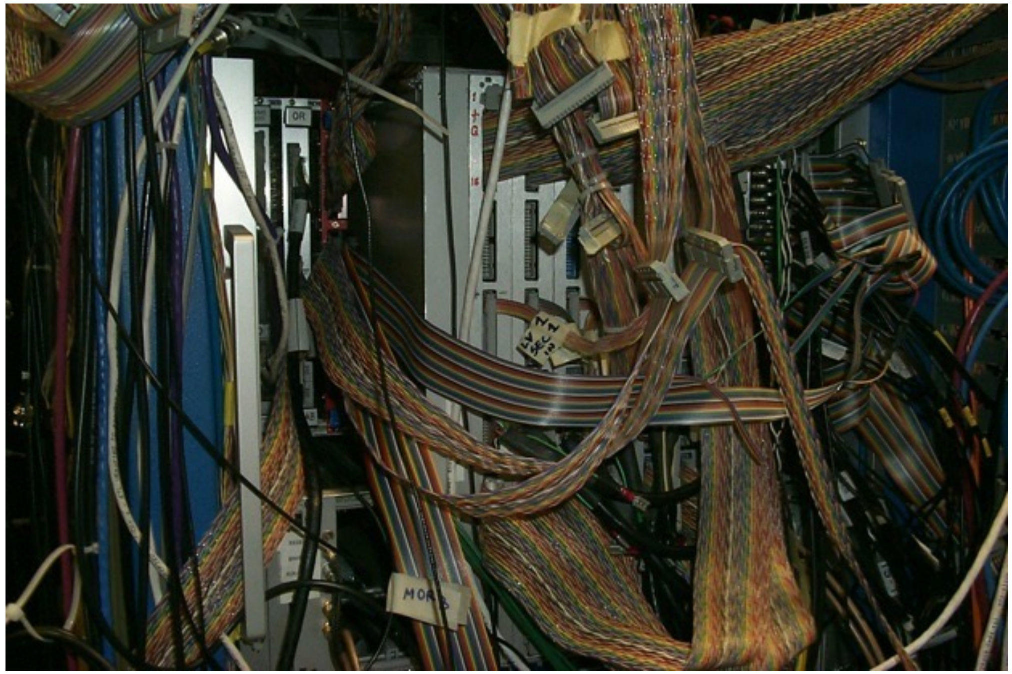
Data Acquisition at JLab (TJNAF)



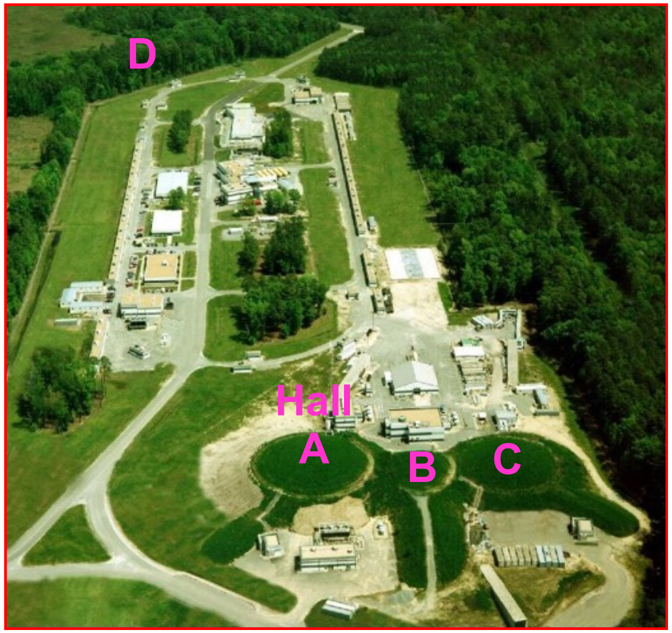
Graham Heyes

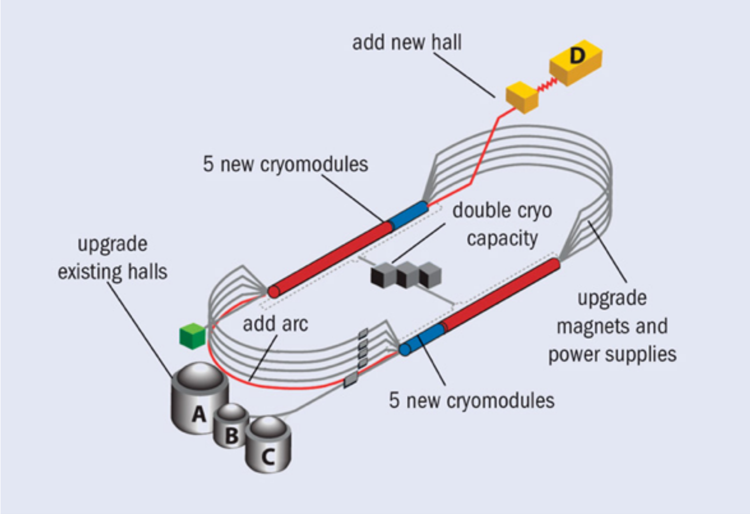
Data Acquisition Support

Experimental Nuclear Physics

Jefferson Lab

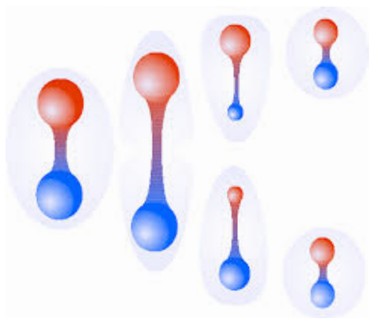
- The principle goal of Jefferson lab is nuclear physics research using the CEBAF electron accelerator.
  - Two superconducting LINACs with recirculating arcs, CEBAF a linear accelerator like SLAC but folded up on itself so that the beam passes multiple times through the two linear accelerators.
  - Simultaneous beam to four halls containing experiments.
    - A, B, C electron beam.
    - D photon beam.
- Each hall has equipment designed to study different, but complementary, aspects of matter in the nucleus.

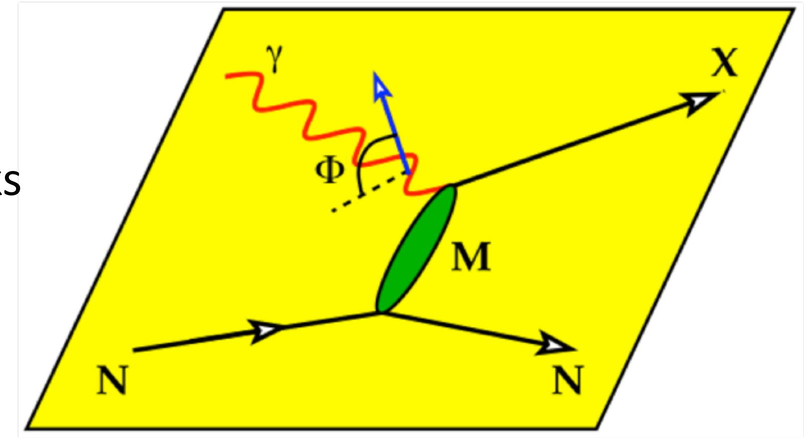




Physics

- In this document I will use the GlueX experiment in hall-D 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 in the nuclei of a liquid target with high energy photons to stretch the glue between the quarks until it snaps and mesons are produced
- Capture many photon interactions and extract physics 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. Since the electron is moving with high momentum and the target is stationary these particles are found primarily in the direction that the electron was traveling

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

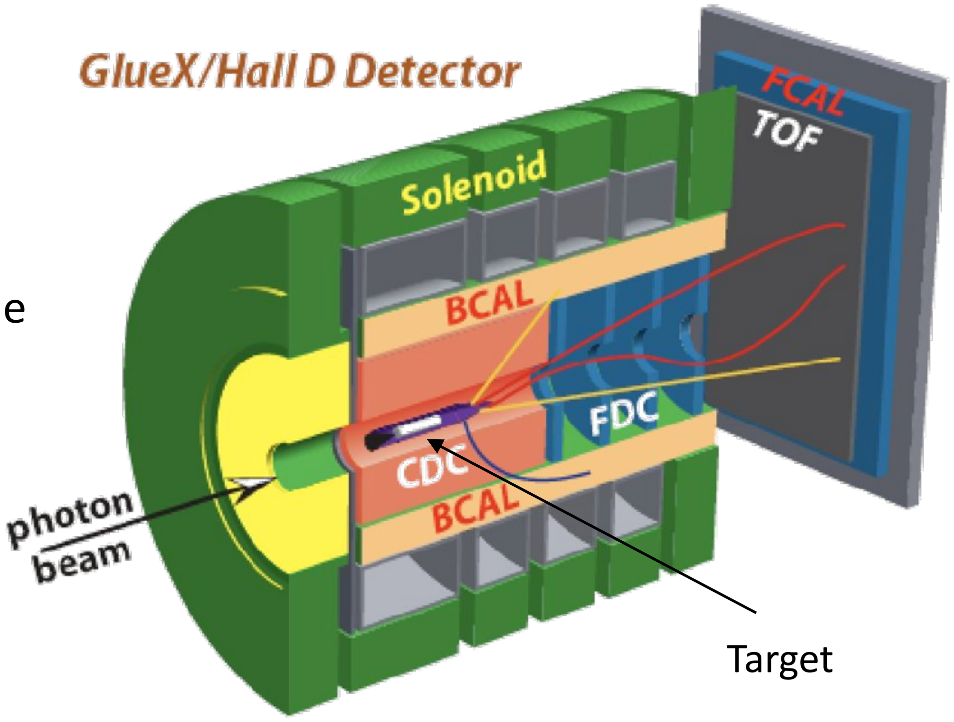
- There are three basic types of measurements are charge, time and count.

- 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.

The GlueX detector

- JLab is a fixed target lab, particles the beam strike a stationary target, as opposed to LHC at CERN or RHIC at BNL which study interactions from colliding beams.
- The GlueX target is a cryogenic target containing liquid hydrogen, deuterium or helium.
- Photons hit baryons in the target. Since the target is fixed particles generated in the interaction travel downstream.
- Various detectors measure properties of the particles.
- Detectors like the CDC and FDC can trace the track of a particle.
- A superconducting solenoid magnet causes tracks of charged particles to curve.
- Direction and radius of curve measures charge and momentum of particle.



Main parts of GLueX detector

Linearly polarized photons
Initial rate: 10^7 ° /s
tagged 8.4-9 GeV (to .1%)
Up to: 10^8 ° /s

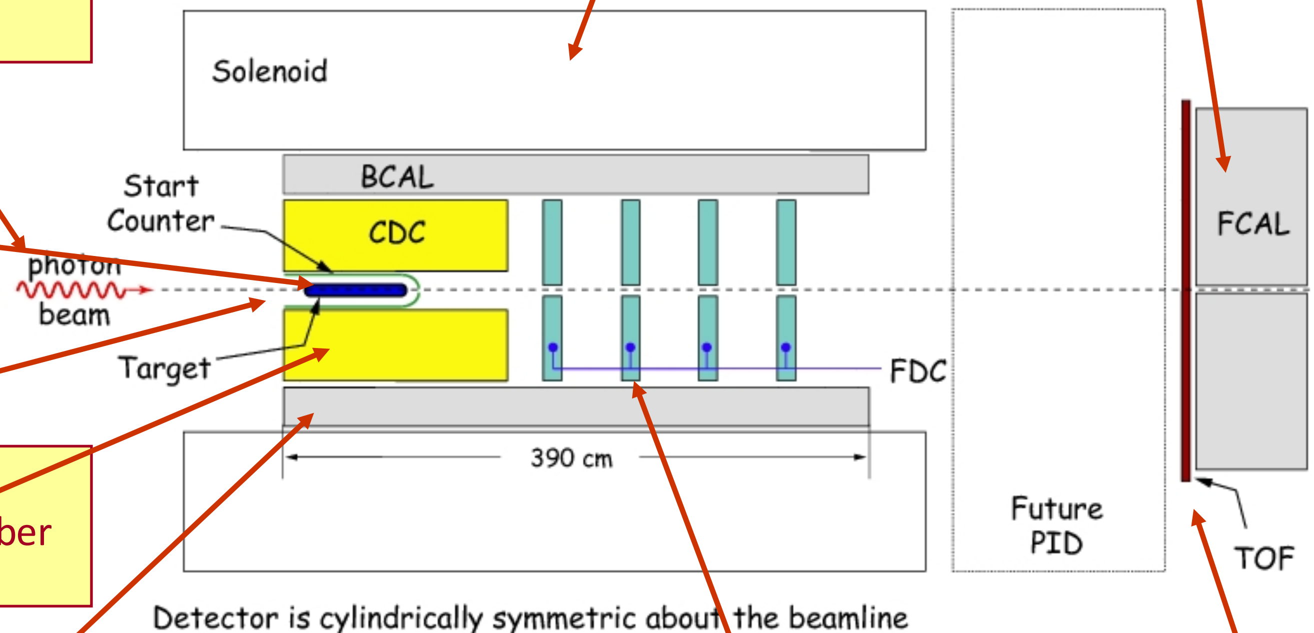
~2.25 T solenoid magnet (refurbished and updated LASS/MEGA magnet).

Pb Glass Calorimeter (glass from BNL E852)

30-cm long LH2 target

Scintillator start counter

Central straw tube drift chamber

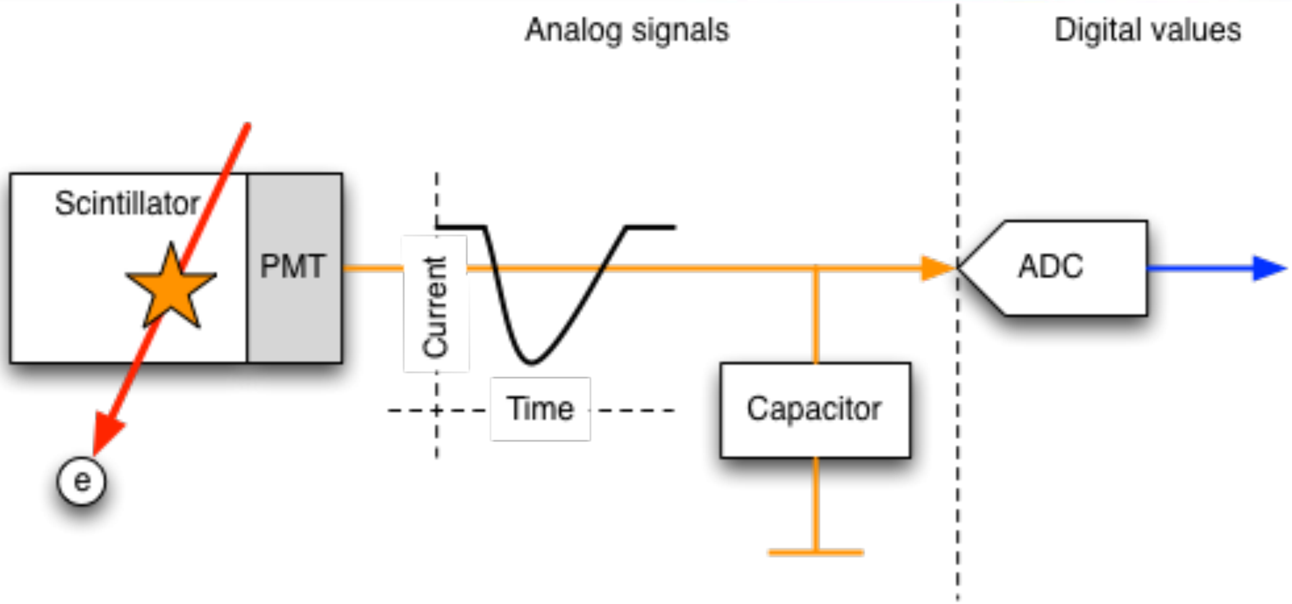


Pb scintillator sandwich calorimeter inside the solenoid. Also measure TOF of charged particles.

Planar cathode drift chambers

Plastic scintillator time-of-flight wall

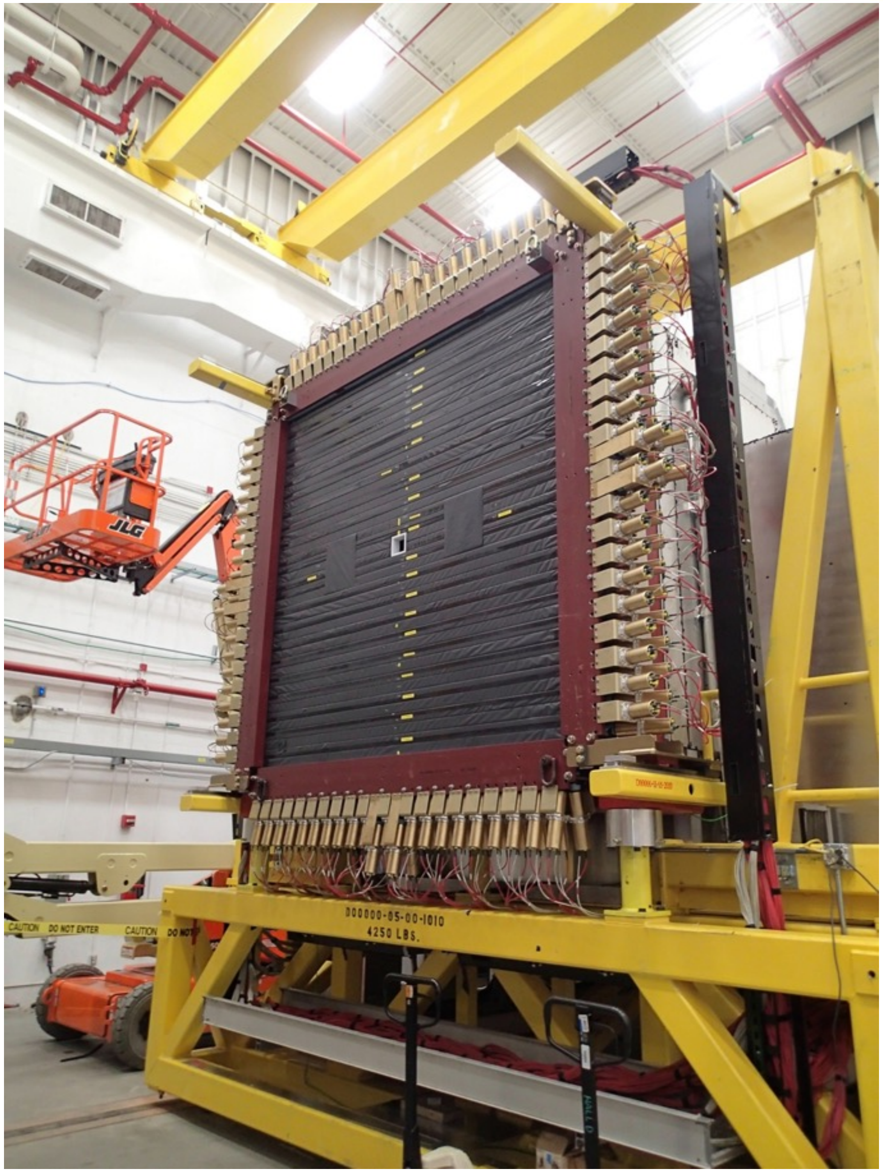
Detector example, a scintillator



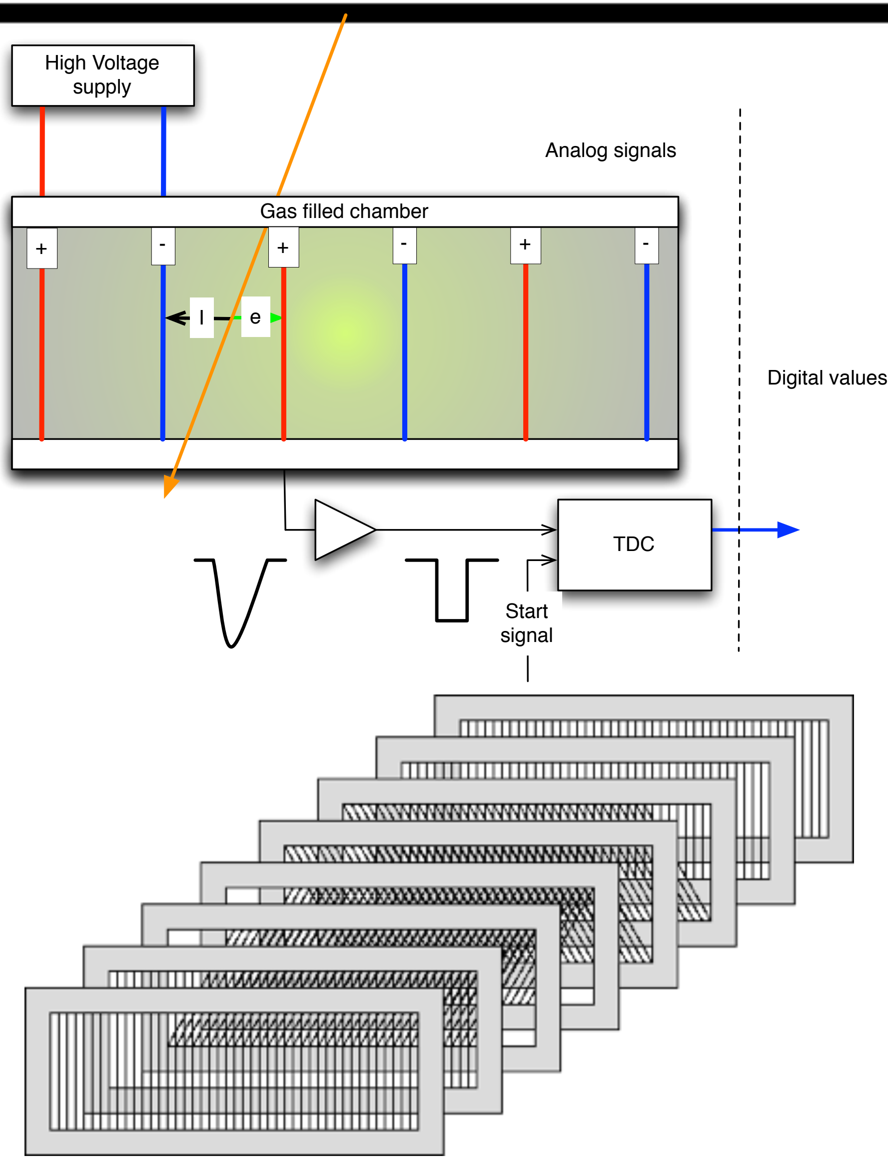
- A particle deposits energy in a scintillating material that converts it into light.
- A Photo Multiplier converts the light into a pulse of electricity.
- The pulse charges a capacitor.
- An ADC converts the voltage into a digital value.

Scintillators in GlueX

- TOF wall has measures energy and time of arrival of particles from the interaction.
  - Setup
    - 2 layers each with 42 scintillator bars (x –y)
    - 6cm x 2.54cm x 252cm
  - Readout:
    - double sided readout
    - XP2020 PMTs
  - Electronics:
    - energy measurement: 250 MHz FADCs (16ch)
    - timing measurement: CDF (16ch) & 62ps F1-TDC (32ch)
  - HV
    - 168 channels
- Start counter, a small detector made up from 40 small scintillators surrounds the target.
  - Setup
    - 40 scintillators
    - 10mm x 500mm, bended with 35^{o} towards beam
    - acceptance 3^{o} to 134^{o} σ_t= 0.5ns
  - Readout:
    - single sided in high magnetic field
    - SiPMs or Hamamatsu R5924-70
  - Electronics:
    - energy measurement: 250 MHz FADCs (16ch)
    - timing measurement: CFD (16ch) & 62ps F1-TDC (32ch)



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.
- GlueX uses a variant based on thin straws but the principle is the same.

CDC - Central Drift Chamber

Setup:

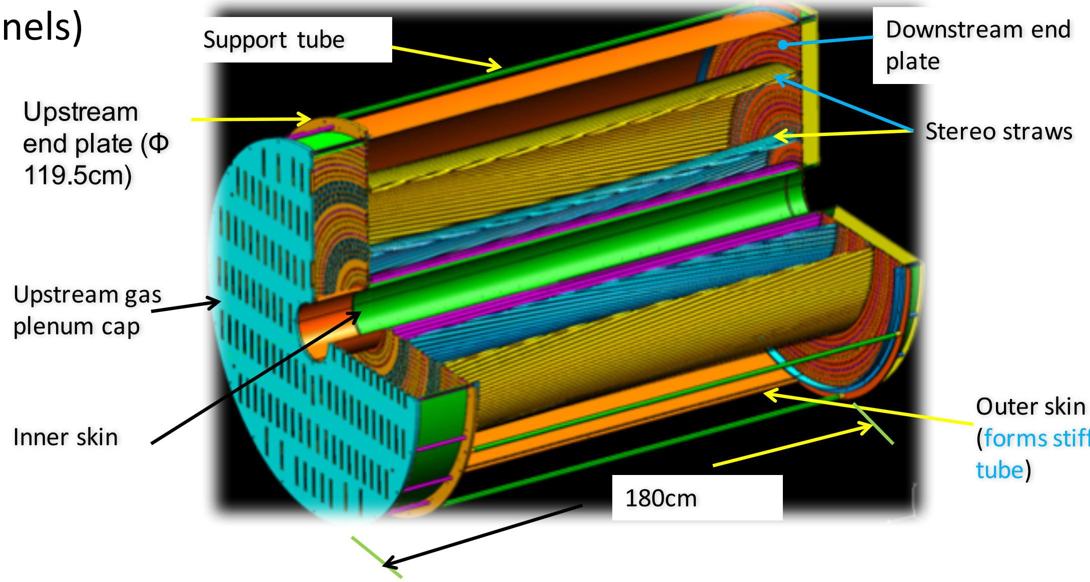
- straw tube tracker
- **3098 straws** (r: 0.8 cm; 100 μm Kapton 5 μm Al)
- radius: inner-10cm outer-58cm length-1.5m
- 4 layers +6°; 4 layers -6°; 16 radial layers

Readout / Electronics:

- preamp cards the same as for FDC based on ASIC
- energy/timing measurement: 125 MHz FADCs (72ch)

HV

- 24 straws / HV channel (130 HV channels)



Endplates, tubes and skin form stiff structure!

Forward Drift Chamber

Setup:

- cathode strip chamber
- 4 packages;
  - ground- cathode(24)-wire(24)-spacer(24)-cathode(24)
- 96 sense + 97 field wires & 216 cathode strips
  - total: **12672 channels**
- wires; u-v strips +/- 75° to wires
- diameter: 1.2m

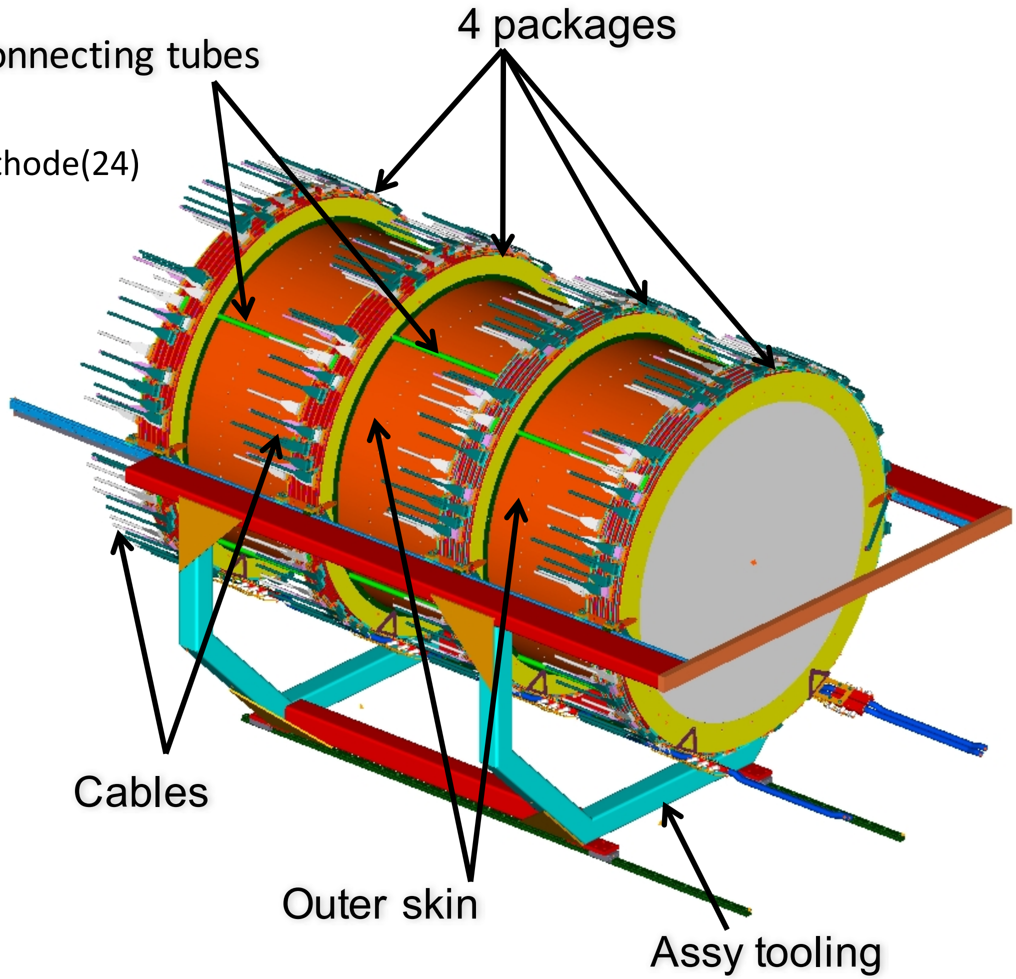
Readout / Electronics:

- Preamp. boards based on ASIC
- cathodes: 125 MHz FADCs (72 ch) 144 modules
- anodes: 125ps F1-TDC (48 ch) 48 modules

HV

- 384 channels

Note the huge channel count!!



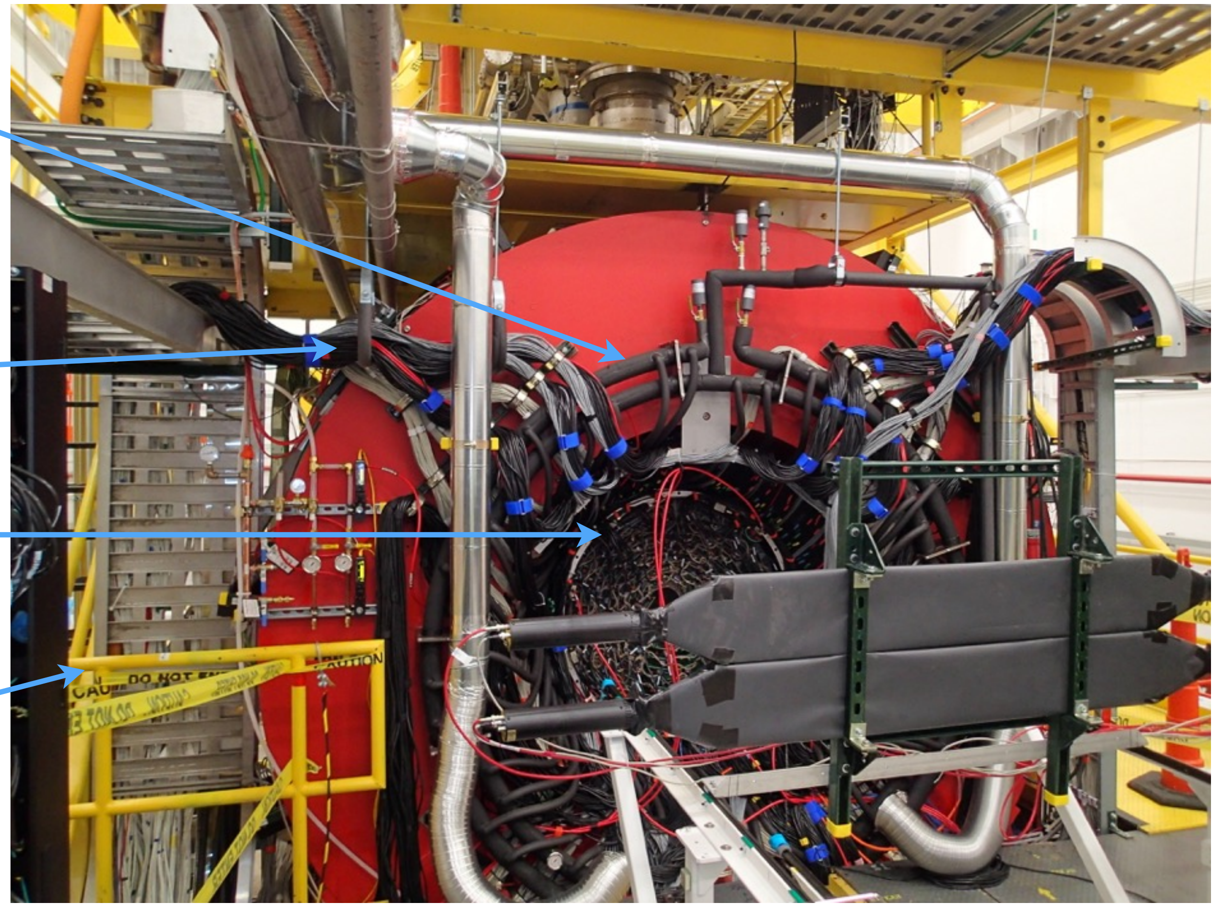
Picture of end of GlueX detector in hall-

Cooling for electronics

Cables

Drift Chambers

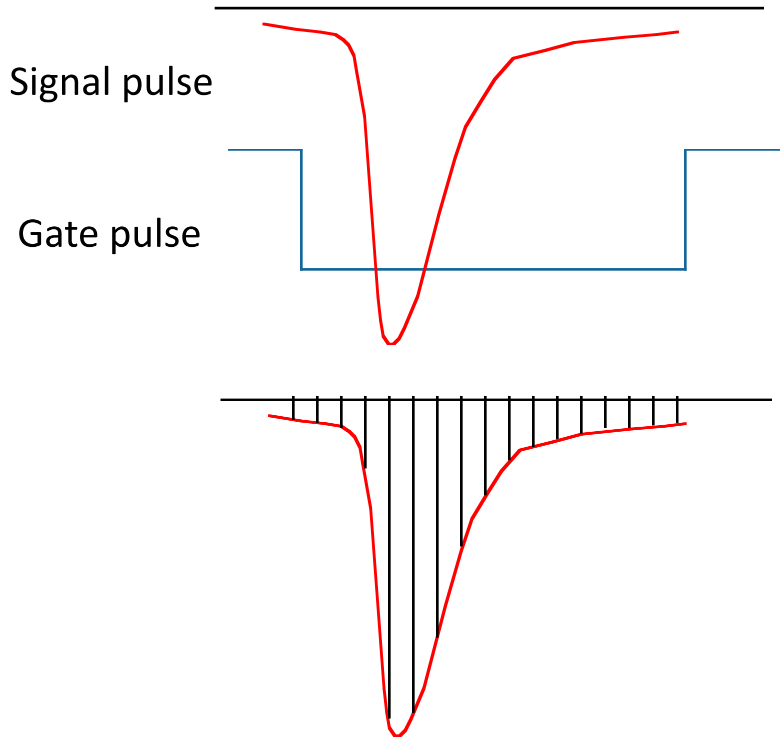
Scale of detector,
this rail is waist high



Sampling vs Integration

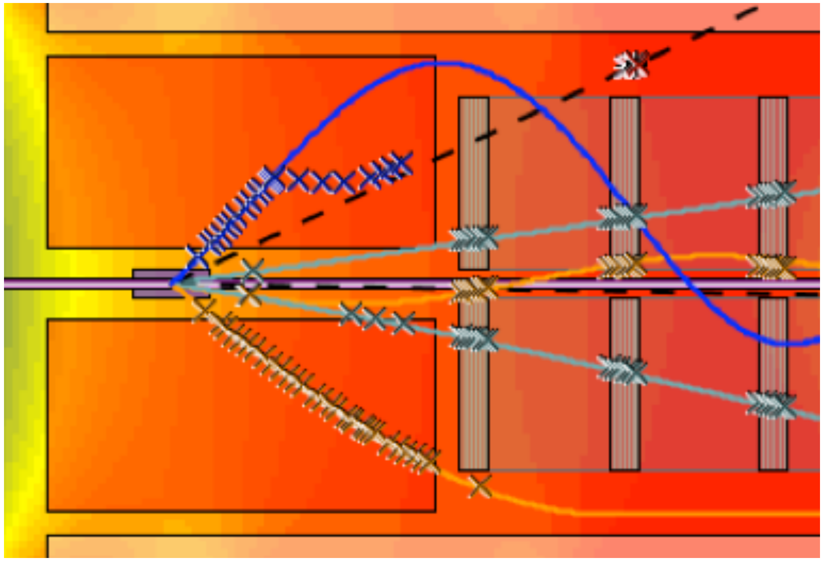
- A traditional "integrating" ADC takes 6 to 10 µsec to digitize a pulse. A gate pulse, generated by other detectors marks the region of interest to electronics electronics that integrates the charge from the signal pulse.
- This type of ADC generates a single measurement representing the charge sum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

- A Flash ADC samples continuously at a fixed rate.
- GlueX uses two types of fADC:
- a 16 input 12 bit 250 MHz ADC.
  - Samples every 4 nsec and generates ~10-15 measurements during the gate describe the pulse shape as well as charge.
- a 72 input 12 bit 125 MHz ADC for the CDC readout, large number of channels but lower sample rate requirement.



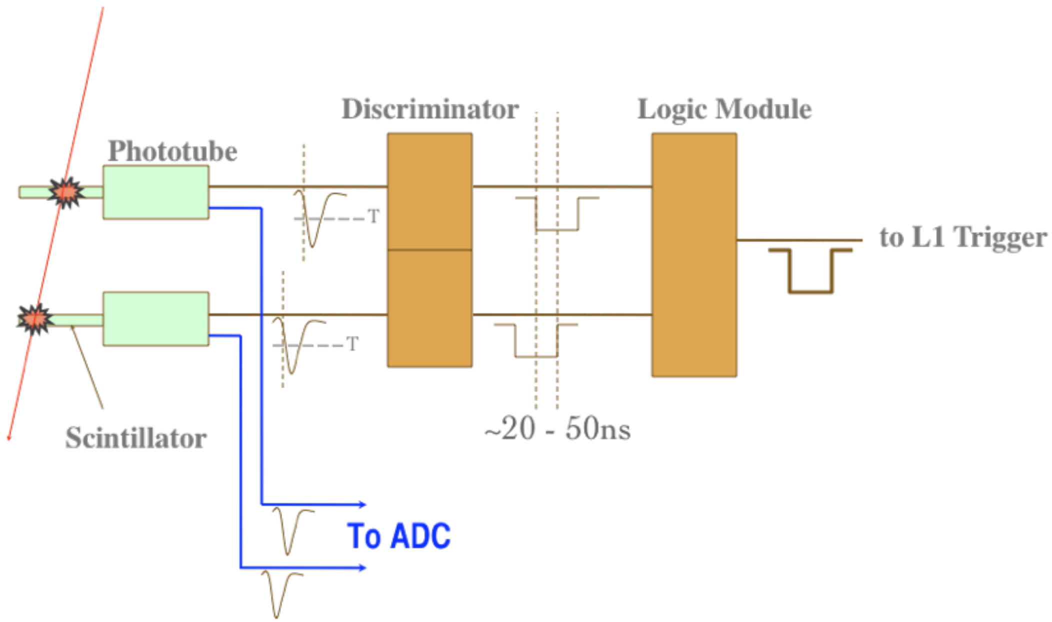
Properties of nuclear physics data

- 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.

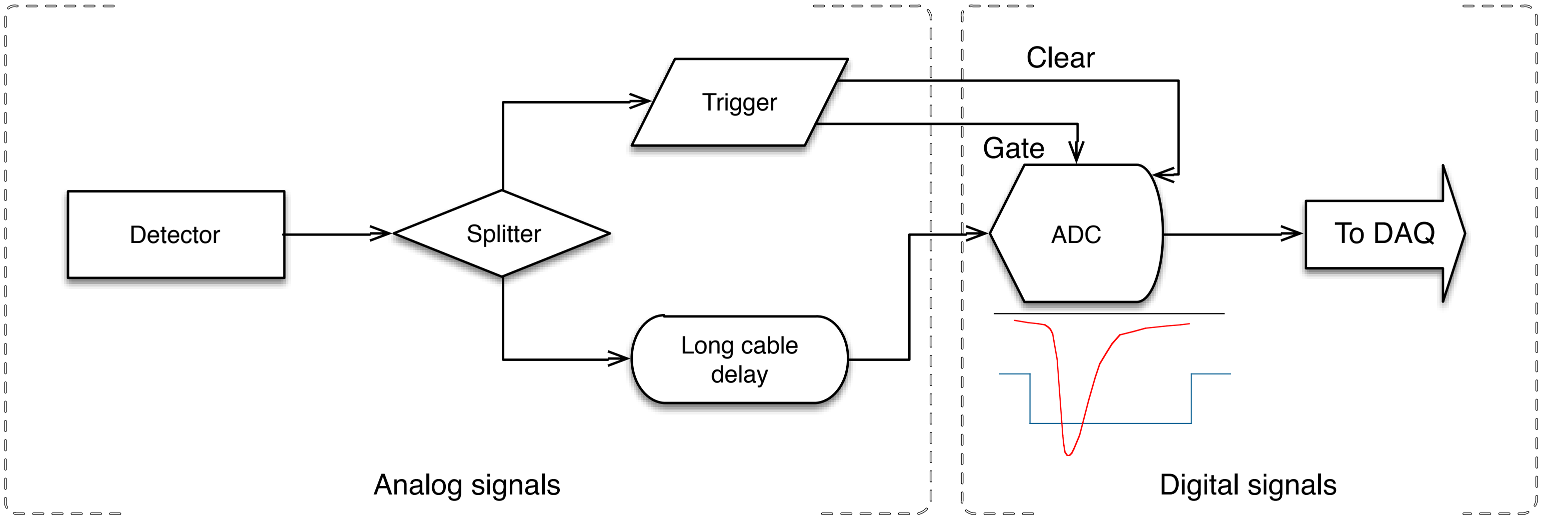


A Simple Trigger

- 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.
- In the simple example shown below a particle hits two scintillators.
  - We generate a digital pulse with width equal to the maximum possible time separation of the two pulses.
  - If these pulses overlap, called a coincidence, it is likely they came from the same particle so a trigger pulse is generated which causes the ADC to digitize.



An analog trigger

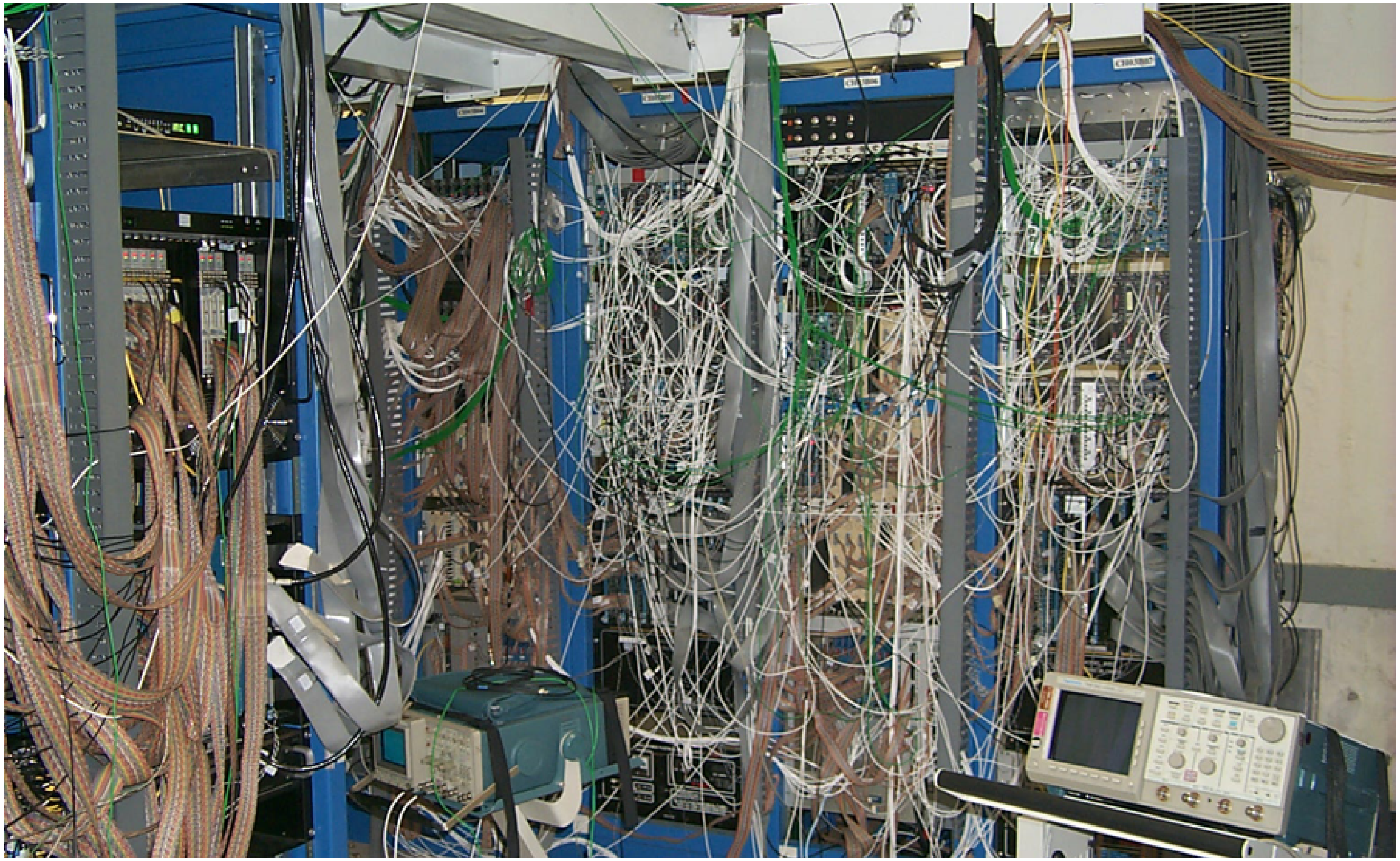


- It takes some time for the trigger logic to decide if a signal should be digitized.
- The analog signal must be delayed so that the gate and signal arrive at the ADC at the same time. Typical coax cables ~1 ns/ft so you could simply delay the signals using long cables.
  - Matching cable lengths is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

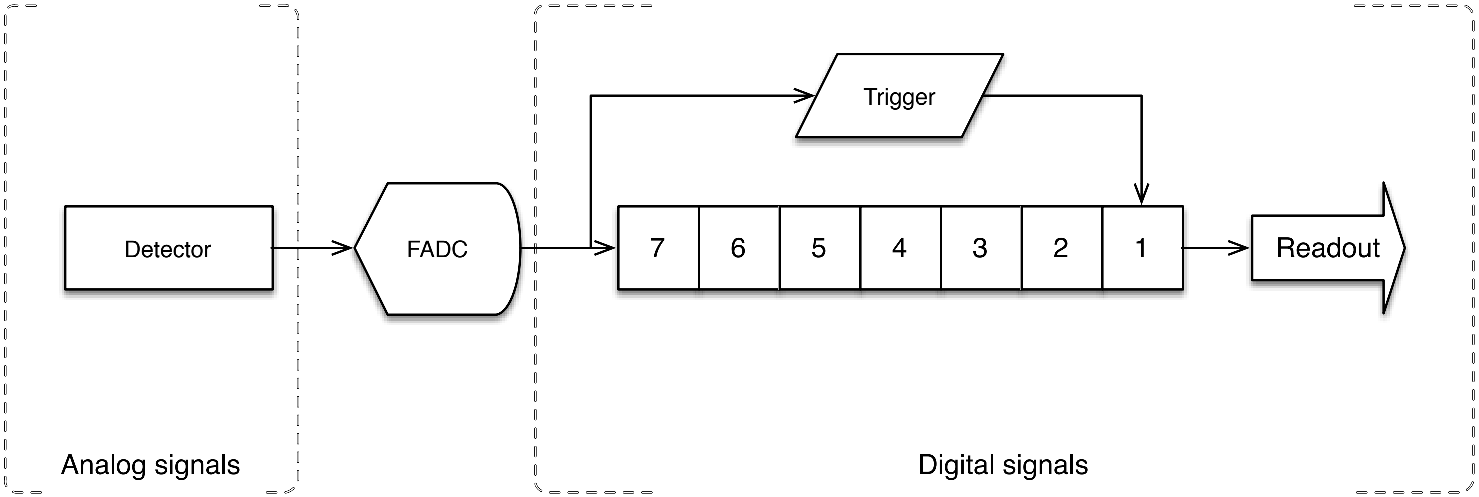
Here’s the long cable in Hall-A.

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.



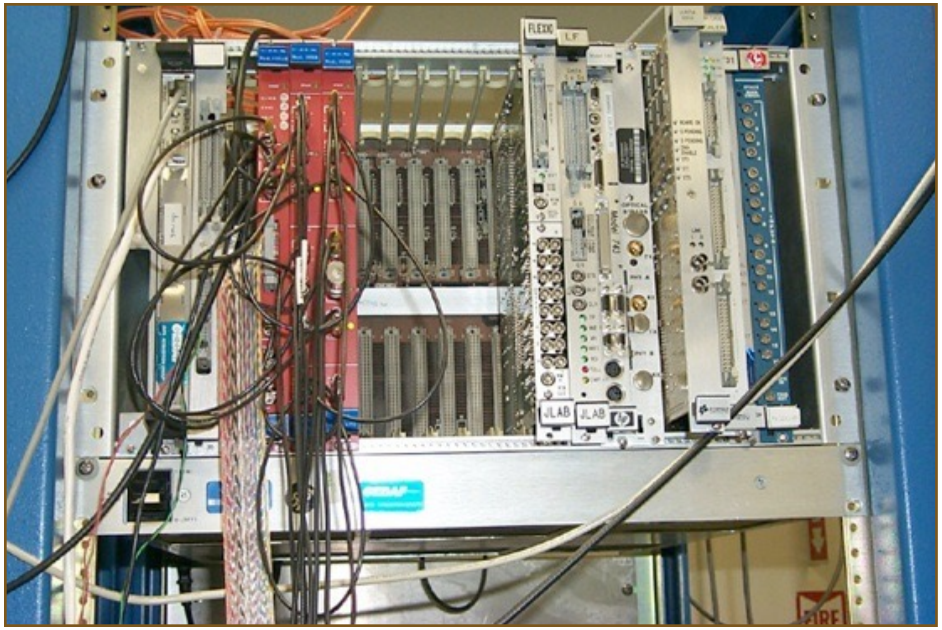
Digital pipeline trigger



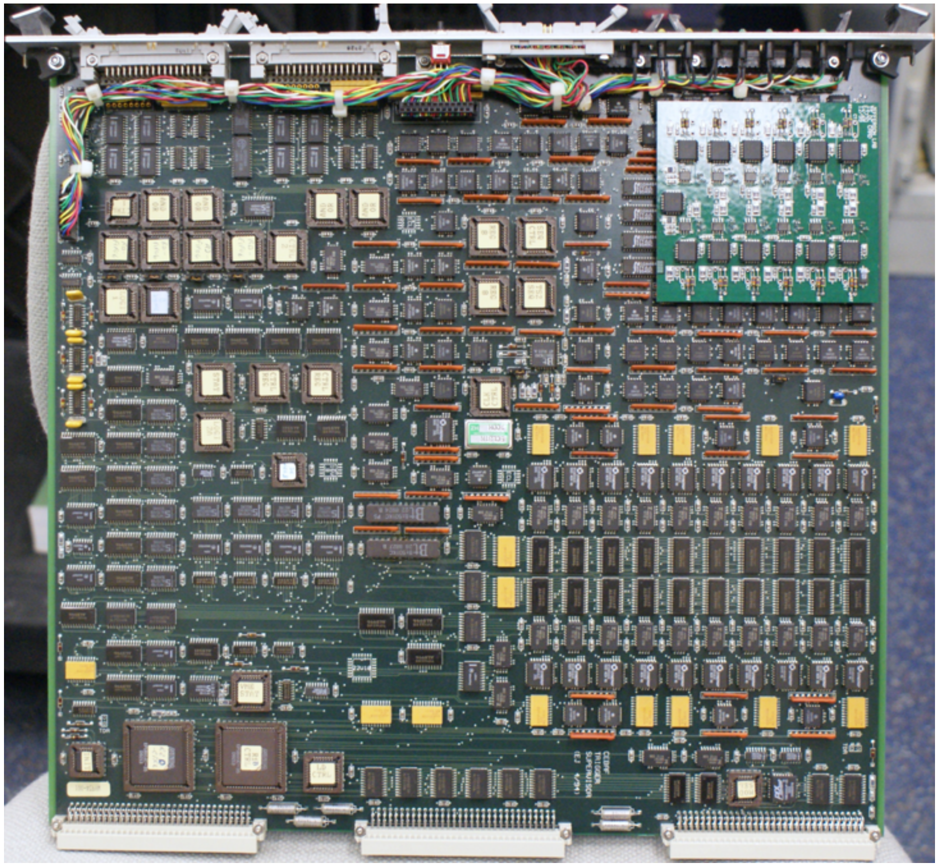
- Replace all that cable with digital memory.
- In a pipelined system a Flash ADC digitizes at a constant rate and stores the values in a memory. Values are clocked into memory at the same rate as the Flash ADC clock which, in the case of GlueX, is 250MHz (4 nS).
- If, for example, the trigger logic 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.
- Needs a synchronized 250 MHz clock and timestamp distributed to every ADC.
- Distribution done digitally over optical cables.

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.
  - GlueX 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 custom electronics

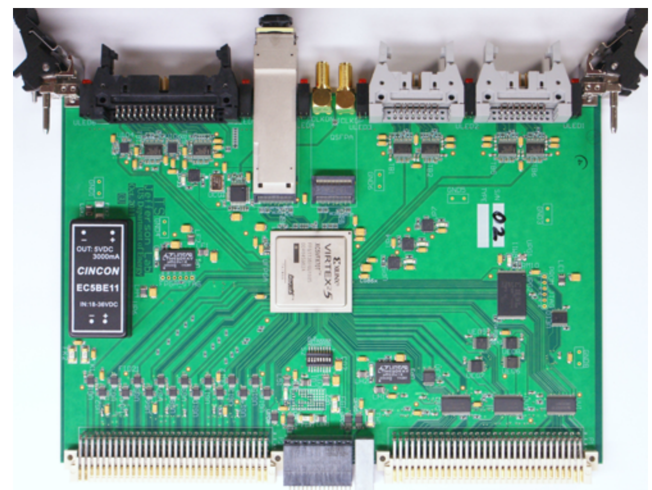


We can now buy programmable logic arrays that allow us to implement complex algorithms in the firmware on a single chip.

The two pictures are of boards with similar functionality.

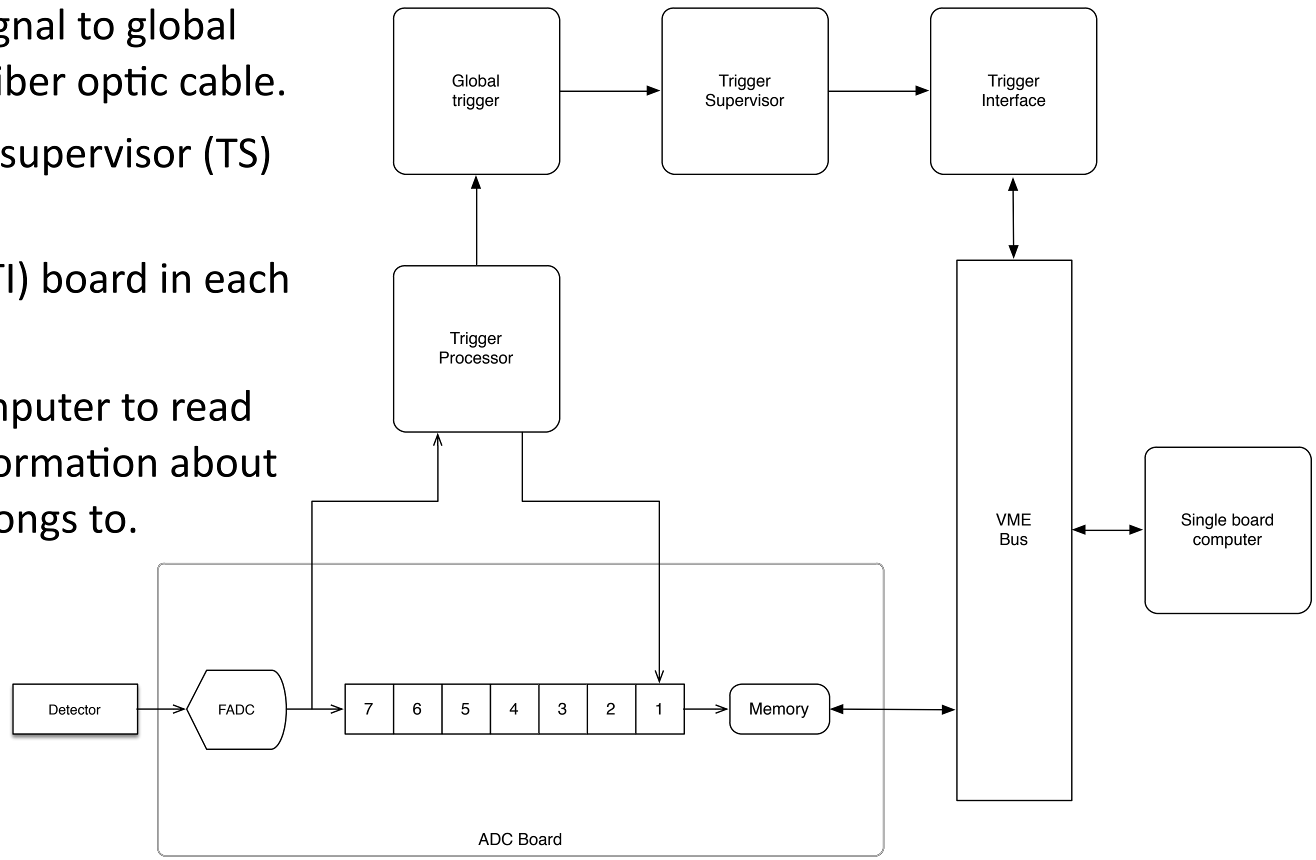
The one on the left designed in the early 1990's.

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.



https://halldweb.jlab.org/wiki/images/e/ec/Rt09_gluex_trigger.pdf

Real world ADCs - fADC 250 MHz

Intel CPU Read Out Controller (ROC) running Linux

Individual ADC channel

16 channels per board

17 boards Per crate so 272 channels per crate

Trigger interface

Boards are connected to CPU via a backplane bus.

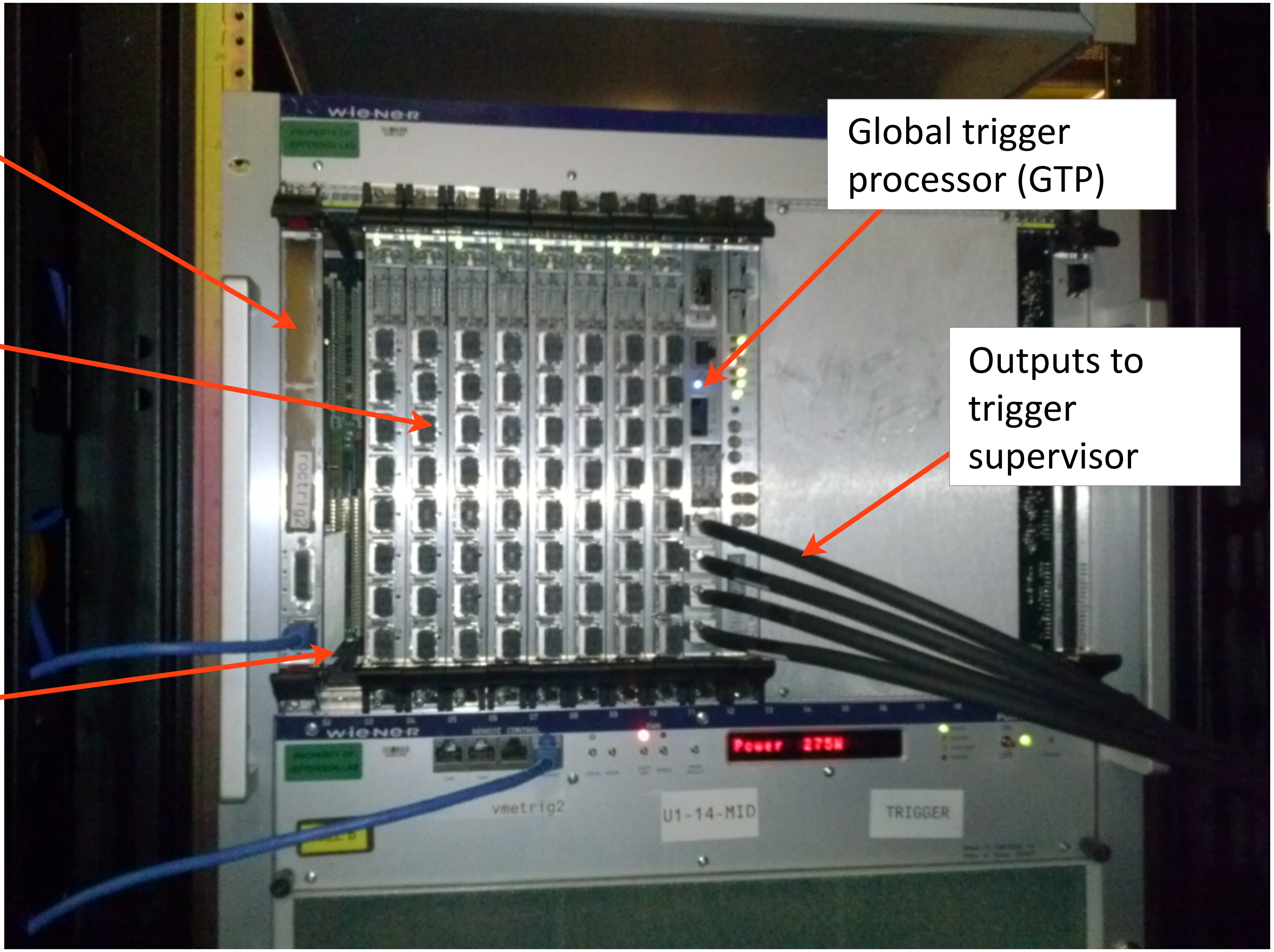
Board sending signals to trigger over fiber

Global trigger crate

Intel CPU controller

Sub-system processor board (SSP)

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



Global trigger processor (GTP)

Outputs to trigger supervisor

Trigger supervisor crate

Signal Distribution board (SD)

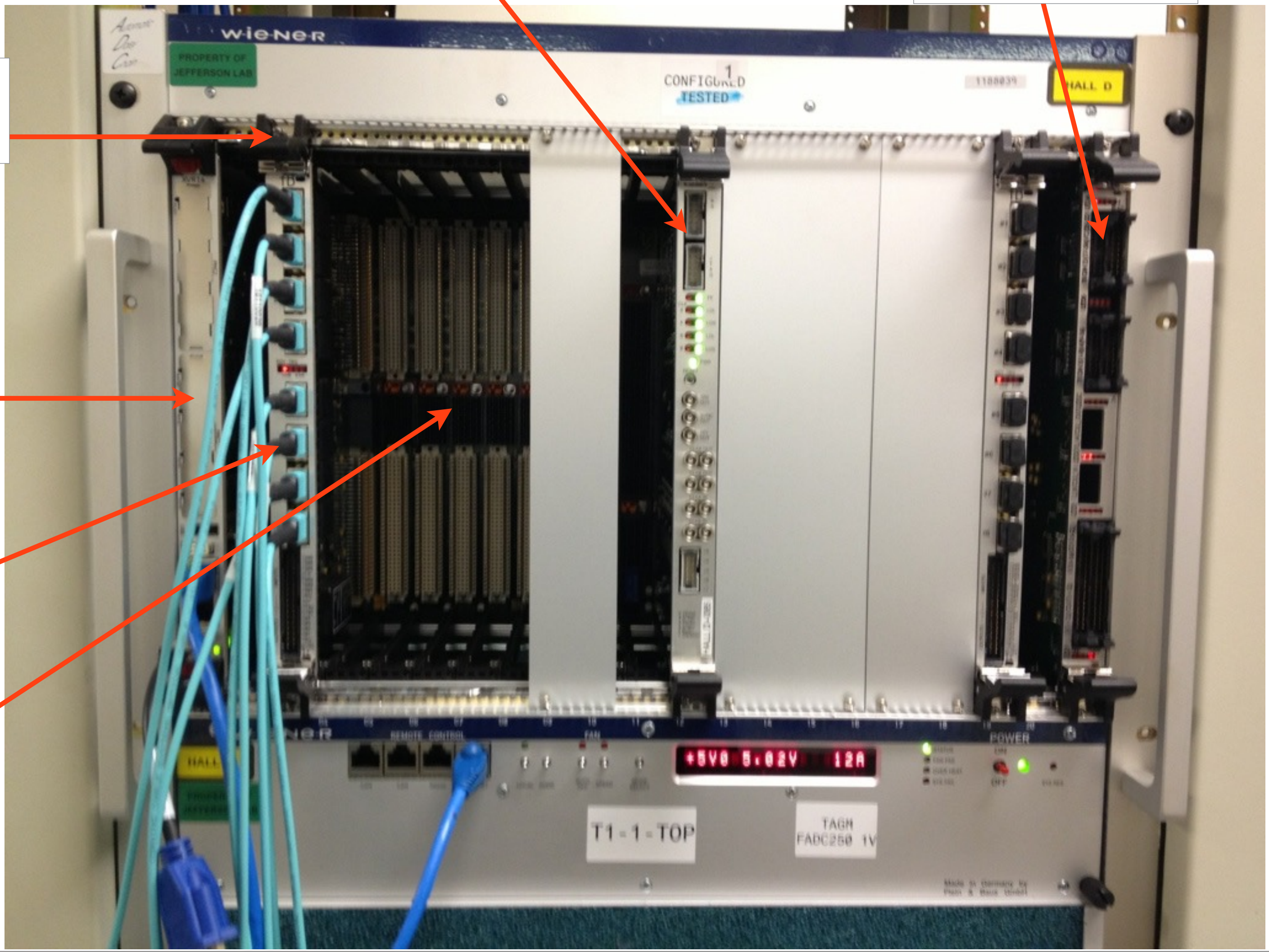
Trigger Supervisor (TS)

Trigger Distribution board (TD).

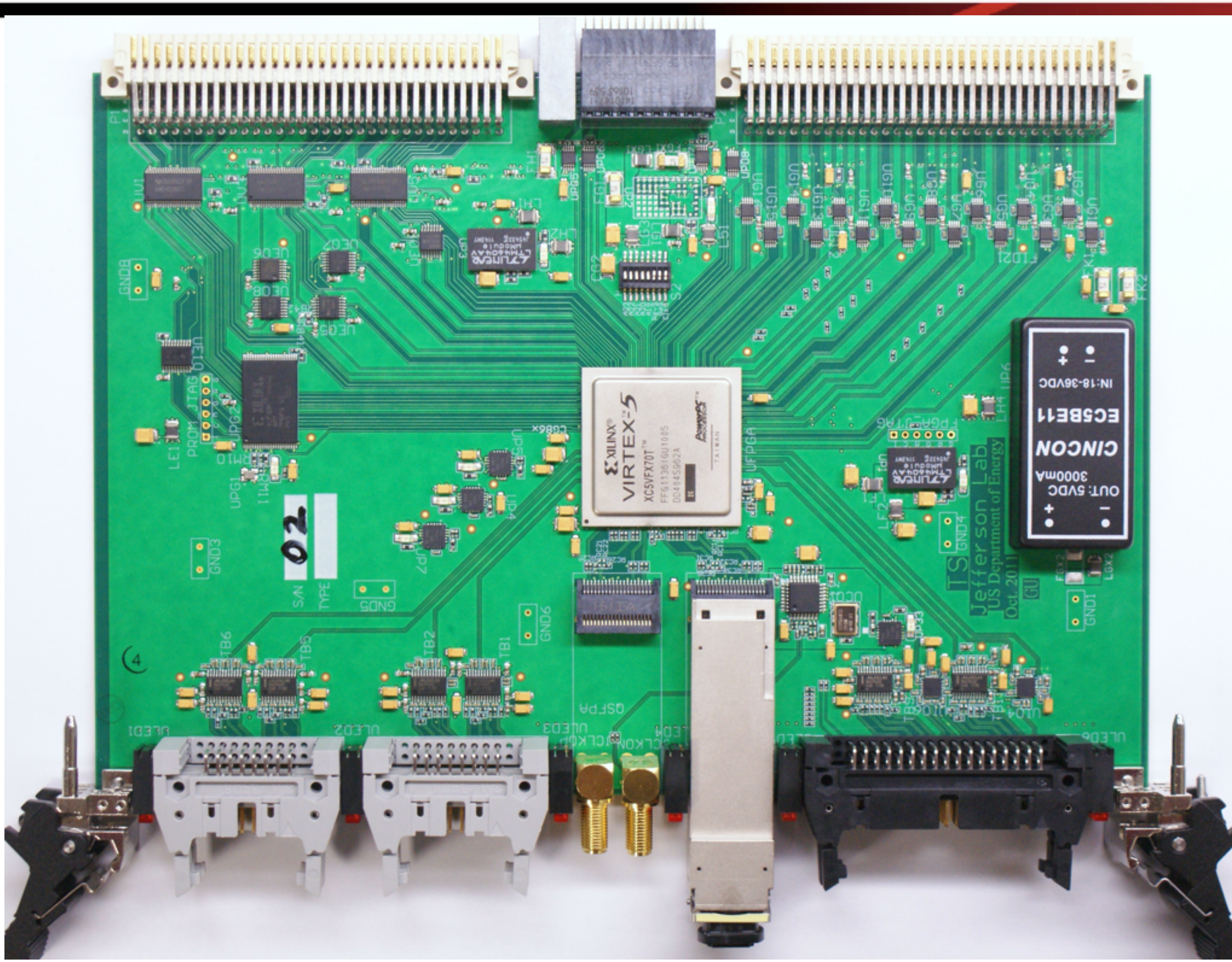
Intel CPU for control and configuration.

Optical trigger link back to crates.

VXS serial backplane



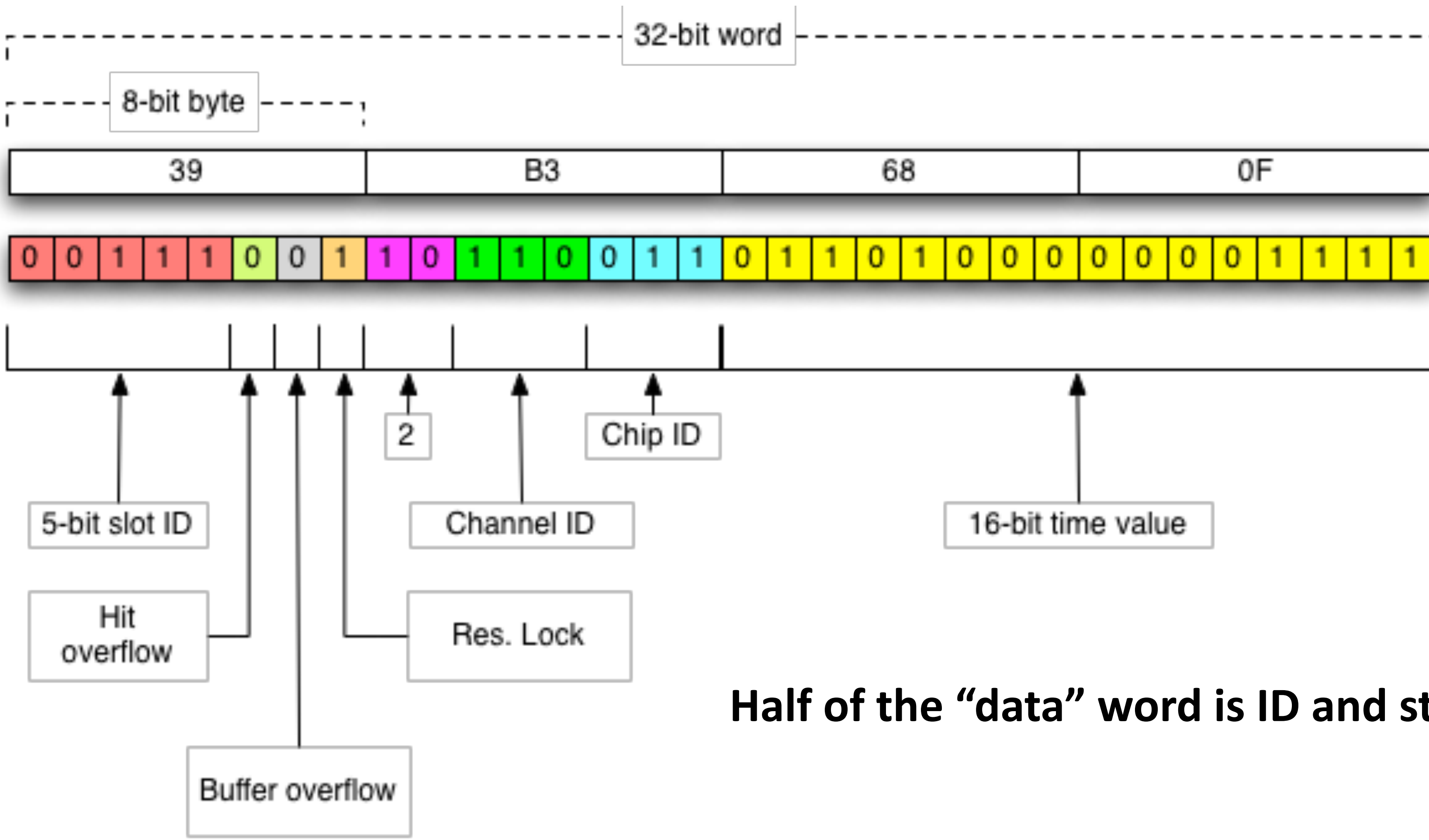
Trigger Supervisor board



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 the data into it's own memory and wraps it in a format that contains the trigger data, which board 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.

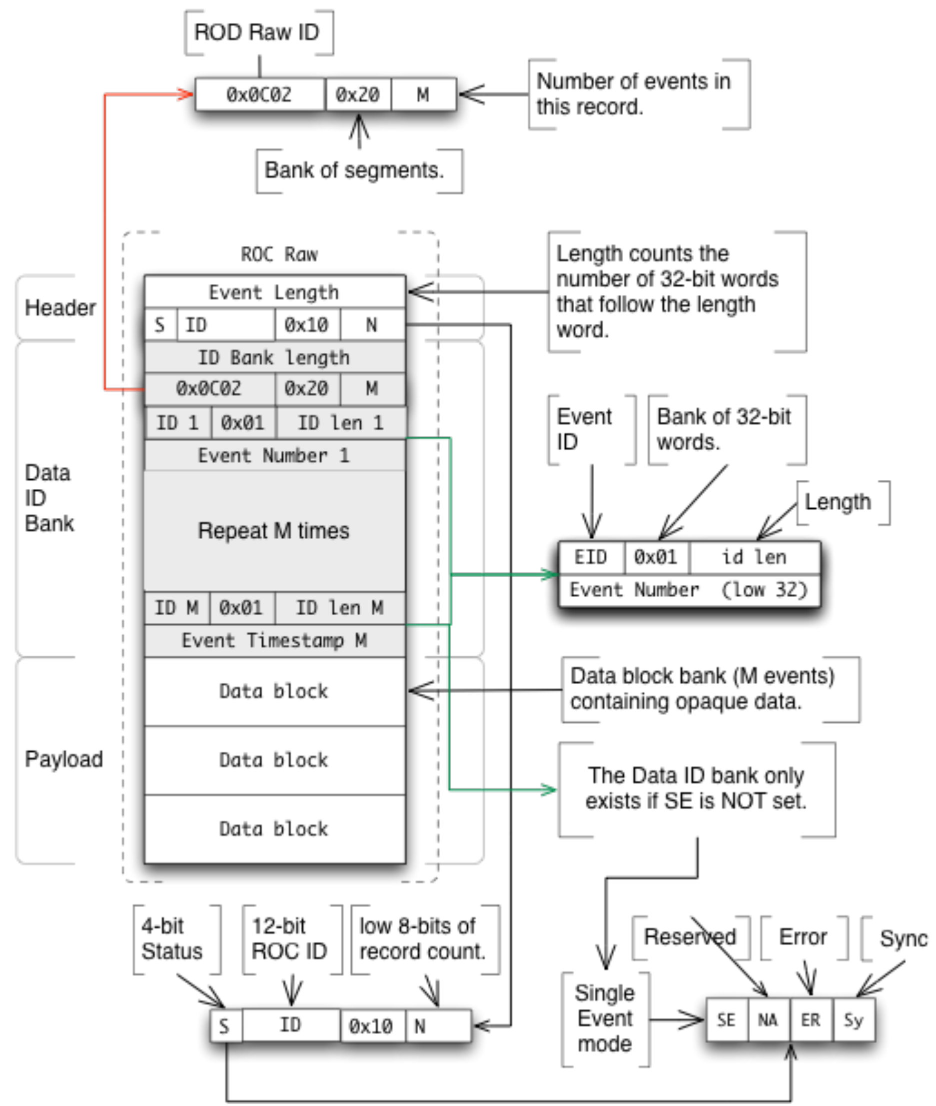
What does the data look like? TDC example.



Half of the “data” word is ID and status.

Data format, a real example

- Bank - container for other data.
- Each bank starts with a header.
  - 32-bit Length.
  - 32-bit Description of content.
- Banks can contain data or other banks.
- Type field 0x10 in outer header tells us this bank contains banks.
- The first inner bank has type 0x20, a bank filled with segments. (Segments are small banks with 16-bit header and length).
- This 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

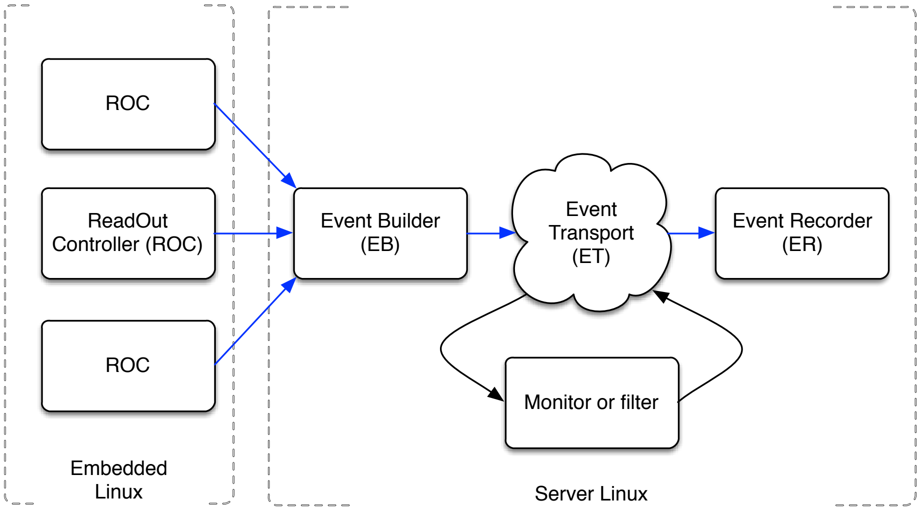
- The GlueX detector and hardware trigger are capable of producing trigger and data rates that are a challenge for existing technology.
- The GlueX collaboration came up with a design specification for the DAQ based on estimates of how the detector should perform.
  - Data spread over 50+ front end systems (crates).
    - 15 kByte events
  - Design luminosity of $5x10^7$ γ/s (photons per second)
    - Event rate of 200 kHz survive trigger, data rate off detector = 3 GByte/s.
  - Design calls for a rate to storage = 300 MByte/s.
  - This reduction is 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.
- How do we implement a data acquisition system that can do that?

CODA

- CODA is the software that all experiments at JLab use to implement DAQ systems.
- CEBAF Online Data Acquisition
  - Toolkit 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 physical volume of space.
- CPUs in crates send data via a network to one or more servers running Linux.
- Bits and pieces of events arrive at different times from different places.
- All the parts of the event need to be collected together and packaged with other information needed by the analysis.
- This is done by the Event Builder which may be thought of as a sophisticated collating software.
- The software represented by the rectangles are called “CODA components”



Data transport

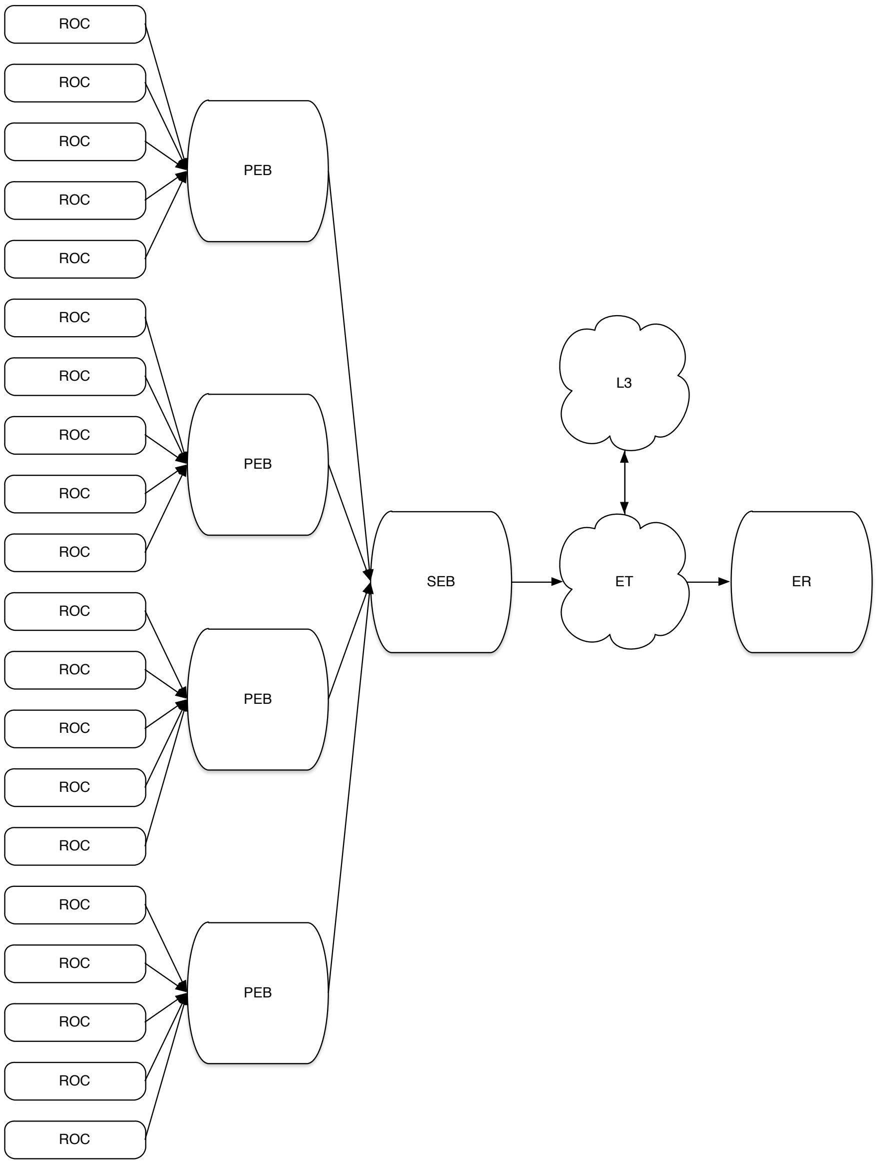
- In the GlueX DAQ implementation of CODA data is read via DMA along the VME backplane into the memory of an Intel CPU running Linux and the CODA ROC software.
- The ROC software sends the data to the EB over the 1 GB/s network link on the front panel (blue cable on the photo).
- We have tested replacing the 1 GB/s link with a 10 GB/s link. The ROC can send at 500 MB/s but the VME backplane limitation is lower than this ~ 200 MB/s per crate.

Event Building challenges

- The GlueX design goal event rate is 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 design goal data rate is 3 GByte/s over 50 incoming links.
  - 60 MByte/s average per incoming link.
  - 3 GByte/s output from EB over the network.
  - Since data is copied several times the data rate inside a machine is several times 3 GByte/s.
- That would be a lot for one LINUX server machine to handle
  - Solution: multi stage parallel event builder.

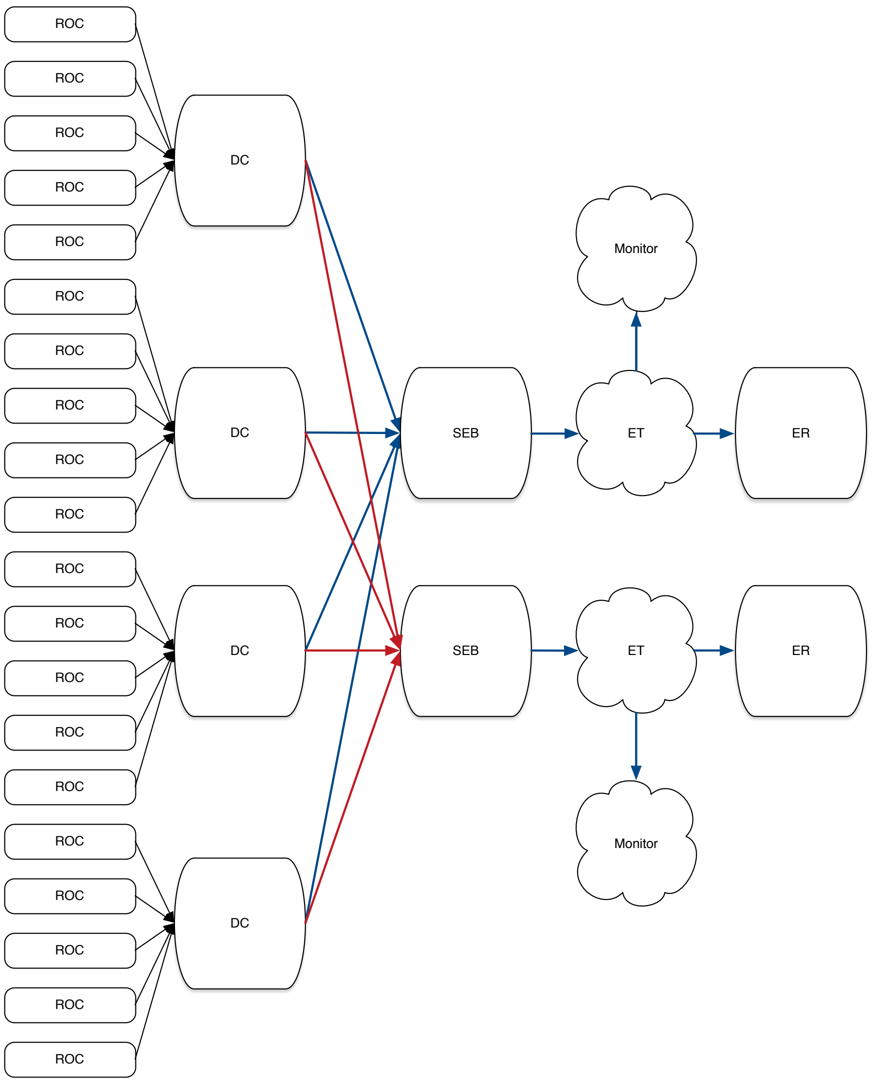
Staged Event Builder, 20 ROC example

- Four Primary Event Builders (PEB) are connected to five ROCs each by 1 Gbit/s Ethernet links (Switched ethernet network).
- Each PEB handles only a quarter of the work load and data throughput.
- The four PEBs are connected to a single Secondary Event Builder (SEB) via 40 Gbit/s Infiniband links (switched IB network)
- 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.



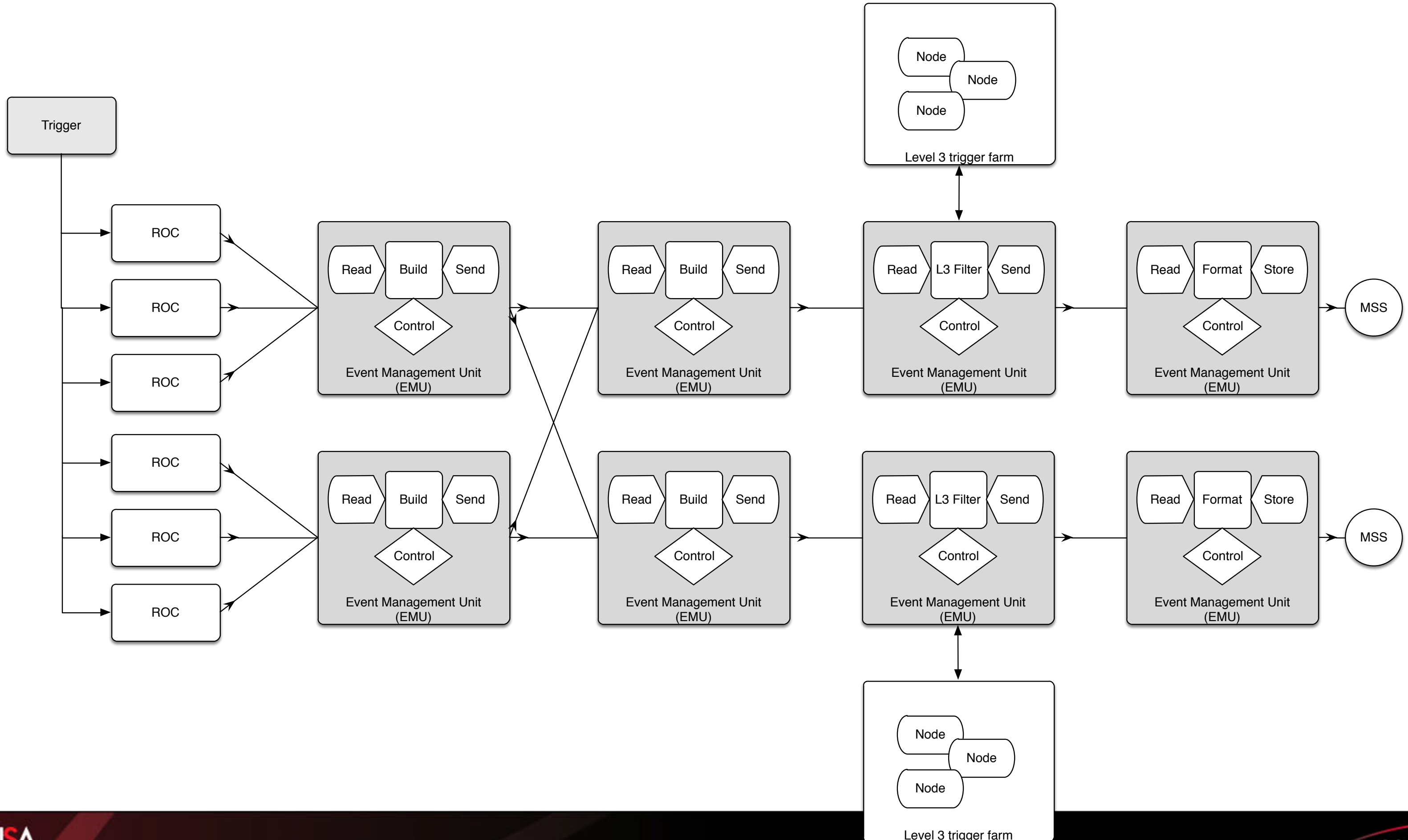
More parallel streams

- CODA allows more than one SEB to operate in parallel.
- The two SEBs each handle half of the data.
- Can add more parallel streams.
- The main challenge is control.
  - How to start and stop such a complex network of components?



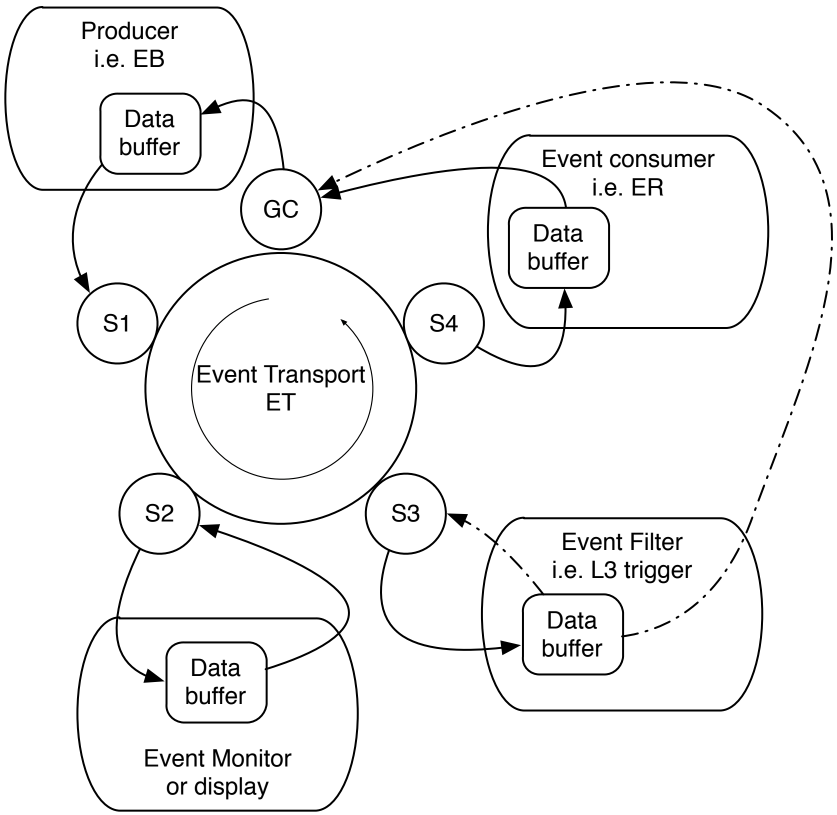
Reusable modular software

- Several parts of the system, the PEB, SEB and ER are very similar in function. Data comes in, is reformatted in some way and is output.
- We wrote a common framework called the EMU (Event Management Unit) that can be configured to perform these functions and any similar function that we may need in the future.



Event Transport, ET

- The ET system solves the problem efficiently giving programs access to the data.
- The system uses a railroad metaphor. Empty data buffers originate at Grand Central. They are filled by one or more 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.
- An event monitor may set up a station, S2, to sample 1% of the events.
- An event filter may set up S3 to stop all events. It then decides which events to keep. Discarded events are sent back to GC good ones move on.
- An event recorder stops all events and, after the data is written to a file sends the buffer back to GC.

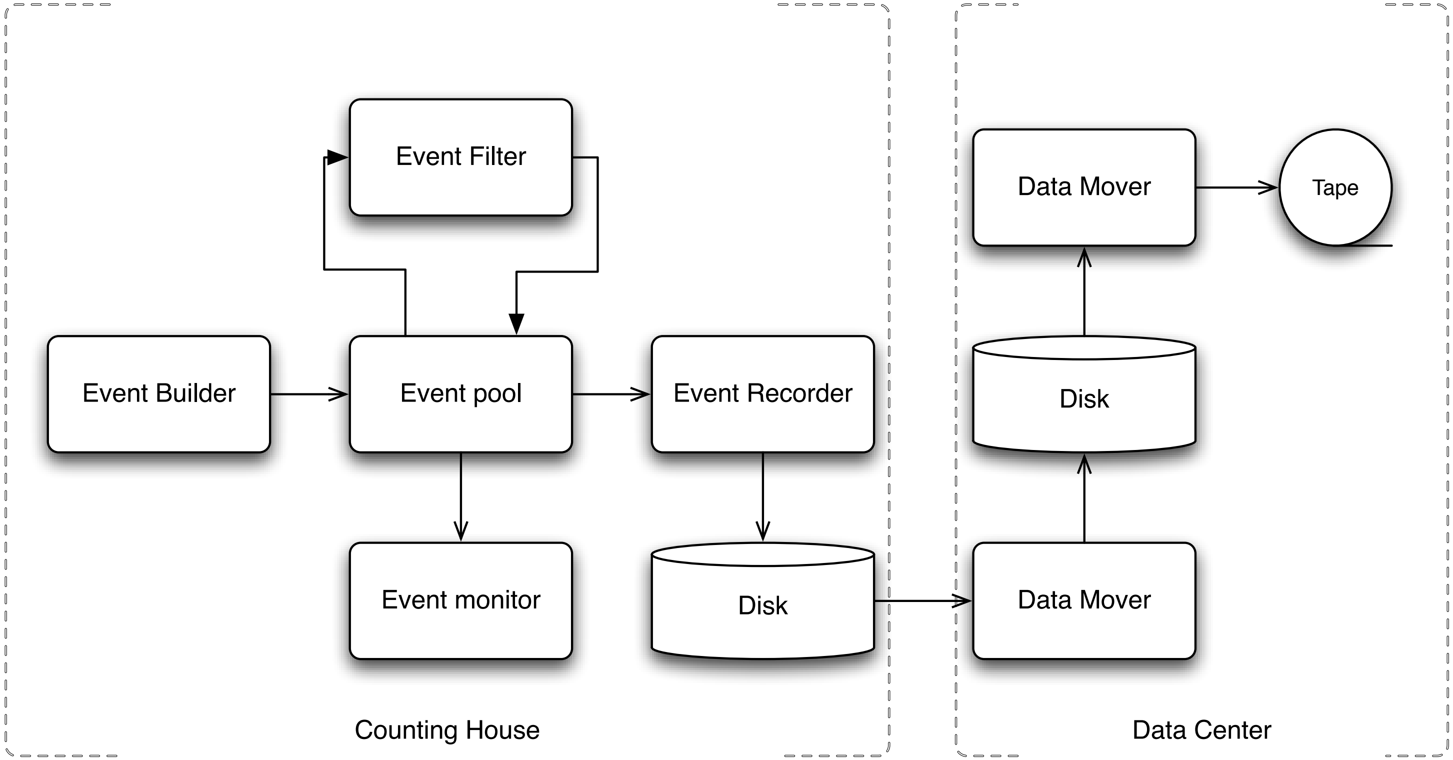


Data storage

- 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.

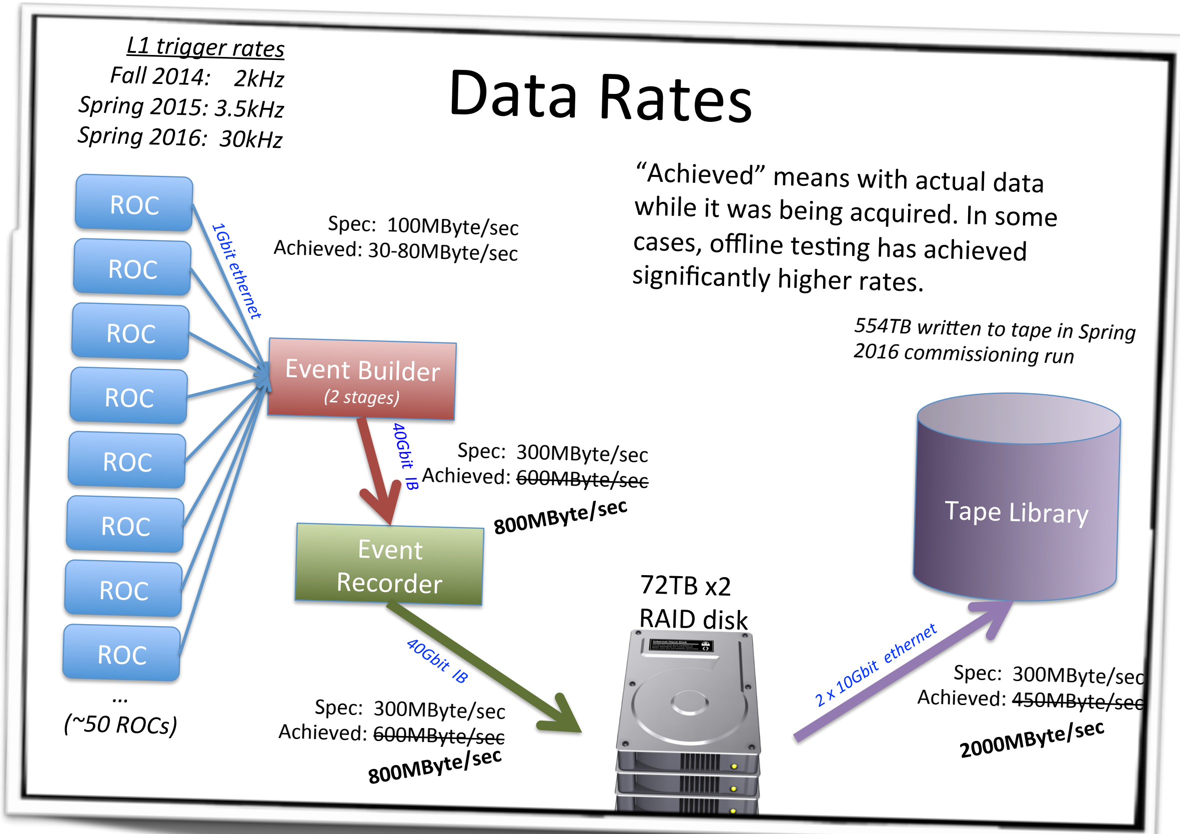
Data flow to tape

- The event recorder writes to several network attached storage devices in the counting house.
- A machine in the CEBAF center datacenter, about a mile away, reads from that disk to a local cache disk from where the data is migrated to tape.

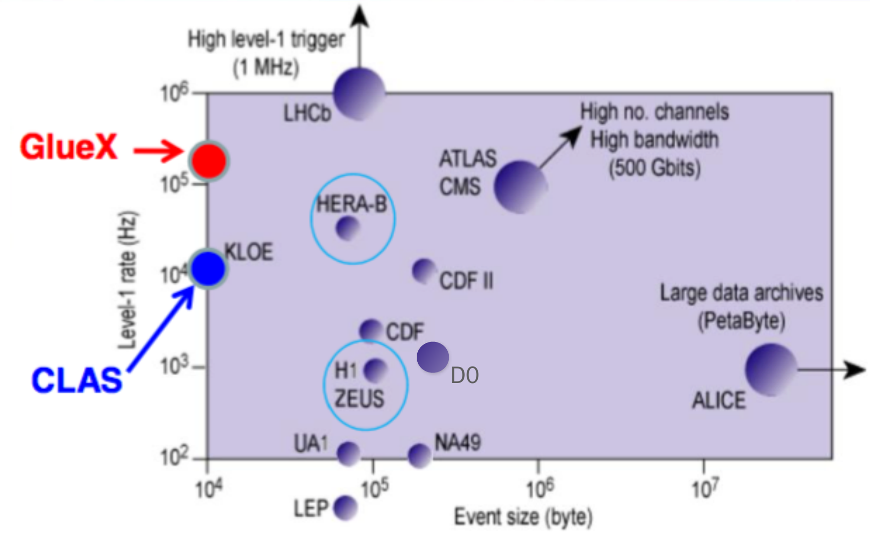


All good on paper but…

- Spring 2016, low luminosity, 0.8×10^7 γ/s GlueX ran at 800 Mbyte/s !!
  - Remember the design goal was 300 Mbyte/s
  - Accidental hits add quadratic term to data rate vs luminosity.
- Wondering what rate for 5×10^7 γ/s will be? Maybe 9 GByte/s instead of the expected 3 GByte/s?
- Fall 2016 test the L3 trigger.
- Fall 2018 high luminosity.



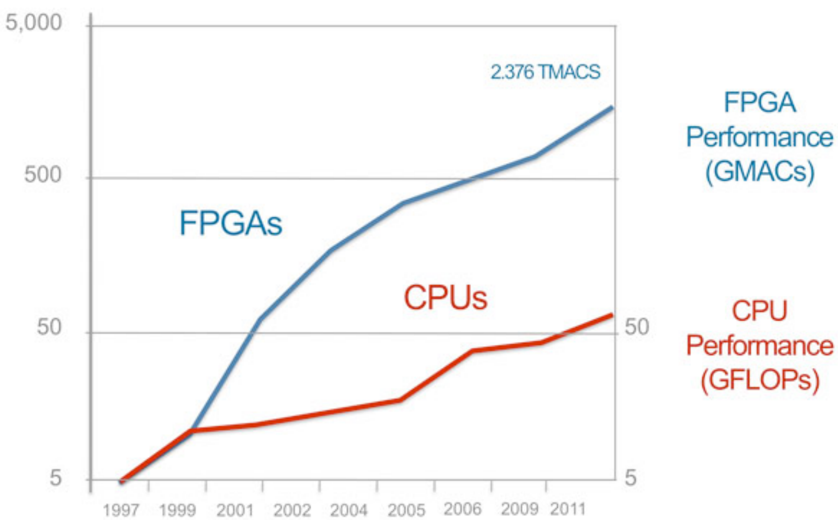
The Future - Trends in experiments

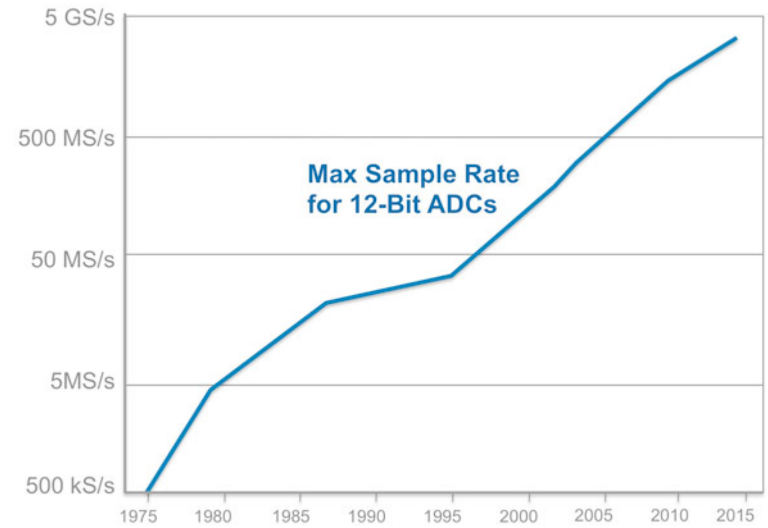


- Look at historical trigger and data rates.
- At JLab
  - mid 1990's CLAS, 2 kHz and 10-15 MB/s
  - mid 2000's - 20 kHz and 50 MB/s
  - mid 2010's
    - HPS, 50 kHz and 100 MB/s
    - GlueX
      - 100 kHz, 300 MB/s to disk.
      - (Last run 35 kHz 800 MB/s)
- FRIB - odd assortment of experiments with 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.

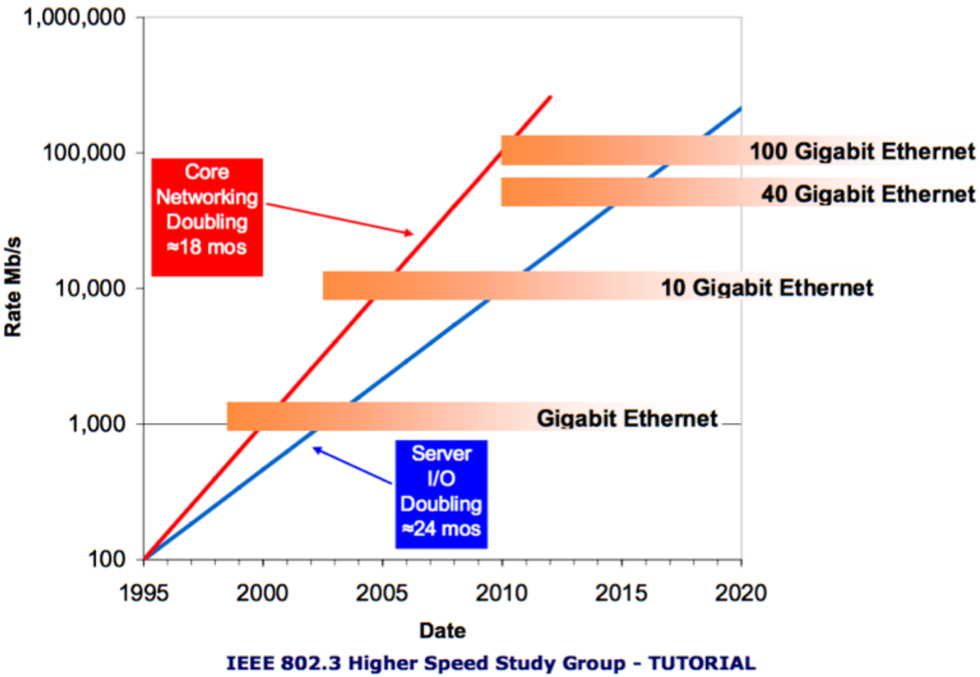




- Current trend is to push some functionality currently performed in software running on embedded processors into firmware on custom electronics. This will probably continue.

Trends in data transport

- Network speed trend 2x every 18 months.
- Server IO trend 2x every 24 months.
- Network technology is shown as a horizontal bar. It is introduced at the left of the bar and becomes cheaper.
  - 10 Gb/s appeared in mud 2000's but we could only afford it in any quantity maybe 2010.



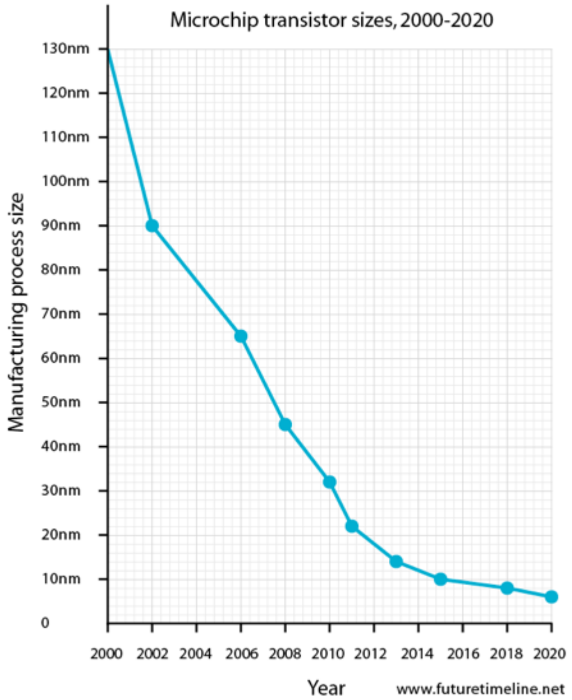
Challenges

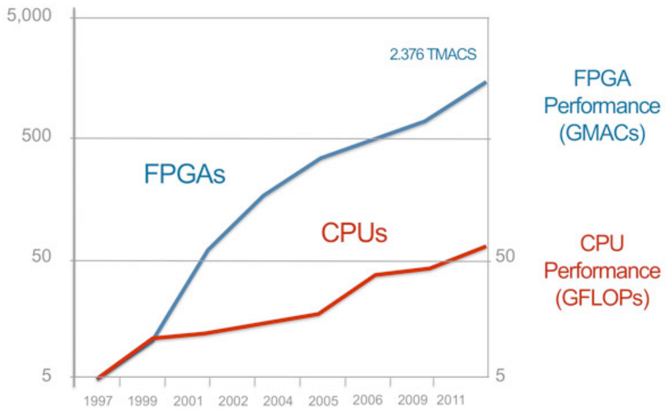
- The precision of the science depends on statistics which leads to :
  - Development of detectors that can handle high rates.
  - Improvements in trigger electronics - faster so can trigger at high rates.

- 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.
  - Increased storage rate to tape.
- 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

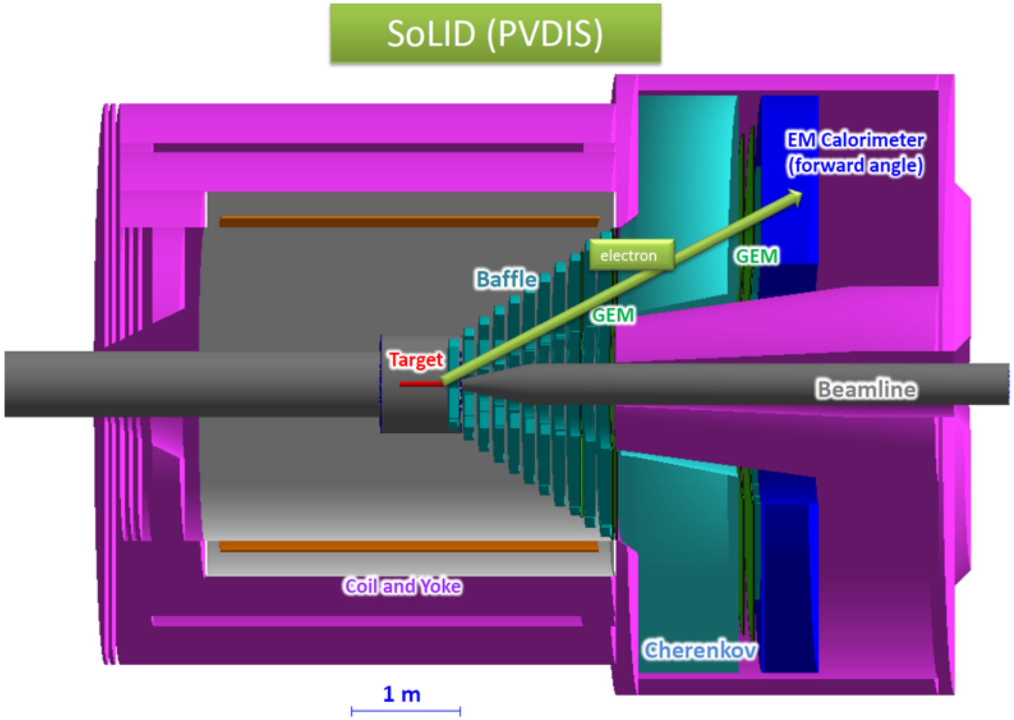
- Manufacturers are struggling shrink transistors.
  - How much further can Moore's law continue?
  - When does this trickle down affect the performance of other DAQ electronics?
- Use of mobile devices is driving tech in a direction that may not be helpful to NP DAQ, low power and compact rather than 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.

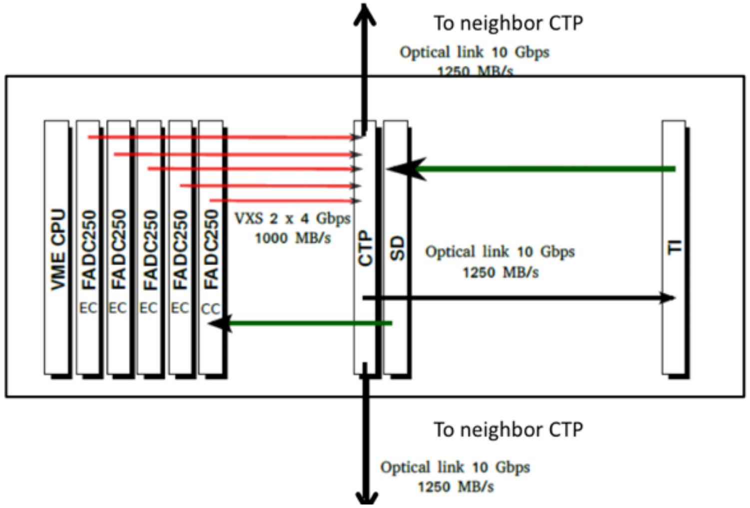




Future experiments, JLab - SoLID

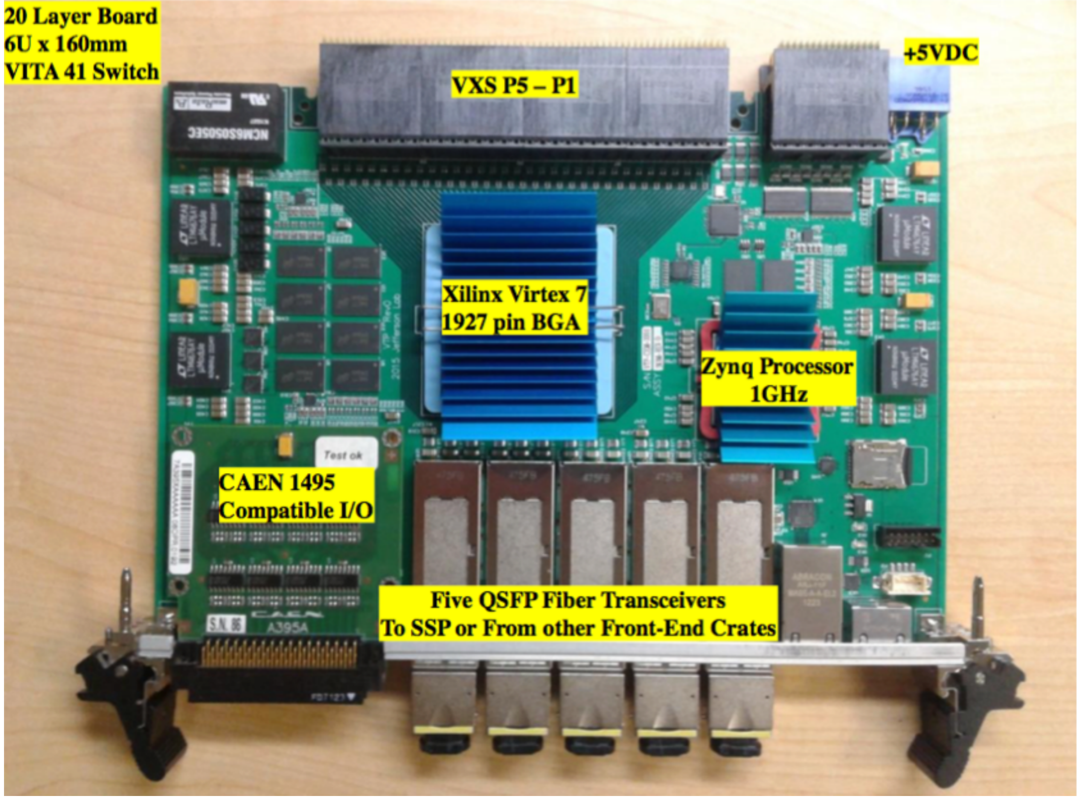
- SoLID is an experiment proposed for installation hall-A at JLab.
- The detector has two configurations. In the PVDIS configuration electrons are scattered of a fixed target at high luminosity.
- The detector is split radially into 30 sectors, the single track event topology allows 30 DAQ systems to be run in parallel at rates of up to 1 GByte/s each, 39 GByte/s aggregate.
- L3 trigger reduces final rate to something reasonable to send to mass storage.





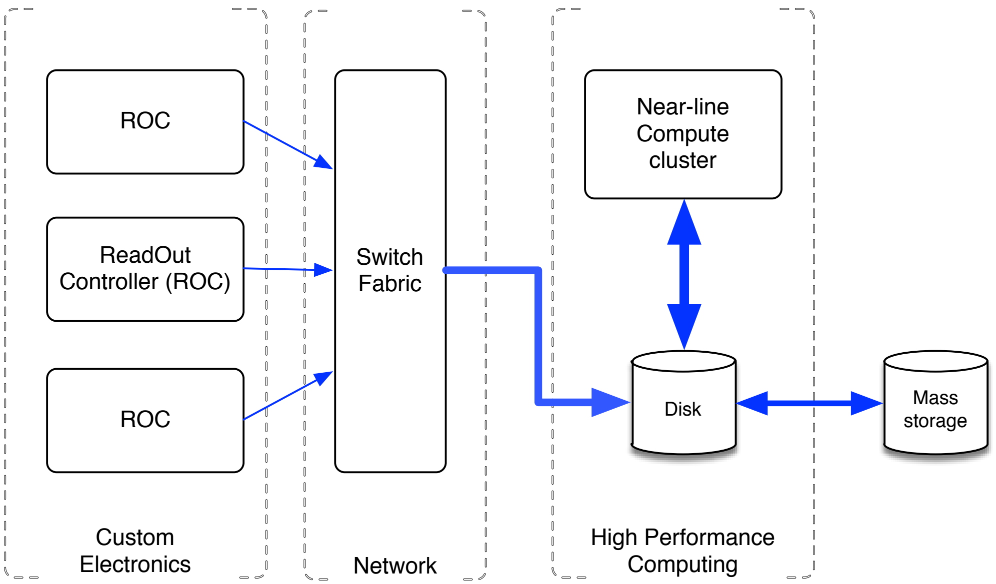
Possible hardware

- 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 (July 2016!).
- Need to program the Xilinx Vertex 7 chip to handle the full data flow rather than just the trigger data.



Solution in ten years?

- Can't escape some sort of crate to put the electronics in maybe MicroTCA to replace VME?
- Stream the time-stamped 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.
  - Time stamped streaming ADCs
  - Terra scale computing.
  - Software to organize time-stamped data.



Summary

- TJNAF, or as we call it JLab, has just finished an upgrade to 12 GeV and added hall-D with the GlueX detector. As part of this upgrade we have developed electronics and software that has allowed us to implement the GlueX DAQ and trigger.
- The same electronics and software is being put to use in the other halls.
- GlueX is on the edge of what is possible with this technology. When other experiments, like SOLiD, come along at the end of this decade we will need to do something different.
- We are interested in the possibility of ADCs and TDCs that are capable of streaming zero suppressed time-stamped data to temporary storage at sample rates of the order of 1 GHz, channel counts in the range of tens of thousands and configurable resolution of 12-bits or more. We would run such a system in an almost trigger-less mode and rely on near-line software on what would now be classed as a supercomputer to identify and process events.
- Our timescale is over five to ten years.