TS) which events are good.
- TS tells Trigger Interface (TI) board in each crate.
- TI signals single board computer to read out crate and provides information about which trigger the data belongs to.



Real world ADCs

Intel CPU Read Out Controller (ROC) running Linux

Individual ADC channel

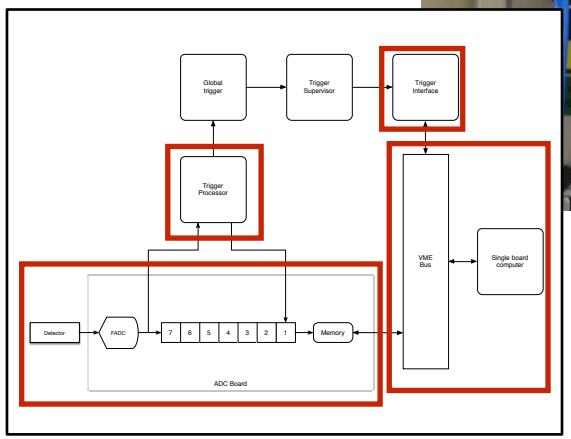
16 channels per board

17 boards = 272 channels per crate

Trigger interface

Boards are connected to CPU via a backplane bus.

Board sending signals to trigger over fiber

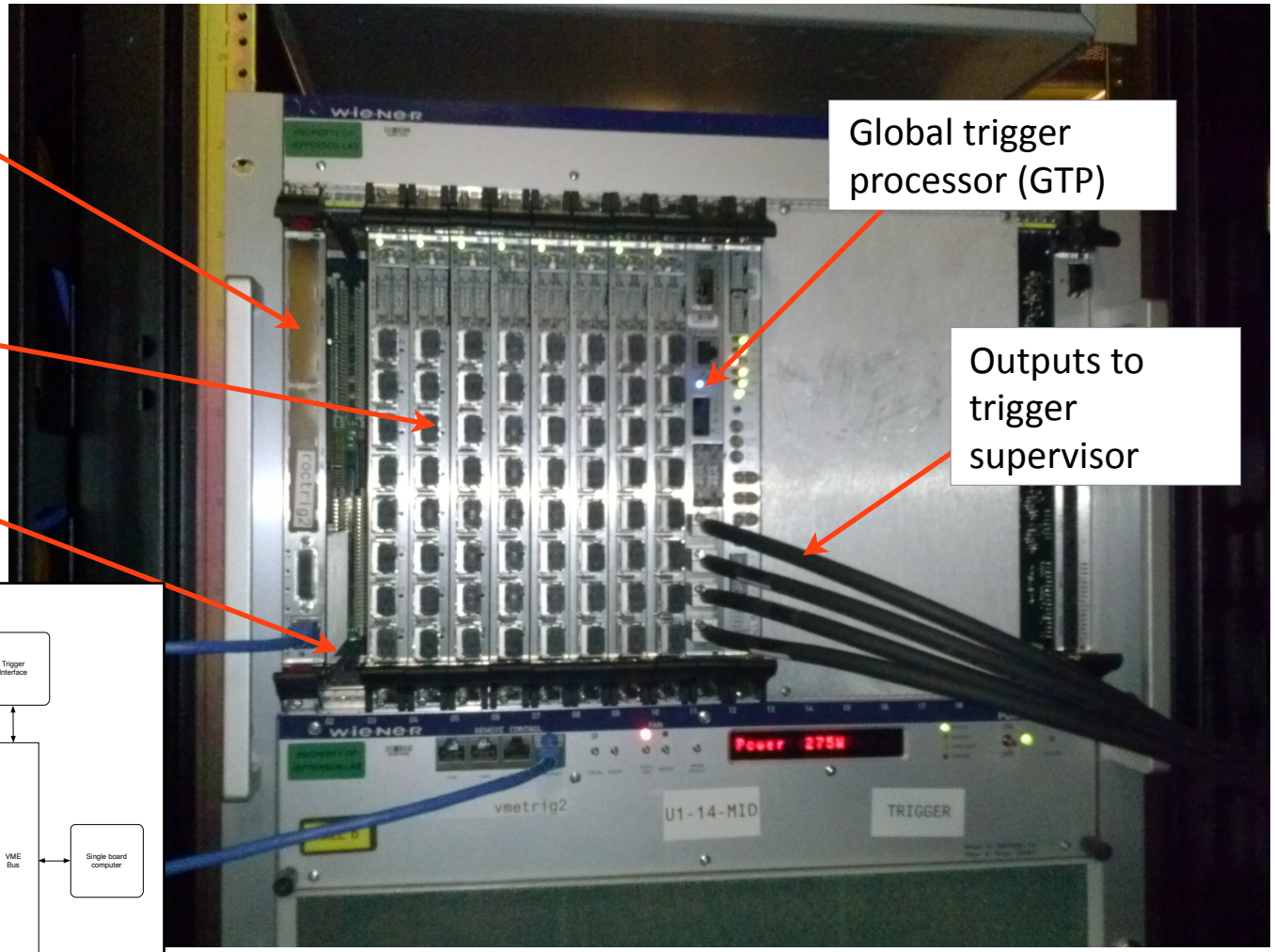


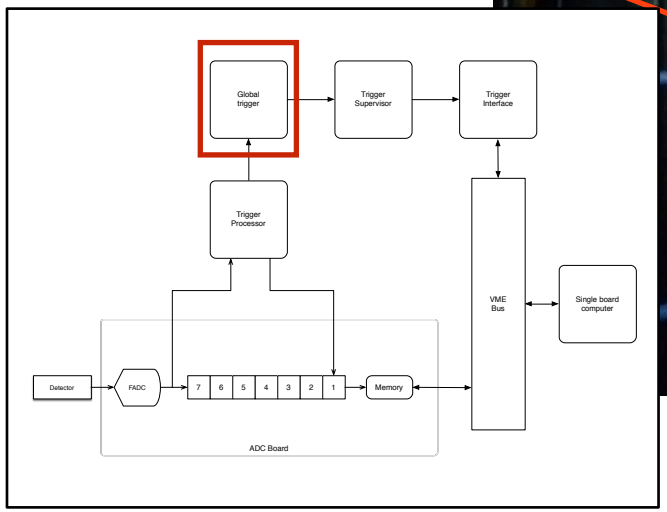
Global trigger crate

Intel CPU controller

Sub-system processor board (SSP)

Eight boards with eight connectors each so up to 64 crates.





Trigger supervisor crate

Trigger Distribution board (TD).

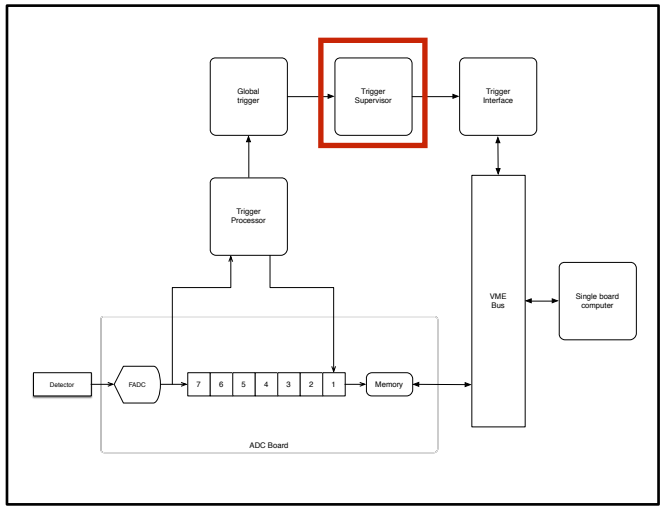
Intel CPU for control and configuration.

Optical trigger link back to crates.

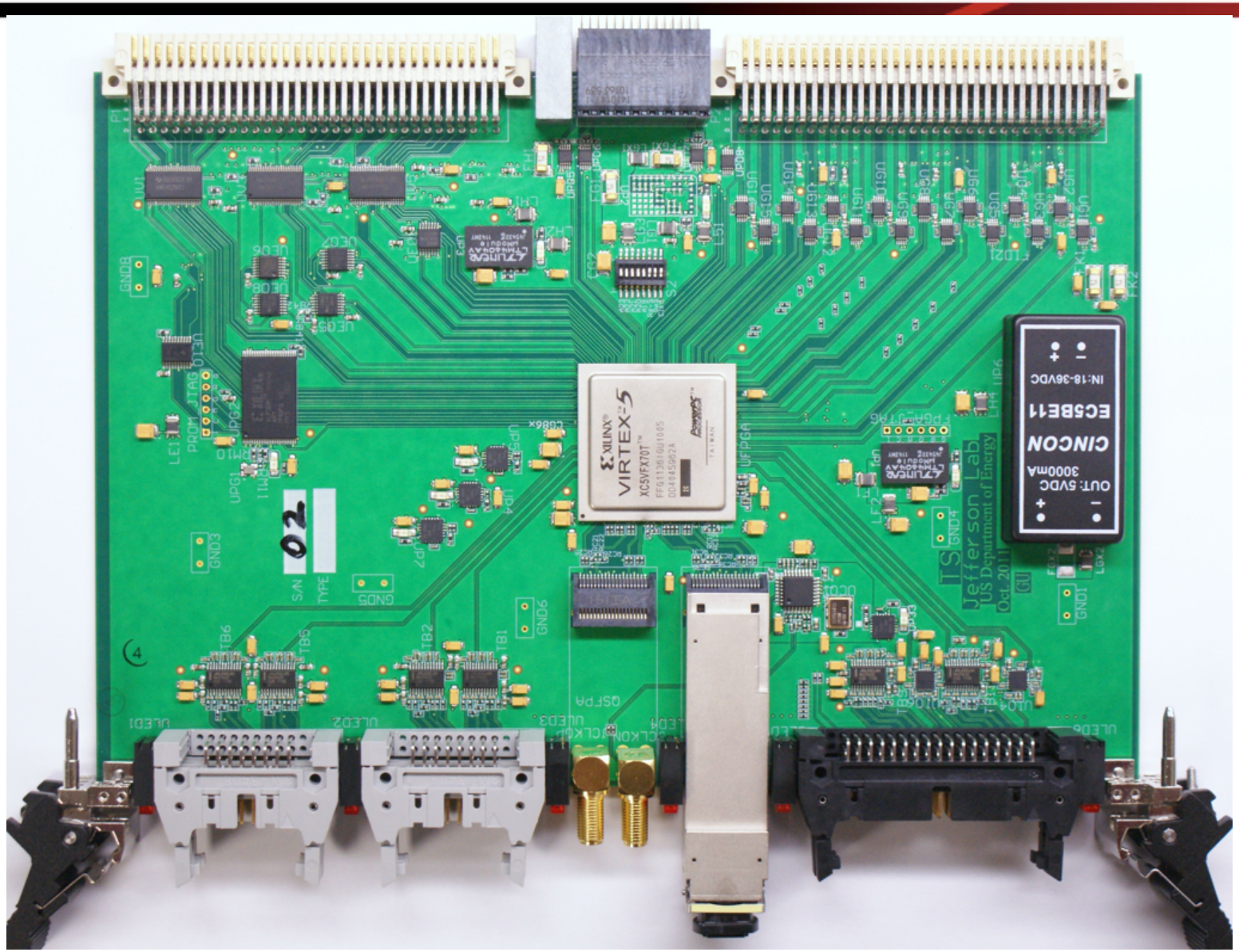
VXS serial backplane

Signal Distribution board (SD)

Trigger Supervisor (TS)



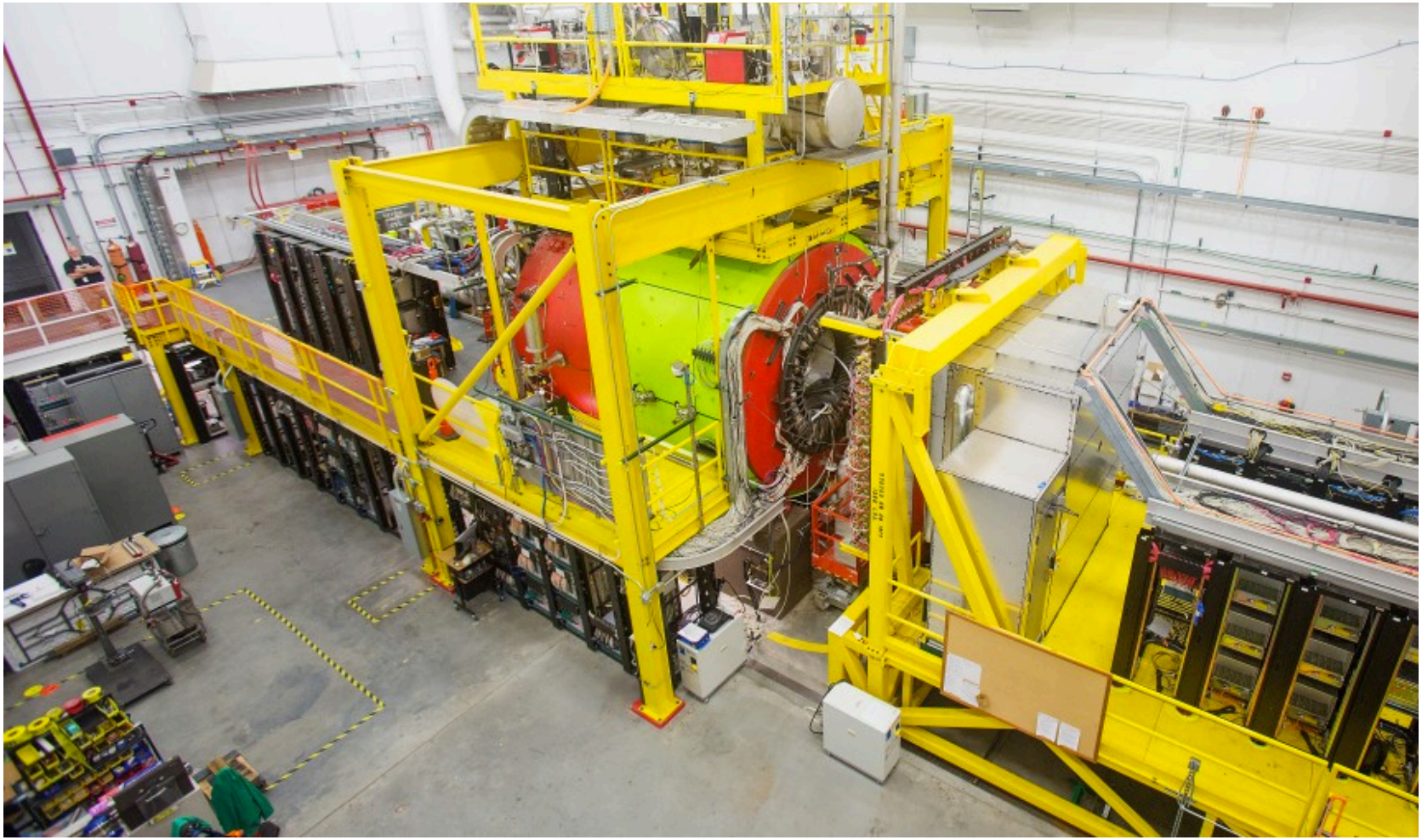
Trigger Supervisor board



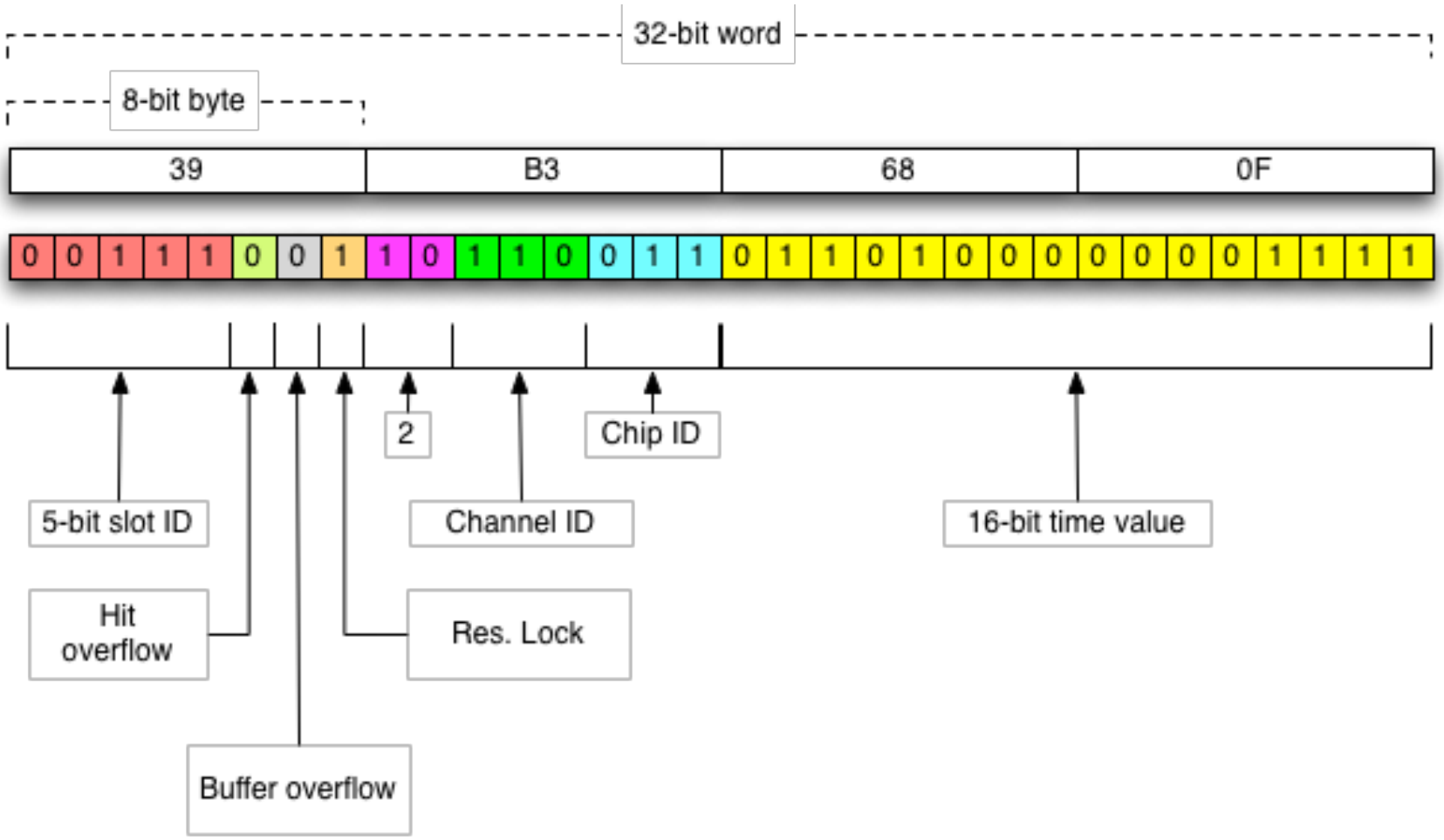
VME readout

- The TI board gets from the trigger supervisor:
  - Signal to read the memory of the ADC boards.
  - Trigger data telling the CPU which events the data belong to.
- The CPU copies ADC data into memory and wraps it in a format that contains the trigger data, which ADCs the data came from and which crate this is.
- Periodically the CPU sends the data over the network to the rest of the data acquisition system.

GLUEX DAQ electronics



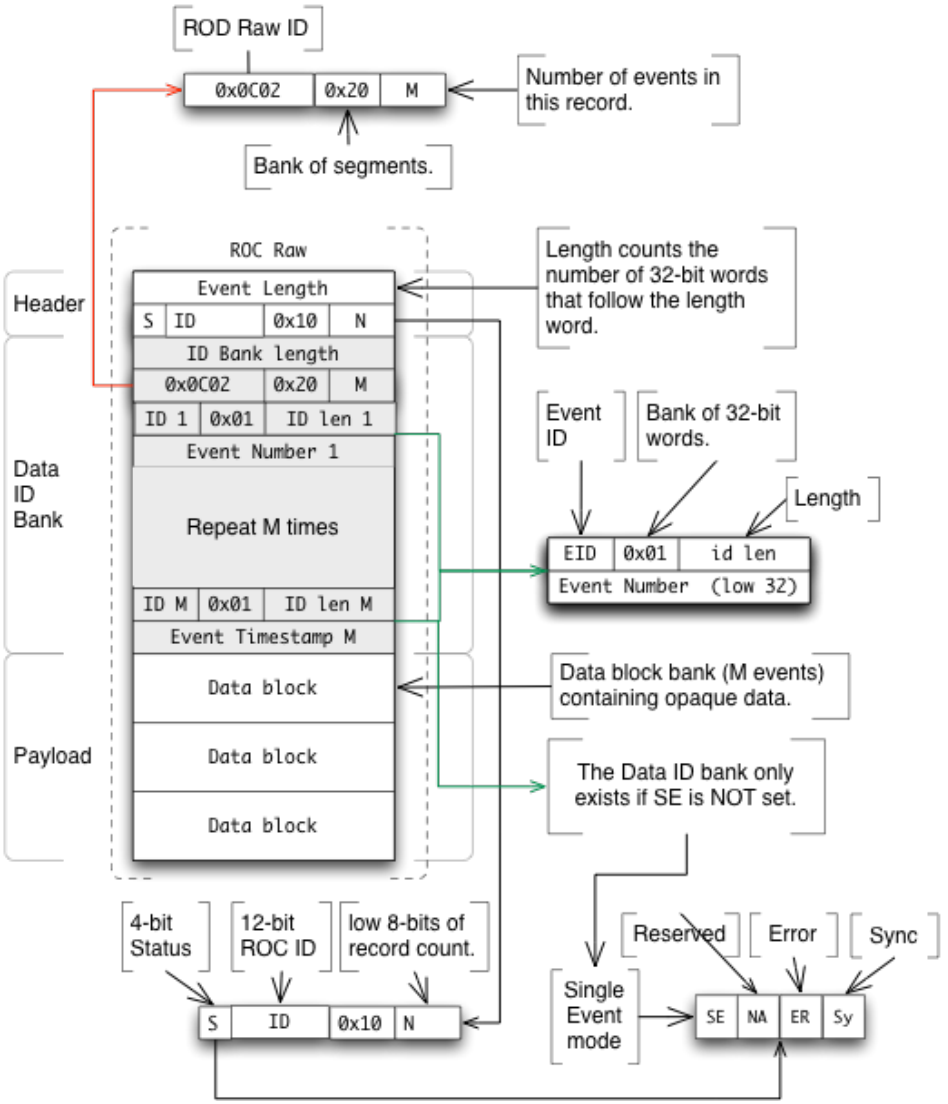
What does the data look like? TDC



Data format, a real example

- Bank - container for other data.
- Each bank starts with a header.
  - Length.
  - Description of content.

- The outer header tells us this bank contains banks.
- The first bank is a list of trigger information for all the events in the block.
- The following “payload banks” contain blocks of raw data read from ADCs or TDCs.

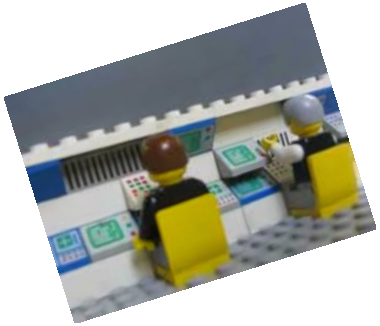


Putting together a big system

- The GLUEX detector and hardware trigger are capable of producing trigger and data rates that are a challenge for existing technology.
- The GLUEX collaboration came up with a design specification for the DAQ based on estimates of how the detector should perform.
  - Data spread over 50+ front end systems (crates).
    - 15 kByte events
  - Design luminosity of $5x10^7$ γ/s
    - Event rate of 200 kHz - data rate off detector = 3 GByte/s.
  - Design calls for a rate to storage = 300 MByte/s.
  - This reduction was to be achieved by partially analyzing events in software as they are taken and rejecting 90%, this is called a level 3 trigger.
  - Start with no Level 3 trigger for the initial running and compensate by using a beam of 10% of design luminosity (beam intensity) to give the same 300 MByte/s storage rate at 20 kHz event rate.
- How do we implement a data acquisition system that can do that?

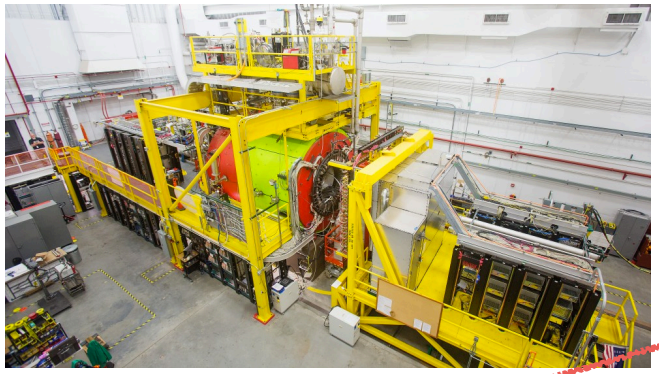
CODA

- CEBAF Online Data Acquisition (see coda.jlab.org)
  - Kit for implementing data acquisition systems.
  - Electronics
    - Custom boards like trigger, TDCs and ADCs.
    - Support for commercial hardware.
  - Software to :
    - Interface with electronics.
    - Readout boards and format data.
    - Move data.
    - Merge data streams.
    - Give users access to data for monitoring.
    - Write data to files.
    - Control the data acquisition system
- CODA is modular, solves big problems by splitting them into smaller ones.

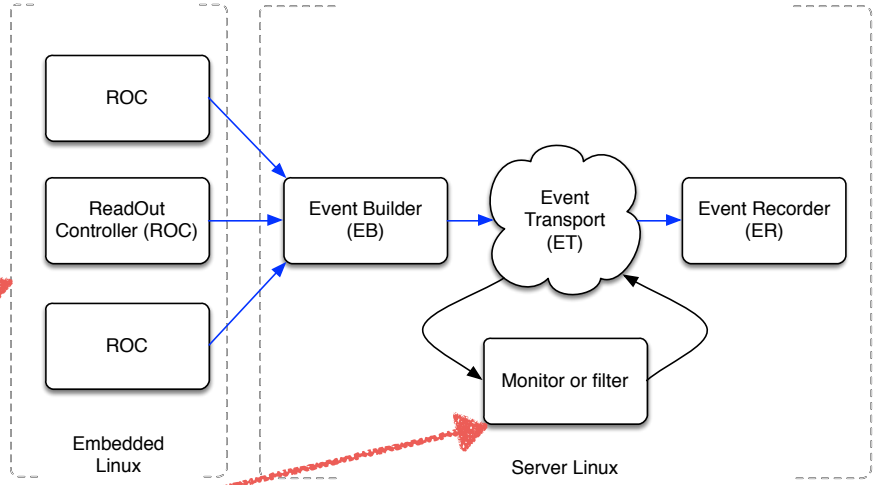


Network based data acquisition

- The detectors are spread over a volume of space.



- ROCS send data via a network.
- Bits and pieces of events arrive at different times from different places but need to be collected together with other data needed by the analysis.
- This is done by the Event Builder.
- A lot of pieces need to work together.
- The software parts represented by the rectangles are called “CODA components”

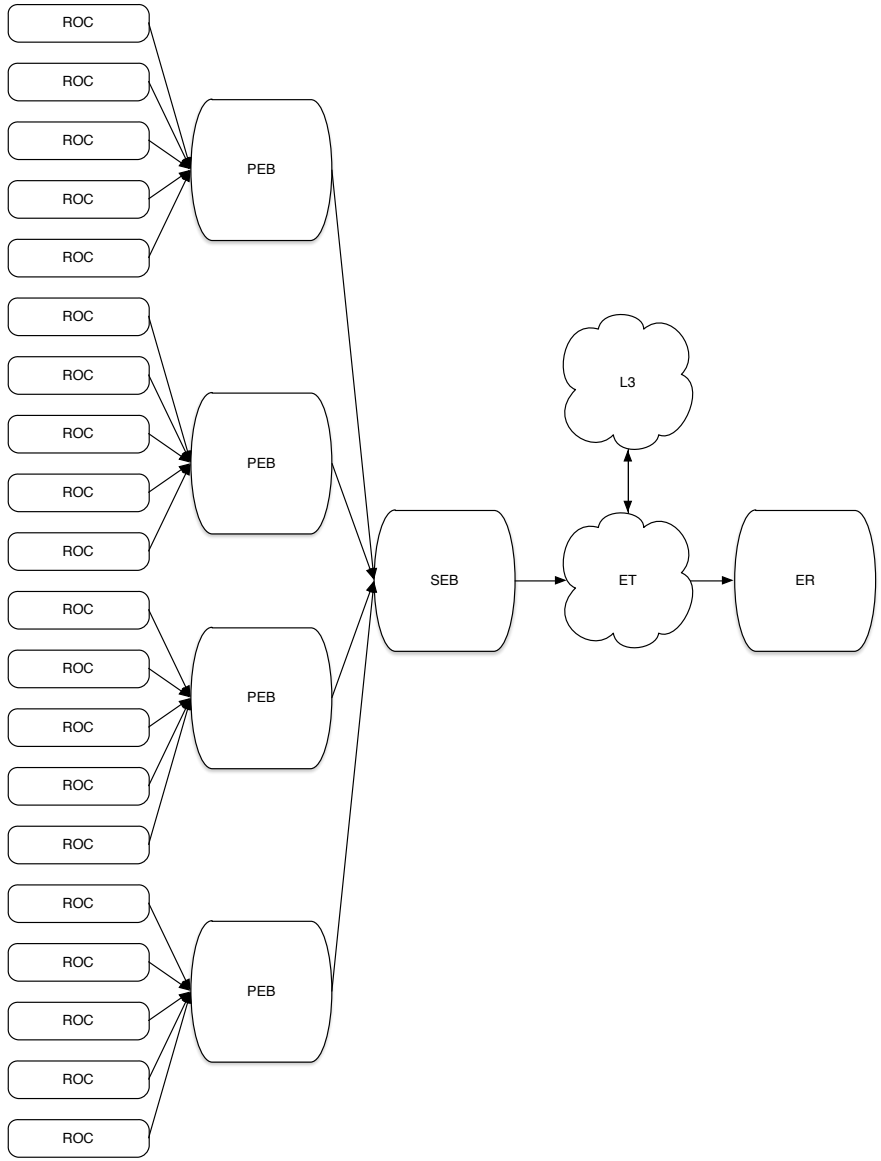


Event Building challenges

- The GLUEX design goal event rate was 200 kHz and there are 50 crates.
  - The Event Builder has 5 µS to:
    - Find all 50 parts of an event.
    - Decode the incoming data headers.
    - Check for errors.
    - Generate new headers for the assembled full event blocks.
    - Copy all of the data into place.
- The GLUEX goal data rate was 3 GByte/s.
  - 60 MByte/s average per each of the 50 incoming links.
  - 3 GByte/s through the EB, ET and ER.
  - Since data is copied several times the data rate inside machine running EB is several times 3 GByte/s.
- That could be a lot for one machine to handle.
  - Solution: multi stage parallel event builder.

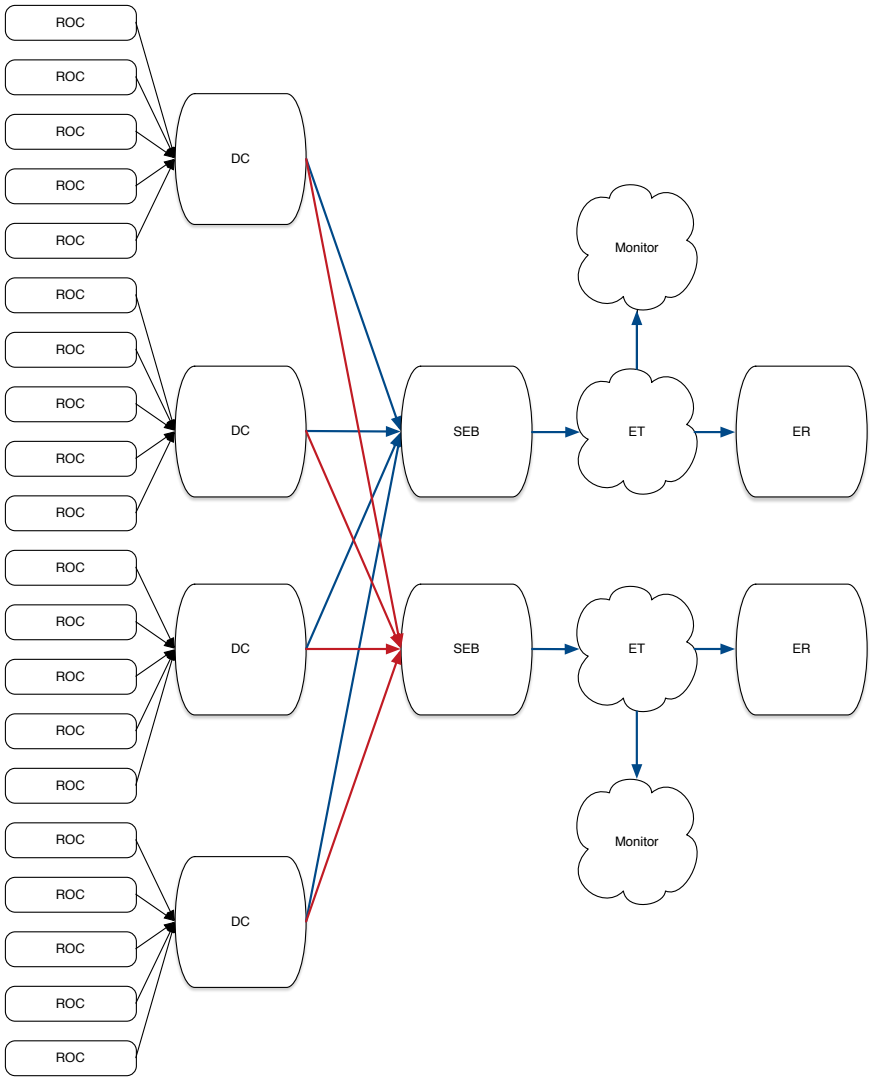
Staged Event Builder, 20 ROC example

- Four Primary Event Builders (PEB) are connected to five ROCs each by 1 Gbit/s Ethernet links.
- For each PEB this divides the formatting/ checking work load and the data throughput for by four.
- The four PEBs are connected to a single Secondary Event Builder (SEB) via 40 Gbit/s Infiniband links
- The SEB has to handle the full 3 GByte/s but only has four incoming streams to handle.
- If this is too much we can use two SEBs in parallel.
- The SEB outputs to a system called ET which distributes the events to the Level 3 trigger and any online monitors.
- A final program called the Event Recorder (ER) writes data files to disk, 100+ TB RAID.



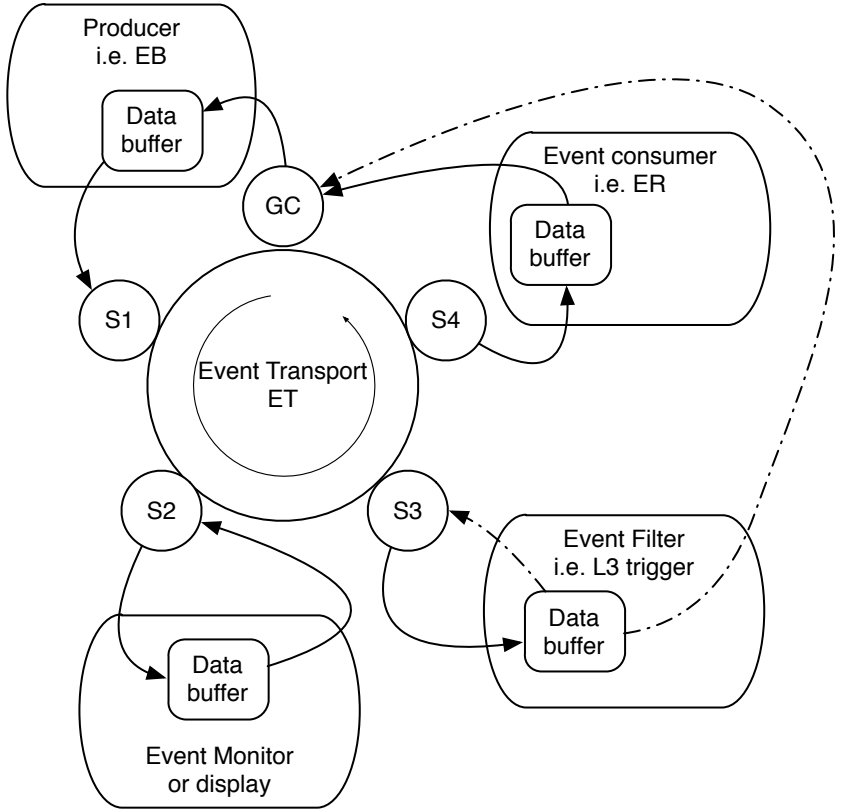
Multi stage and parallel

- Point pushing the full data stream through a single computer could be a bottleneck.
- Fortunately computers are evolving fast enough in CPU speed and IO bandwidth that as soon as we get close to this limit it moves away.
- In case we ever hit the limit CODA has the ability to split the data stream.
  - Faster event rates because more computers are event building.
  - Faster data rates because IO bandwidth goes up.



Event Transport, ET

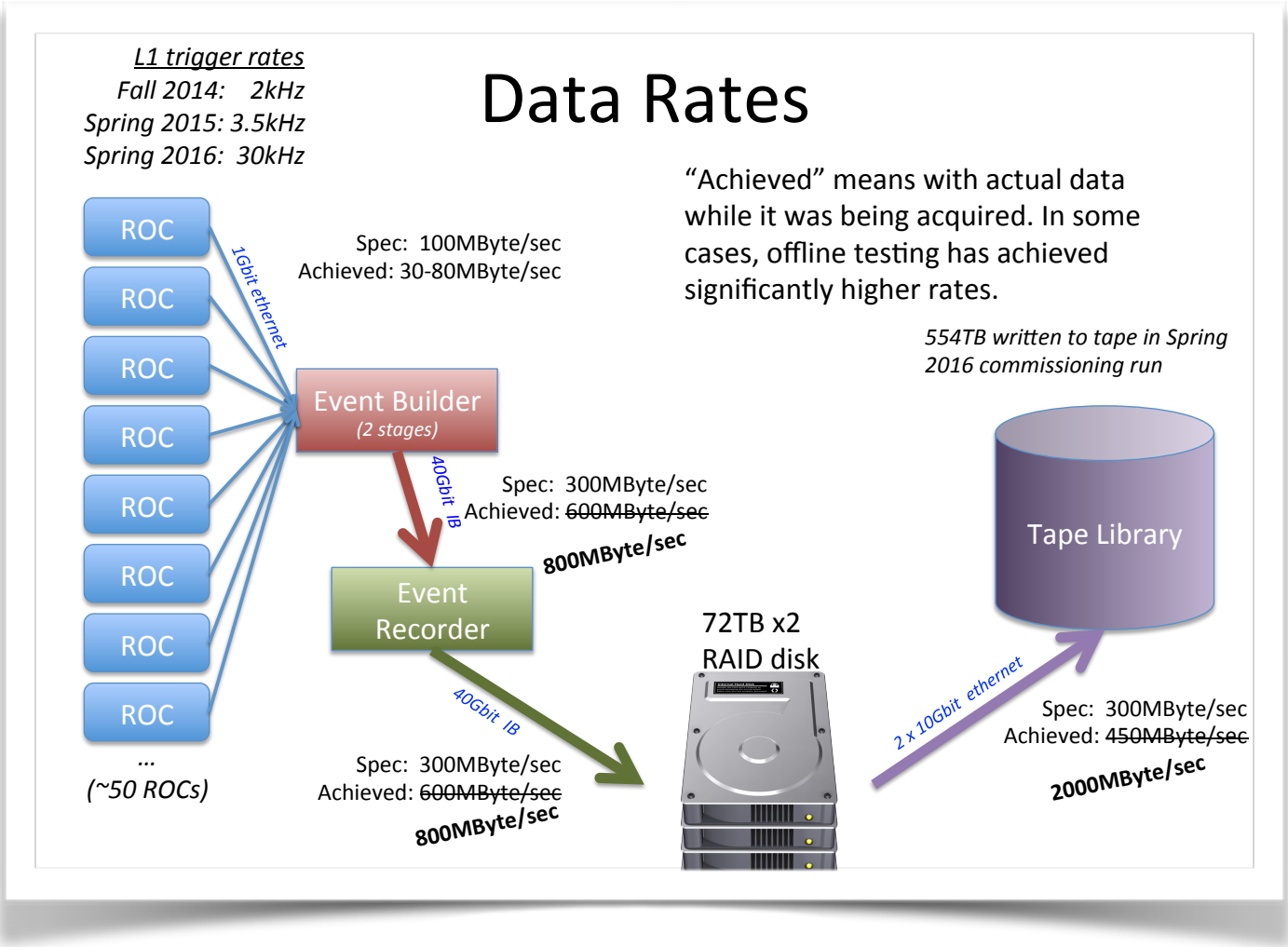
- Allocating and freeing buffers is time consuming.
- The ET system gives programs access to data via preallocated shared buffers.
- The system uses a railroad metaphor. Empty data buffers originate at Grand Central. They are filled by data producers and tagged to describe the content.
- The buffers “move” around a circular track and at each station the tag is checked to see if the buffer should stop at the station.
- An event monitor could set up station, S2, to take 1% of the events.
- An event filter could set up S3 to take all events. Discarded events are sent back to GC good ones move on.
- An event recorder takes all events and, after the data is written to a file sends the buffer back to GC.



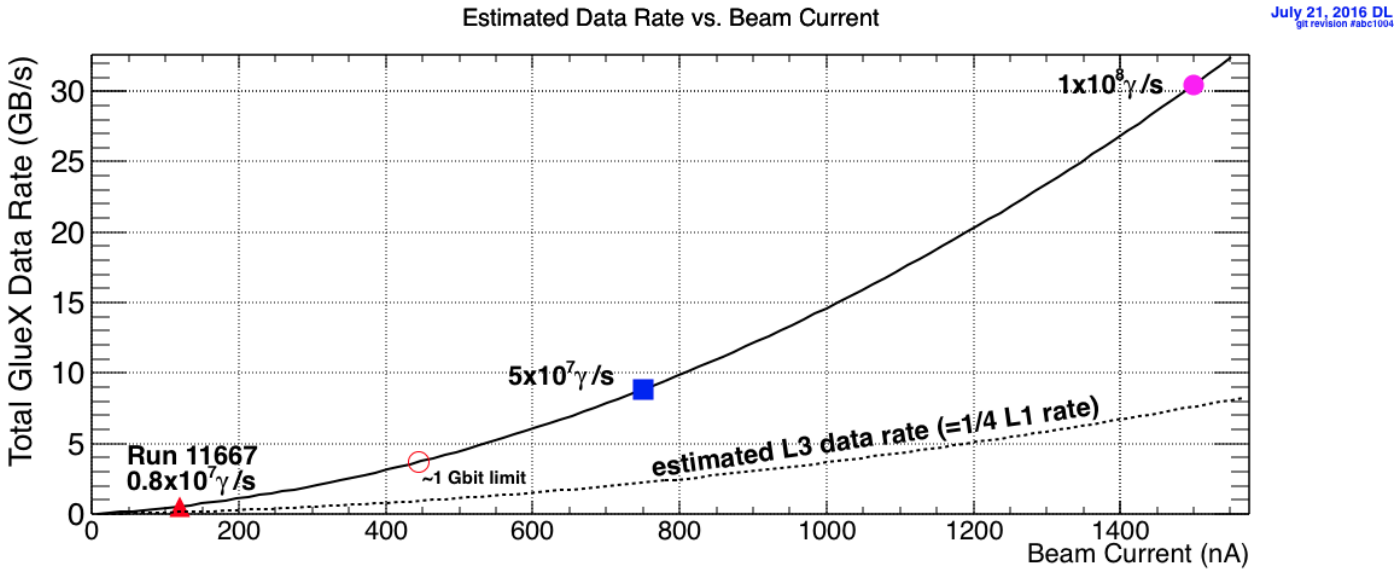
- 6 GeV experiments ran at tens of MB/s.
- 12 GeV experiments, hundreds of MB/s.
- Generate tens of petabytes per year.
- Tape is cheap but disk is faster.
- We write data to disk then copy from disk to tape later.
  - Tape speed only needs to handle average rate over a 24 hour period.
  - Tape drives and library robots are expensive and fragile. Writing to disk allows data taking to continue if the tape system breaks.
  - We typically have enough disk to hold three days of raw data.

All good on paper but…

- Spring 2016, low luminosity, 0.8x10^7 γ/s GLUEX ran at 800 Mbyte/s !!
  - Remember the goal for low intensity running was 300 Mbyte/s, 20 kHz



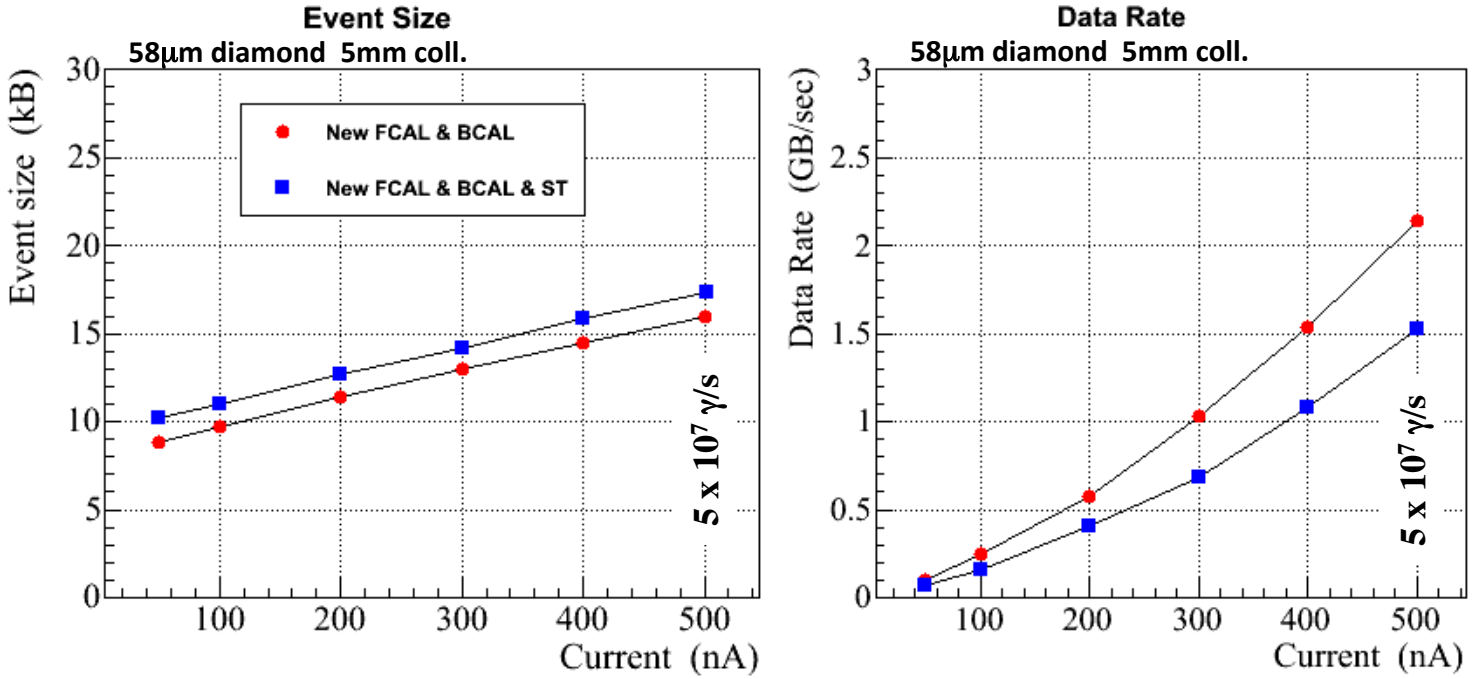
Trouble



- Tests during the Spring 2016 running showed that the event size grew roughly linearly with beam luminosity (current).
  - Caused by background hits.
    - data rate = event rate x event size
    - **If both event rate and size are proportional to luminosity data rate is proportional to the square of the luminosity. - A bad thing…**
- In 2019 GLUEX must run at 5×10^7 photons per second or 750 nA.
  - This would be 9 GByte/s but would be unachievable because at 3×10^7 γ/s (450 nA) some crates would reach the 1 Gbit/s limit of their ethernet connection.

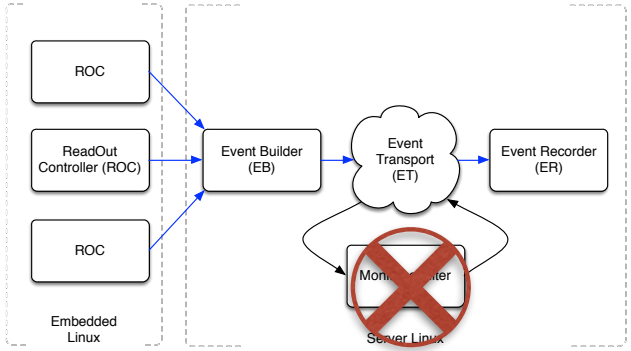
The Fix

- Fine tune the L1 trigger.
- Add a start counter to the trigger.
- Tune detector thresholds to reduce data per module.
- Change ADC and TDC firmware to remove unnecessary information
  - Event size ~17 kByte, data rate ~1.5 GByte/s



Goodbye to L3, again

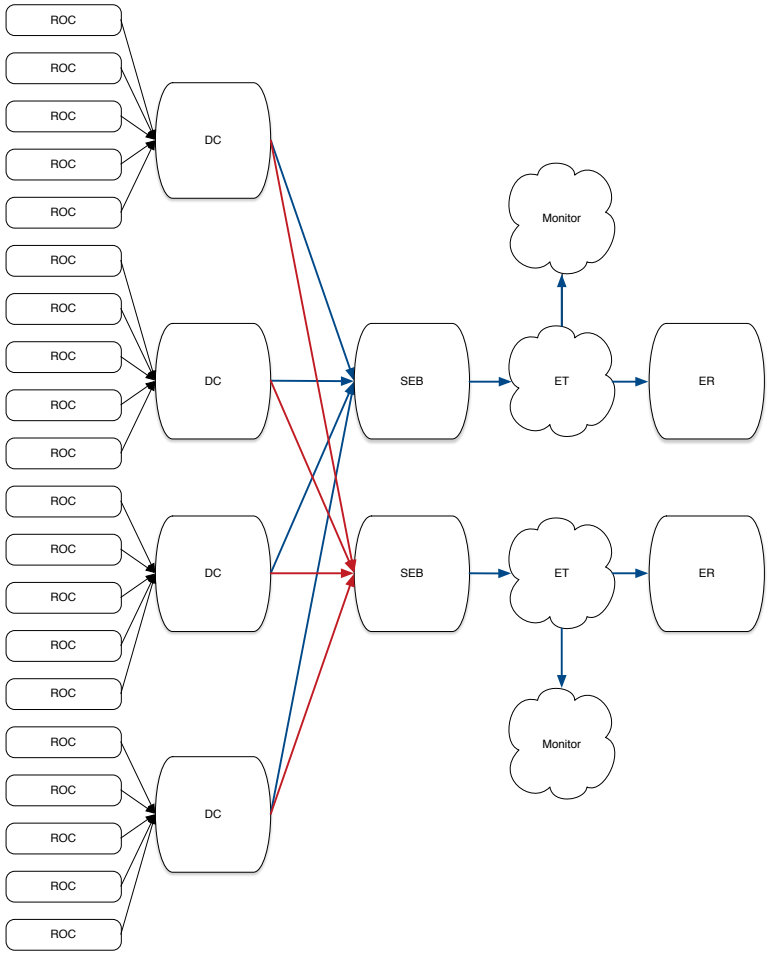
- To get the event and data rate down to a point where the data could be stored the GLUEX DAQ was designed with a L3 trigger to deal with high luminosity running.
  - Cut 3 GB/s data rate to 300 MB/s,
  - Cut 200 kHz event rate to 20 kHz.
- In the real world with a properly tuned L1 trigger and readout thresholds and no L3 we predict we will have:
  - Data rate 1.5 GB/s
  - Event rate ~90 kHz
- We have a DAQ that can handle these rates.
  - Why do we need a L3?
  - GLUEX dropped the L3 requirement earlier this year.



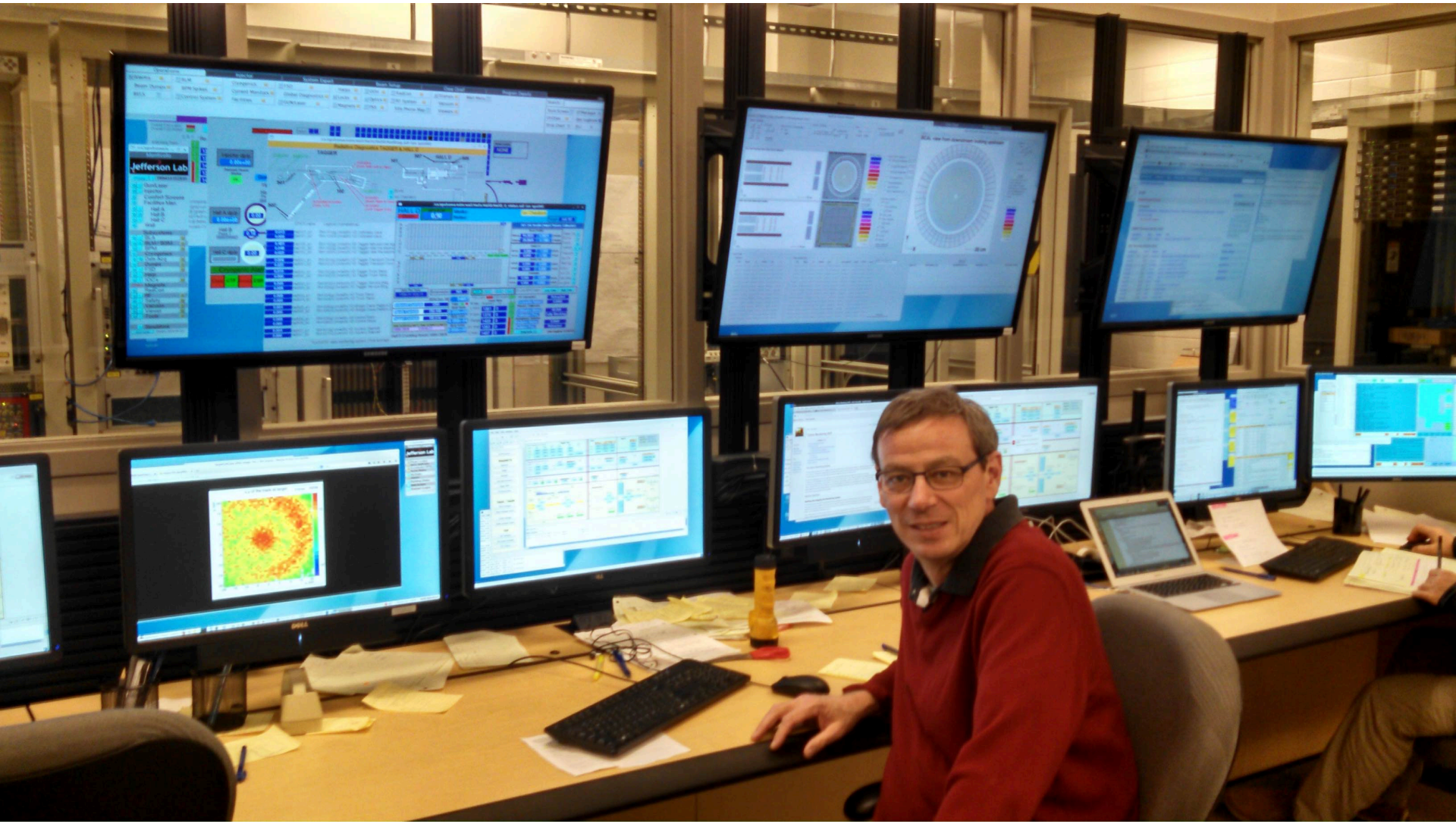
- "again" is in the title because this is exactly what happened to CLAS in the 6 GeV era.
  - When experiments are designed the only guide we have is the technology that we guess we will have. Often, in the real world, we are pleasantly surprised.

Spring 2017 GLUEX testing

- After thinking about it for 20 years, for the first time we tried out parallel event building with a real experiment.
- We only had one disk that was fast enough so we didn't write to a file - second disk is on order.
- Results:
  - Event rate 100 kHz
  - Data rate 2.7 Gbyte/s
  - 93% average live time.
- This is twice the data rate GLUEX currently claim to need.
- So, we think we are set for 2019.

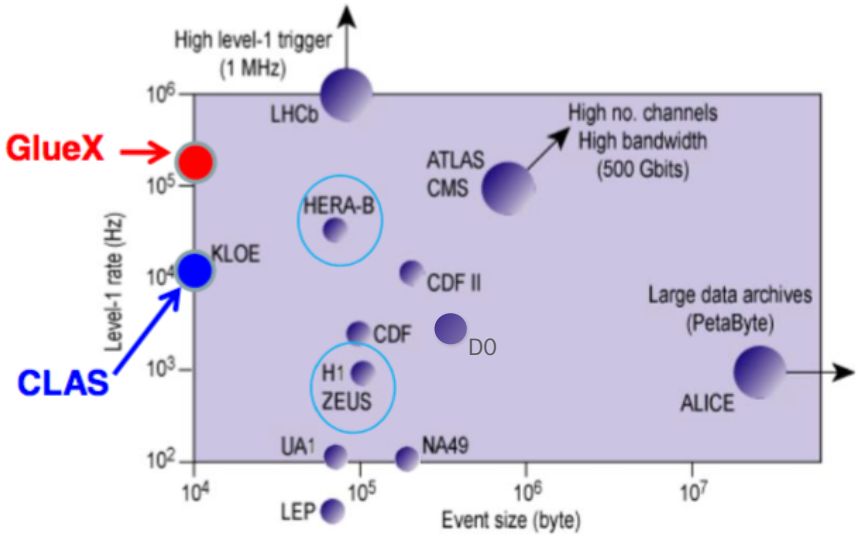


GLUEX - happy for now



The Future - Trends in experiments

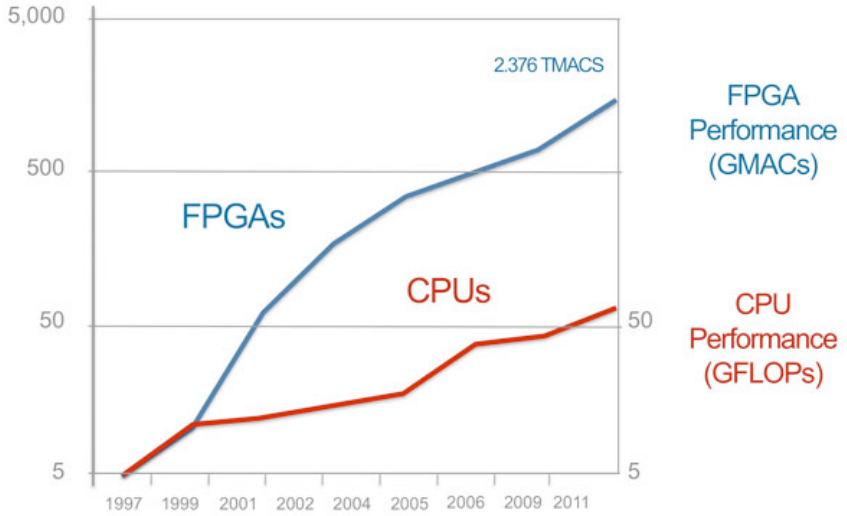
- Look at historical trigger and data rates.
- At JLab
  - mid 1990's CLAS, 2 kHz and 10-15 MB/s
  - mid 2000's - 20 kHz and 50 MB/s
  - mid 2010's
    - HPS, 50 kHz and 100 MB/s
  - 2019
    - GLUEX, 90 kHz, 1.5 GB/s to disk.

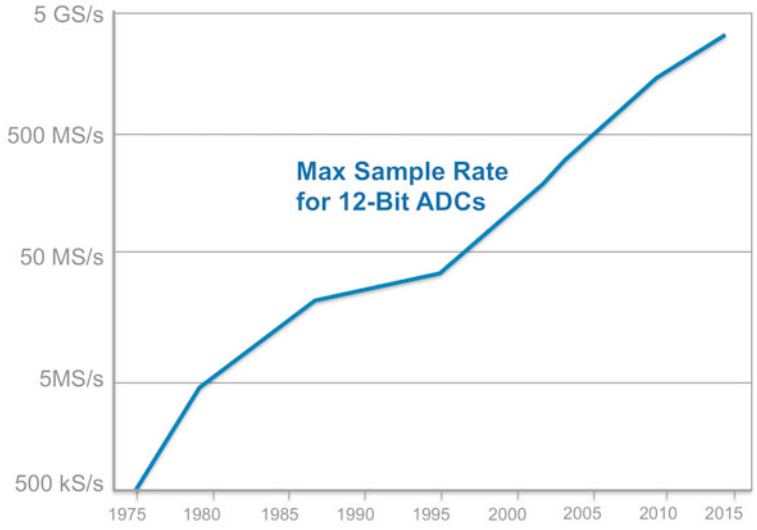


- FRIB - odd assortment of experiments with varying rates
  - LZ Dark matter search 1400 MB/s
  - GRETA 4000 channel gamma detector with 120 MB/s per channel. (2025 timescale)
- RHIC PHENIX 5kHz 600 MB/s
- RHIC STAR - Max rate 2.1 GB/s average 1.6 GB/s
- Looking at the historical trends the highest trigger rate experiments increase rate by a factor of 10 every 10 years.

Trends in trigger and electronics

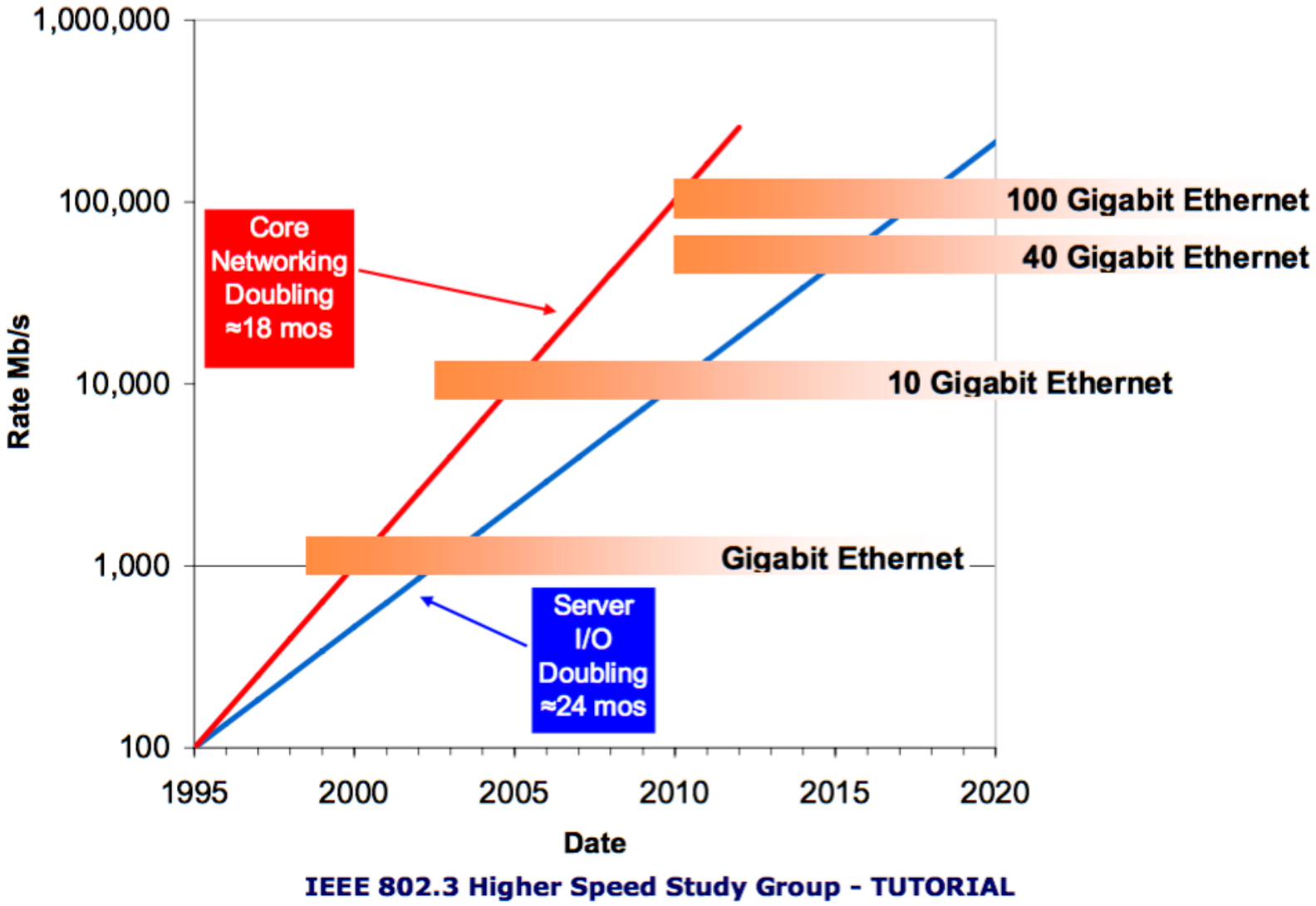
- FPGA performance is increasing faster than CPU performance. Why? There is a delay between when technology is developed and when it becomes affordable for use in custom electronics. So there is room for growth over the next ten years.





- Current trend is to push some functionality currently performed in software running on embedded processors into firmware on custom electronics. This will probably continue.

Trends in data transport



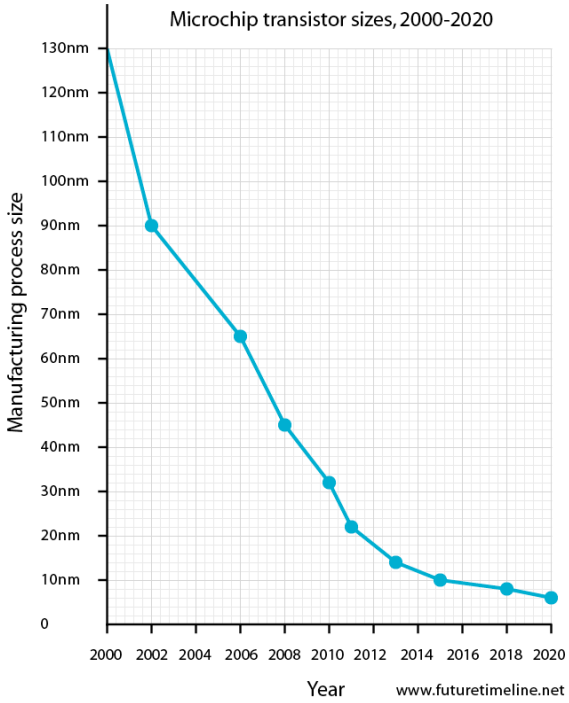
Challenges

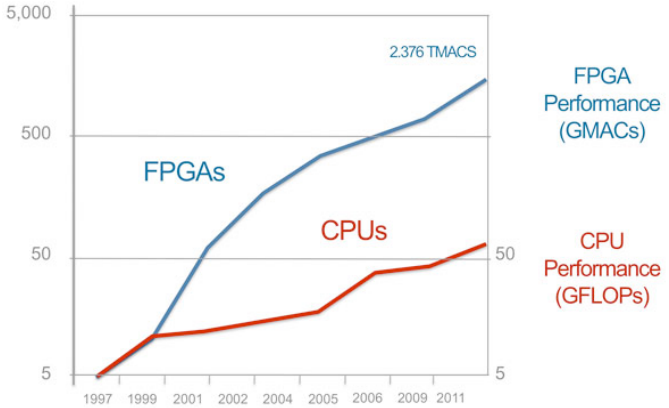
- The precision of the science depends on statistics which leads to :
  - Development of detectors that can handle high rates.
  - Improvements in trigger electronics - faster so can trigger at high rates.

- Beam time is expensive so data mining or taking generic datasets shared between experiments is becoming popular.
  - Loosen triggers to store as much as possible.
- Some experiments are limited by event-pileup, overlapping signals from different events, hard to untangle in firmware.
- Often the limiting factor in DAQ design is available technology vs budget, a constraint shared by all experiments at the various facilities.
  - It is not surprising that trigger and data rates follow an exponential trend given the "Moore's law" type exponential trends that technologies have been following.
  - **What matters is not when a technology appears but when it becomes affordable**. It takes time for a technology to become affordable enough for someone to use it in DAQ.

Challenges

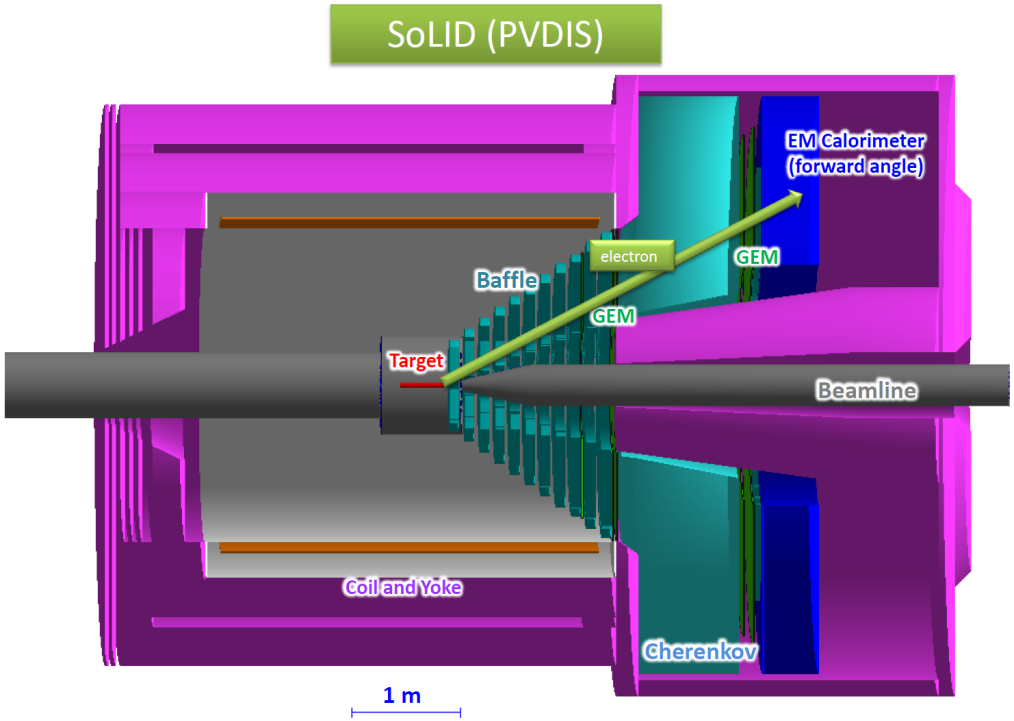
- Manufacturers are struggling shrink transistors.
  - How much further can Moore's law continue?
  - When does this trickle down affect the performance of other DAQ electronics?
- Use of mobile devices is driving tech in a direction that may not be helpful to NP DAQ, low power and compact rather than high performance.
- Are the rates for proposed experiments low because of low expectation?
  - Does the requirement of the experiment expand to take full advantage of the available technology?
  - If we come back in five years from now and look at experiments proposed for five years after that will we see a different picture than the one that we now see looking forward ten years? Probably yes.

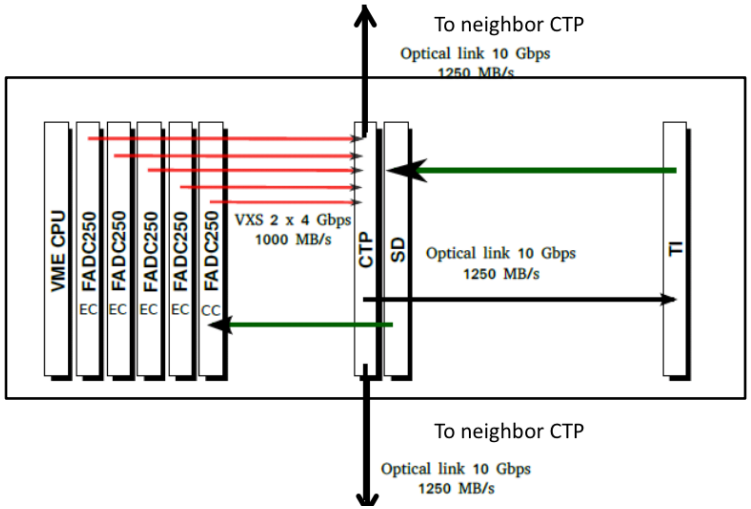




Future experiments, JLab - SoLID

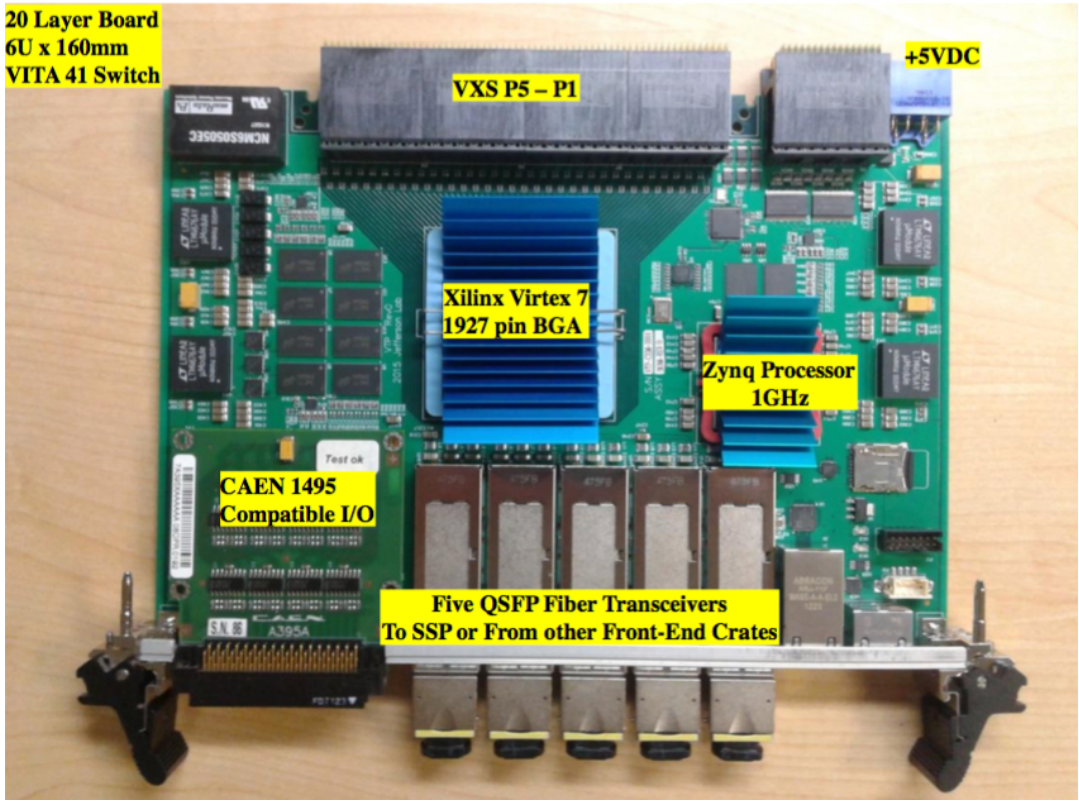
- SoLID is an experiment proposed for installation hall-A at JLab.
- The detector has two configurations. In the PVDIS configuration electrons are scattered of a fixed target at high luminosity.
- The detector is split into radially 30 sectors, the single track event topology allows 30 DAQ systems to be run in parallel at rates of up to 1 GByte/s each.
- L3 trigger reduces final rate to mass storage.





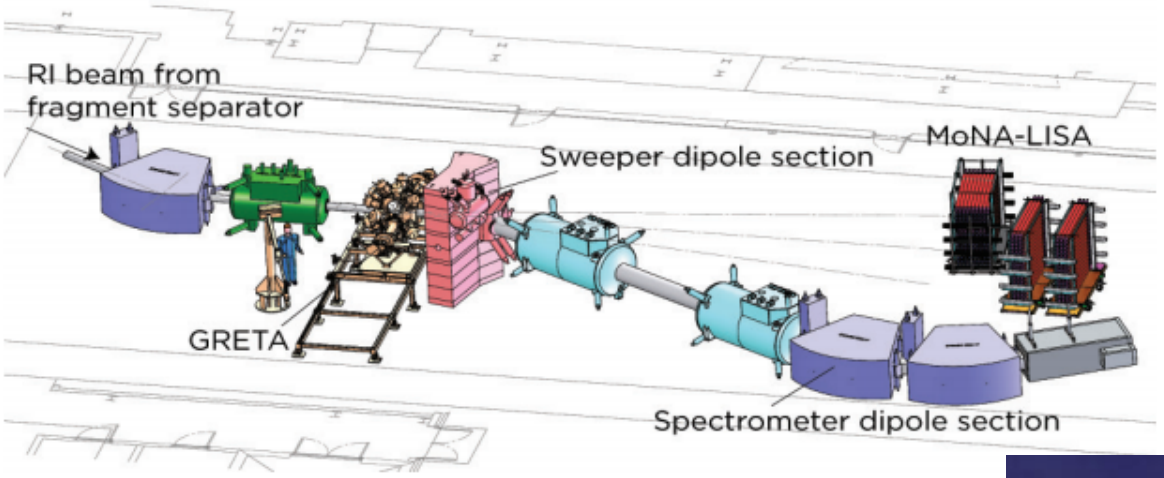
Hardware

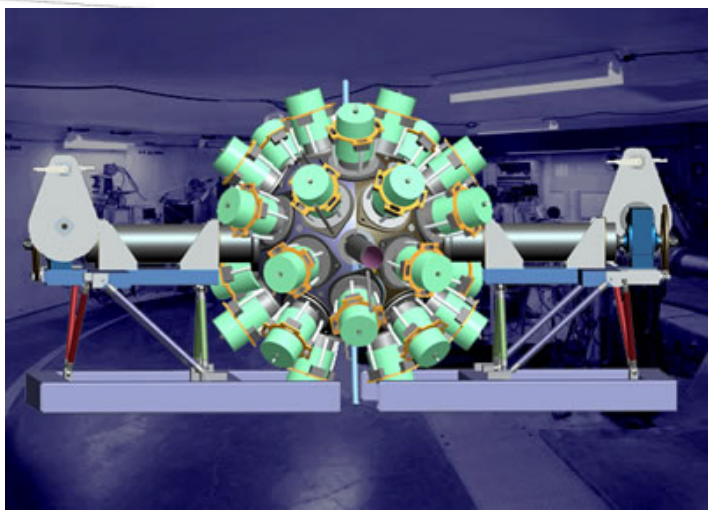
- The CTP and GTP trigger processors for GLUEX were very similar so merged the two designs to form the VTP, a general purpose board to manage serial data.
- The CODA 3 ROC has been compiled to run on the onboard processor (last week!).
- Need to program the Xilinx Vertex 7 chip to handle the full data flow rather than just the trigger data.



FRIB - GRETA

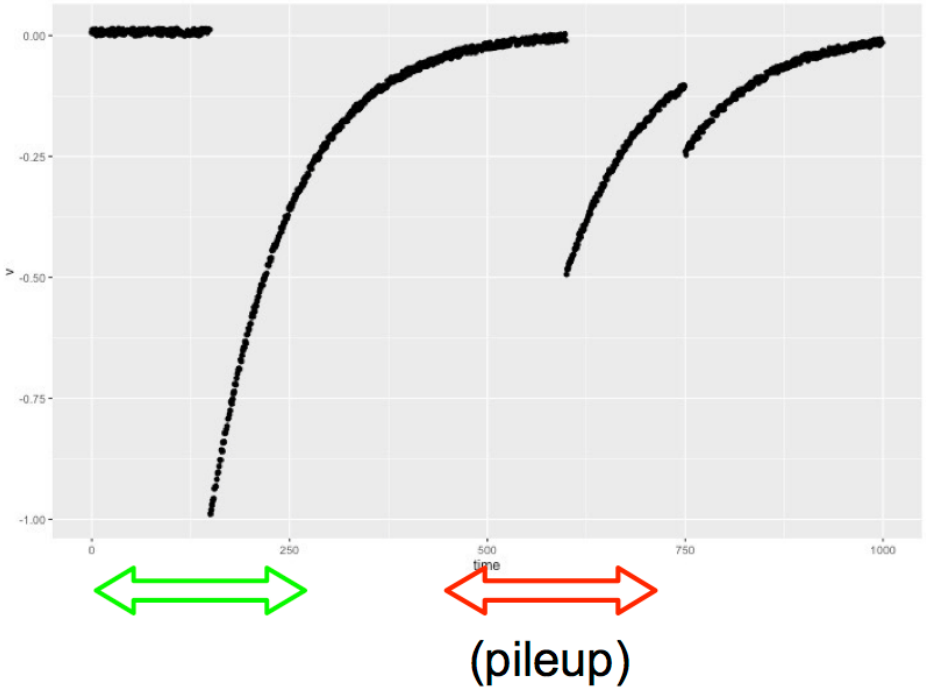
- Gamma ray spectrometer to be used at FRIB.
- Instrumented by 4000 x 100 MHz 16-bit ADCs.
- 2025 maximum I/O rate 100 MB/s per channel, 400 GB/s aggregate.



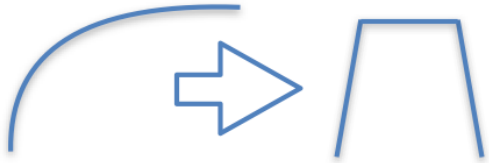


GRETA signal

- Flash ADCs measure periodically at a rate determined by a clock signal.
- The signal is integrated to determine the total energy deposited.
- This becomes complicated if signals overlap.
- Overlap becomes more likely if long integration times are needed.



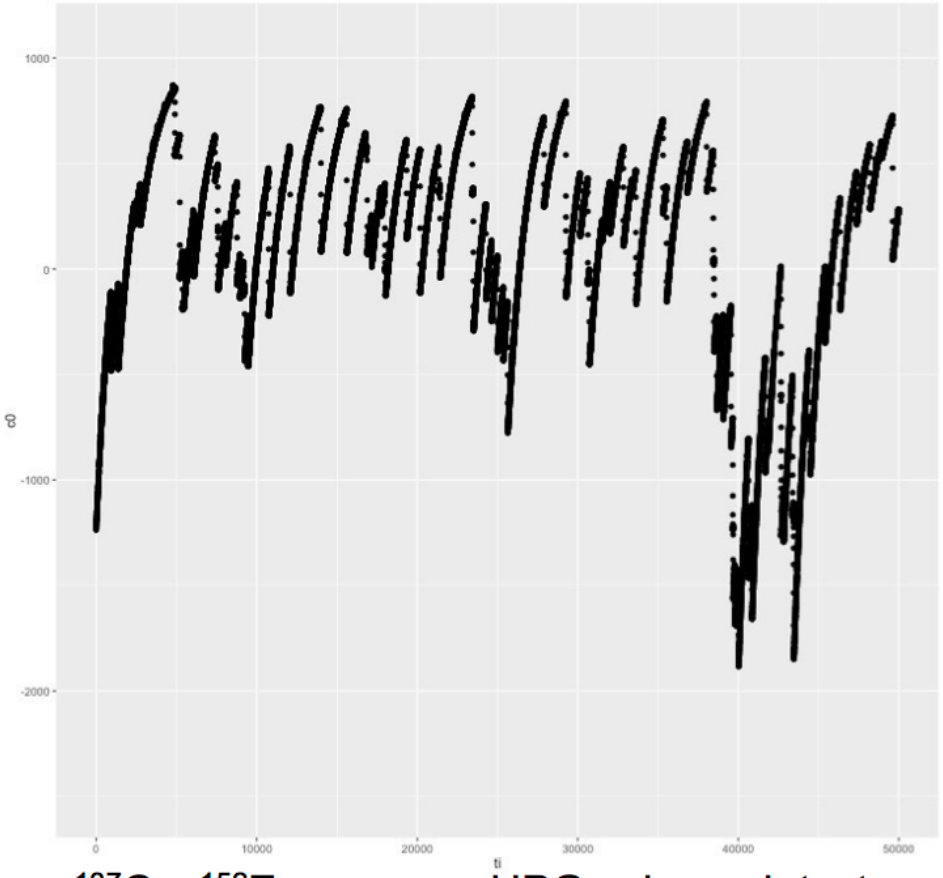
(pileup)



online trapezoidal filter implemented in FPGA memory pipeline (6μs flattop)

- HPGe detectors have very high intrinsic energy resolution (< 0.1 %)
- require long integration times

GRETA at 50+ kHz



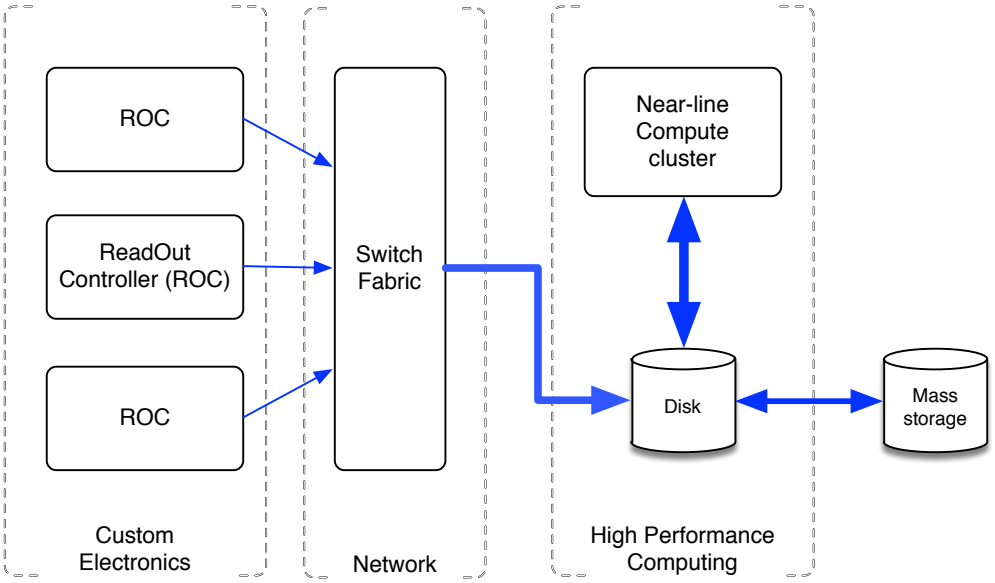
^{137}Cs, ^{152}Eu sources; HPGe clover detector

- Most signals overlap
- Erratic baseline - accounting for electronic response and history is important
- Difficult to maintain 0.1% energy resolution for sizable fraction of events - throughput losses

? - 100 kHz, .. more ..

Solution in ten years?

- Can't escape some sort of crate to put the electronics in - MicroTCA to replace VME?
- Stream the data through a network directly to temporary storage.
- High performance compute system processes the data online implementing a software trigger.
  - Several different triggers in parallel?
- Data surviving trigger or output from online processing migrates to long term storage freeing space for raw data.
- Much simpler architecture - more stable DAQ - but needs affordable versions of :
  - Reliable high performance network accessible storage.
  - High bandwidth network.
  - Terra scale computing.



Summary

- Data acquisition is constantly challenging.
  - Technology changes all the time.
  - Physicists think up experiments with tougher requirements.
  - The boundary between hardware and software is fluid and depends on what is available when a system is implemented.
  - There is always some R&D time to discover new algorithms and techniques.

  - To be honest, we do it because it's fun and we get to play with all of the cool toys!

Expect the unexpected

Engineer: I told the Captain I'd have this analysis done in an hour.

Scotty: How long will it really take?

Engineer: An hour!

Scotty: Och, you didn't tell him how long it would *really* take, did ya?

Engineer: Well, of course I did.

Scotty: Och, laddie. You've got a lot to learn if you want people to think of you as a miracle worker.