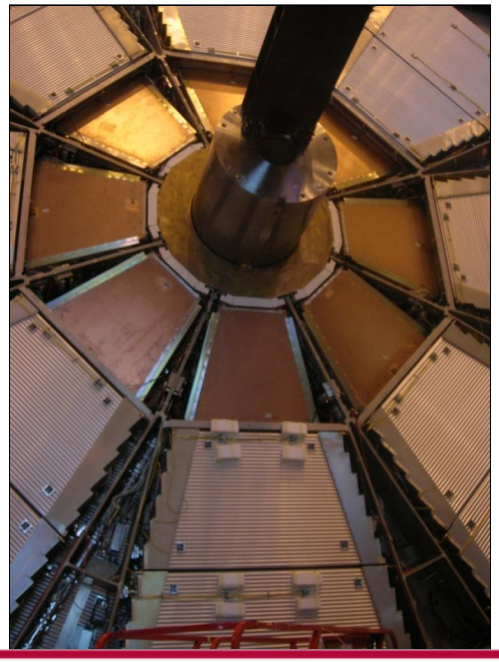
CSC ROD Replacement Conceptual Design Review

Firmware and Software Design of the New ROD Complex (NRC)



Richard Claus†

claus@slac.stanford.edu

CERN, October 8, 2012

On behalf of:

Rainer Bartoldus†
Anthony DiFranzo††
Raul Murillo Garcia††
Nicoletta Garelli†
Ryan T. Herbst†
Michael Huffer†
Andrew J. Lankford††
Andrew Nelson††
James Russell†
Michael Schernau††
Su Dong†

U.S. DEPARTMENT OF ENERGY
Office of Science

† SLAC NATIONAL ACCELERATOR LABORATORY
†† UCI

Outline

SLAC

Overview of the system

- The system as a collection of well defined components (Mike's talk) bound together through firmware and software
- Will show the flow of data (of various kinds) through the system
- Will itemize the external interfaces and show how we meet them

Software framework

Development Infrastructure

- Trigger Simulator
- CSC Emulator

Configuration management

- Code management
- Release method

Performance

Summary

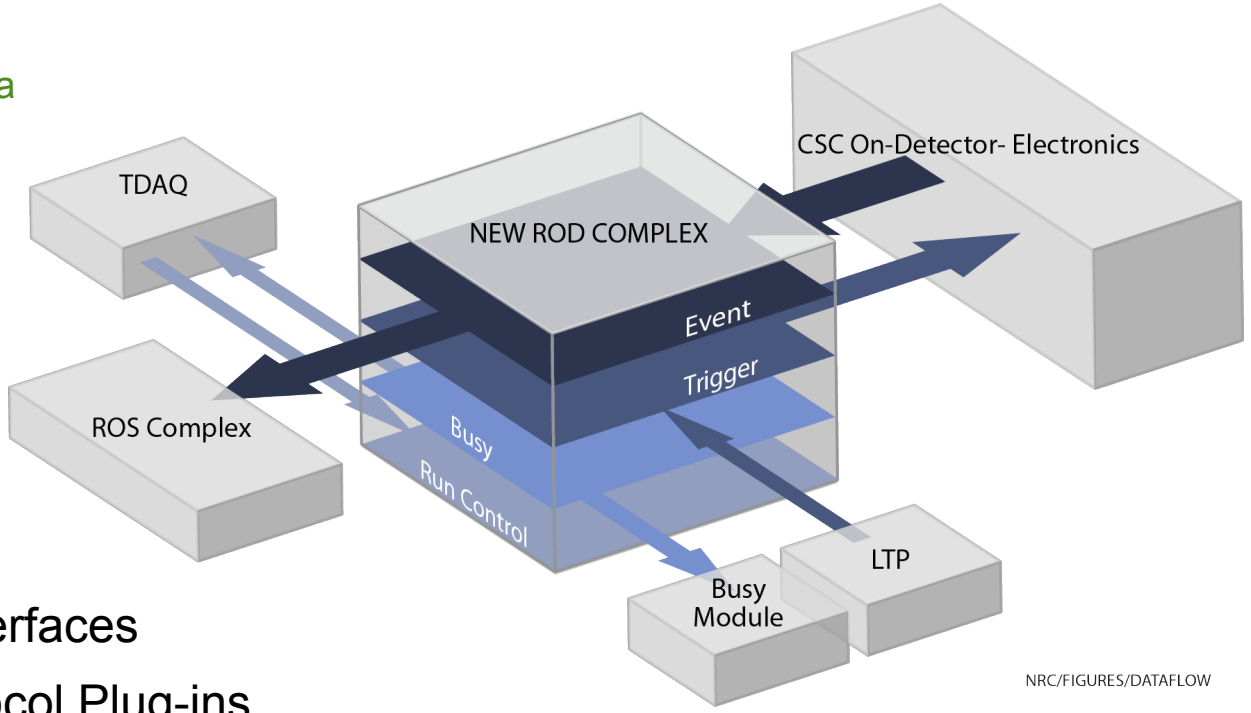
Overview of the NRC

System is decomposed into a set of independent *planes*

Planes are associated with the type of data they transport

RCEs interact on all planes

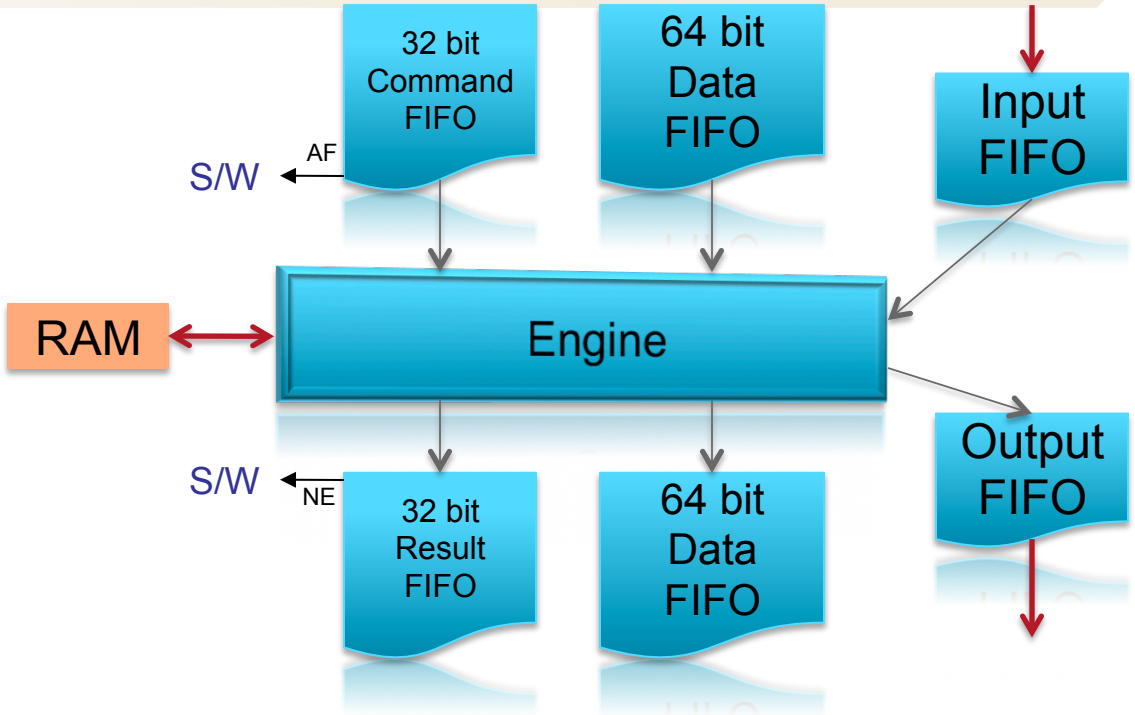
- Event
  - Physics and Calibration data
- Trigger
  - Trigger information
- Busy
  - Back-pressure
- Run Control
  - Control, monitoring and configuration



Must satisfy the external interfaces

Accomplished through Protocol Plug-ins

The PPI Plug



Communicates between f/w & s/w

Uses user defined instructions:

- Auxiliary Processor Unit (APU) load/store
- 3 CPU cycles per instruction (CPI)

Can interrupt CPU, e.g., Not Empty (NE)

Status flags, e.g., Almost Full (AF)

Command/Result contain length & location

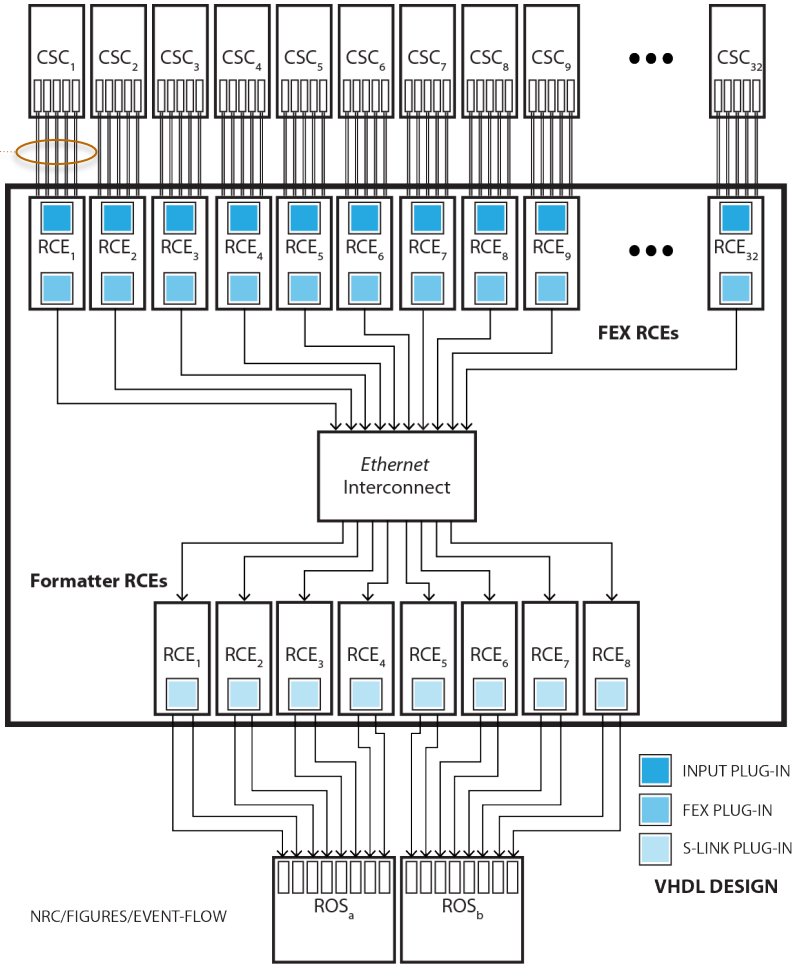
Location can be:

- Data in the Data FIFO
- Address of a block in RAM to DMA

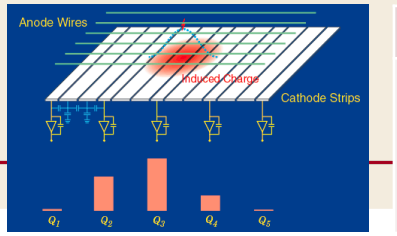
Allows trade between DMA setup overhead and throughput

Event data flow

- Input of data on the FEX RCE
  - For each time-slice (samples), a chamber produces 5 layers x 192 channels x 12 bits
    - A nominal event is 4 time-slices = ~6 KB
  - *All* chamber data comes into the NRC through 5 1.28 Gbit/second links
    - With 4 time-slices per trigger:
      9 µS = 360 BC = 111 KHz
- Perform feature extraction (FEX)
  - Reduce the input data volume
  - Threshold & timing cut and composing clusters of hits
    - Expect ~600 Bytes per chamber at <µ> = 80
- Forward to Formatter RCE
  - Via Ethernet and the Fabric Interconnect
- Format the data
  - Package up the data from 2 chambers into an ATLAS CSC event contribution
- Send it to the ROS



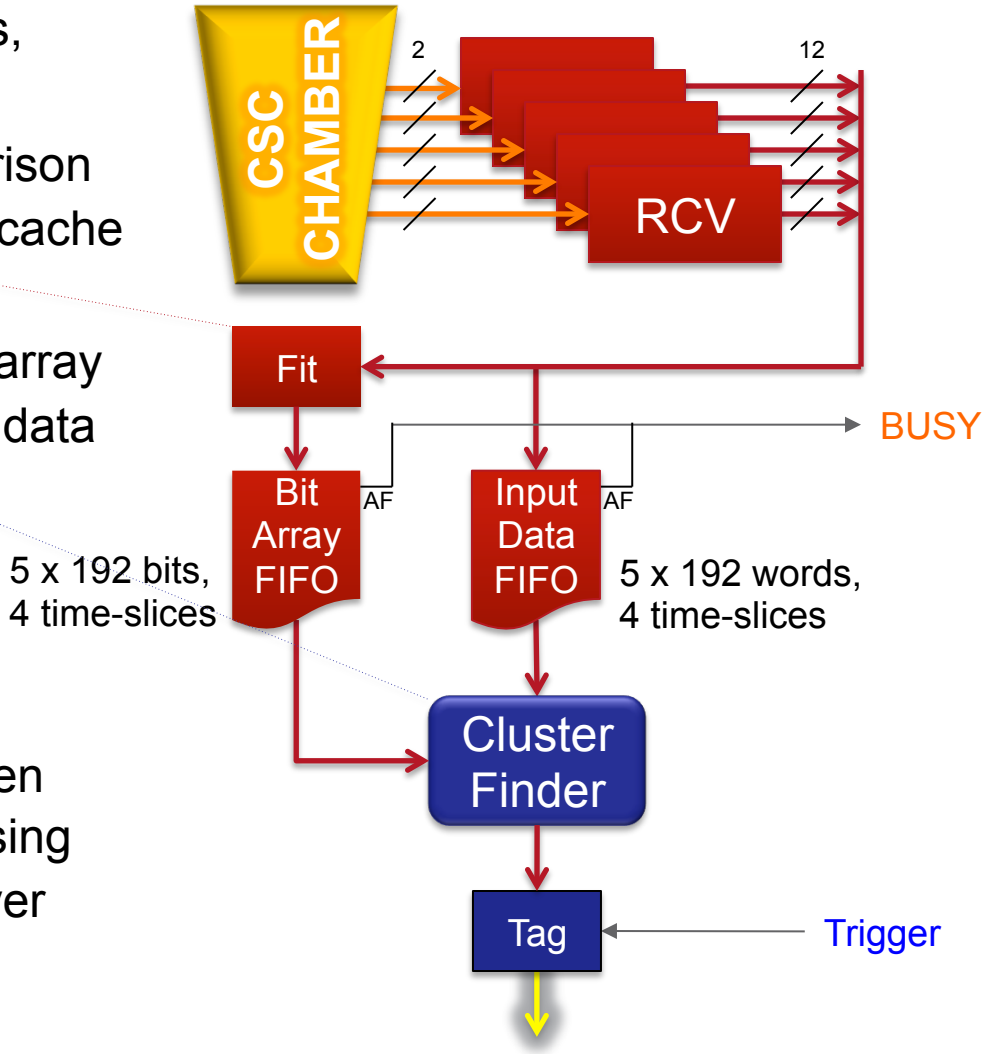
Event data flow via Input and FEX PPIs



x192	Layer				
Time-slice	3.4	3.3	3.2	3.1	3.0
	2.4	2.3	2.2	2.1	2.0
	1.4	1.3	1.2	1.1	1.0
	0.4	0.3	0.2	0.1	0.0

SLAC

- RCV receives data from the fibers, takes up arrival time differences, builds parallel data structure
- FEX f/w does threshold comparison and Out-of-Time cut to build a (cache friendly) bit array of "hits"
- Cluster Finder s/w uses the bit array to determine which of the Input data to forward to its Formatter RCE



- Tag data with trigger info
- Manage transfer latency between FEX and Formatter RCEs by using low-level, point-to-point Link layer protocol

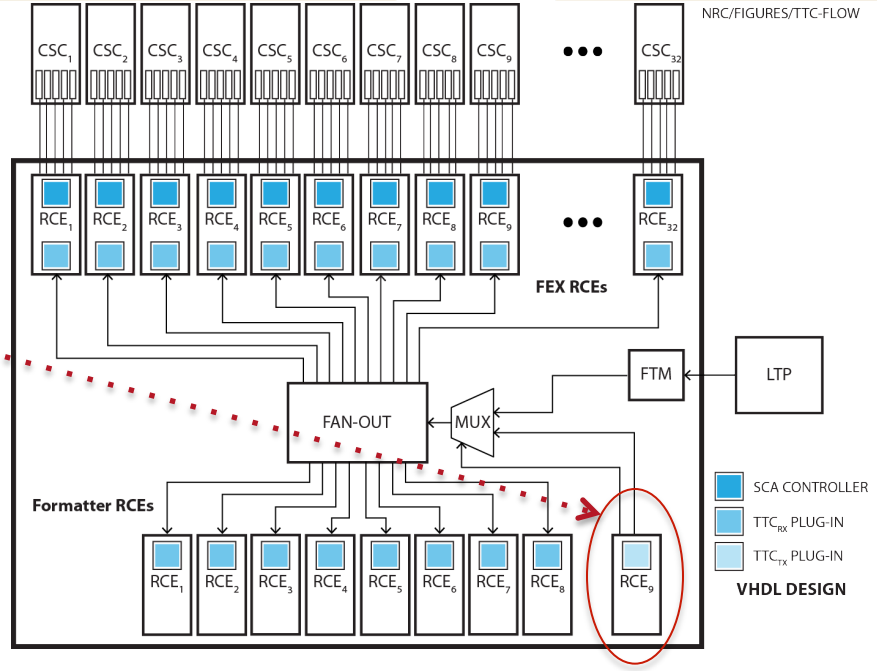
S-Link PPI

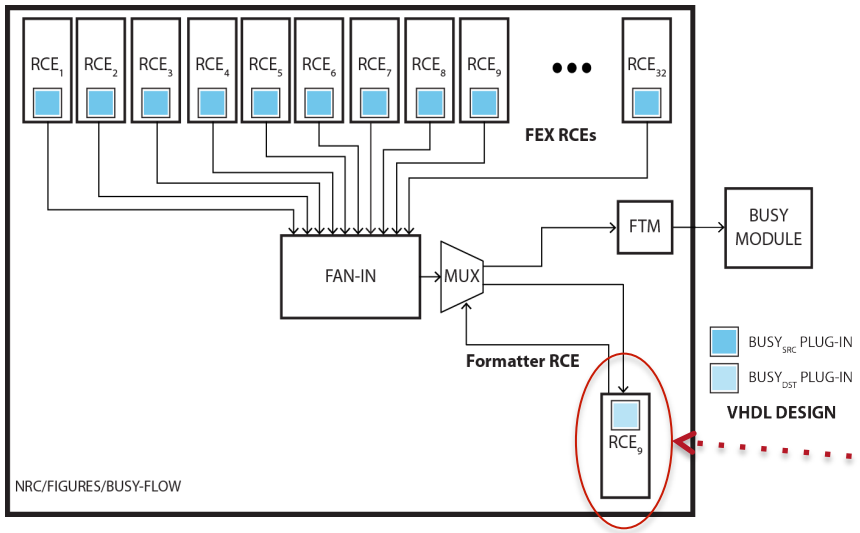
SLAC

- Is the ROL interface
- Follows CERN's specification
- Full duplex, 160 MByte/second link
  - One duplex carries the data
  - The other carries the flow control

Trigger System

- TTC_{RX} Plug-in
  - Receives a trigger message
- TTC_{TX} Plug-in
  - Emits a trigger message
  - Used only for the Trigger Simulator





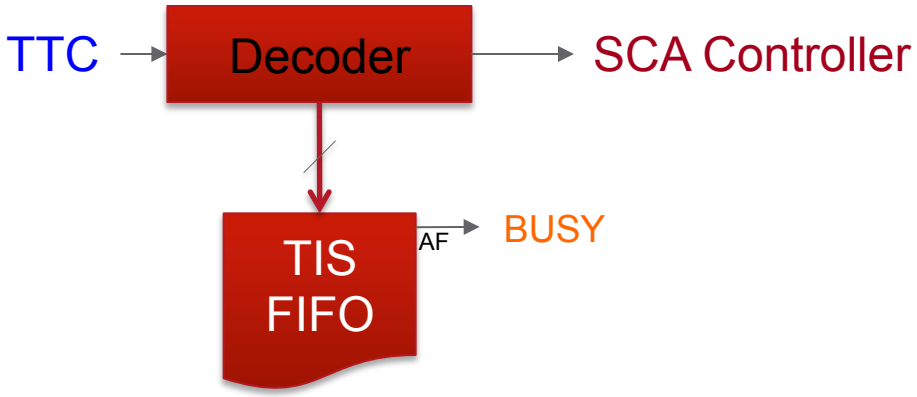
- $BUSY_{SRC}$ Plug-in
  - Collects and forwards back-pressure
- $BUSY_{DST}$ Plug-in
  - Handles back-pressure on the Trigger Simulator

TTC_{RX} PPI

SLAC

Wraps CERN's TTC_{RX} firmware in a PPI plug

Recovers Beam Crossing clock



Trigger Information Summary (*TIS*):

- L1-Accept ID
- Bunch Crossing ID
- Trigger Type
  - L1A
  - ECR
  - Calibration Strobe
- Orbit count

SCA Controller

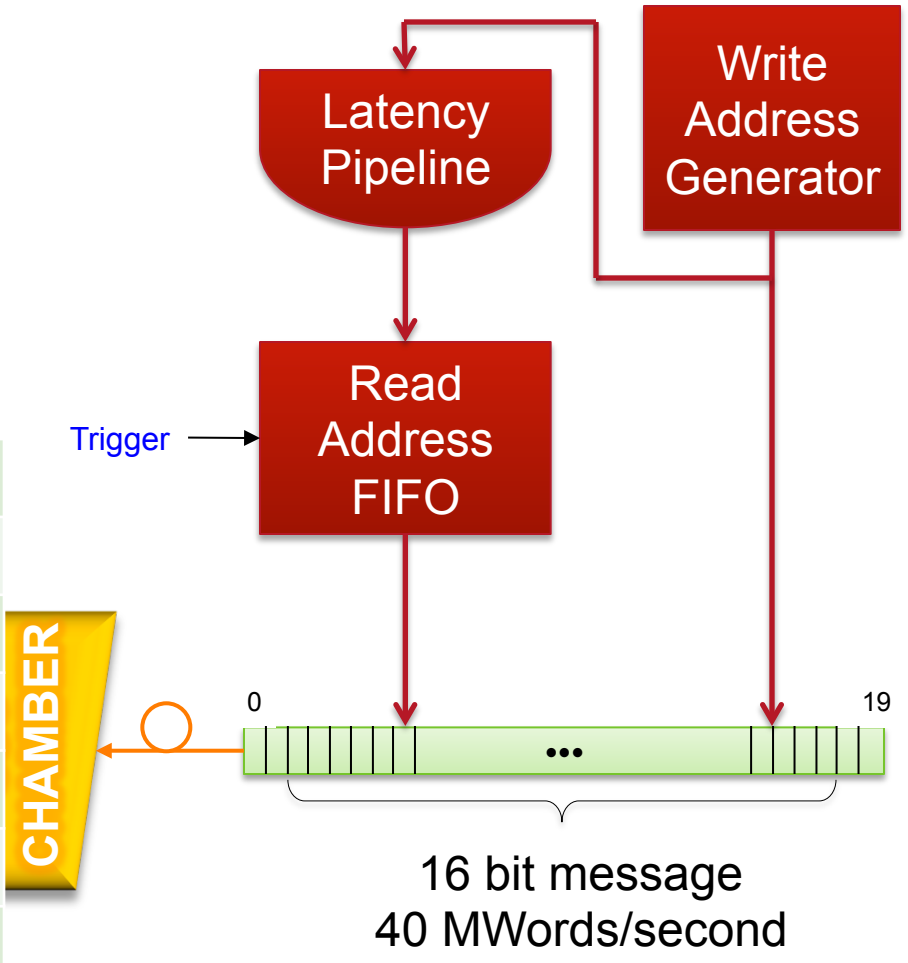
SLAC

Manages 144 SCA cells per channel

- Write Address Generator selects a cell for every (other) beam crossing
- Trigger (whether L1A or CalStrobe) causes Read Address to be emitted

40 MHz clock recovery on chamber

RD_CLK	5 or 6.67 MHz
SD	Serial read address
RD	Read enable
GA[1:0]	Part of read address
WA[7:0]	Write address
TRIG_DATA	Unused (?)
ADC_CLK	Phase shifted RD_CLK
WR_CLK	20 or 40 MHz



BUSY PPI

SLAC

SRC plug-in:

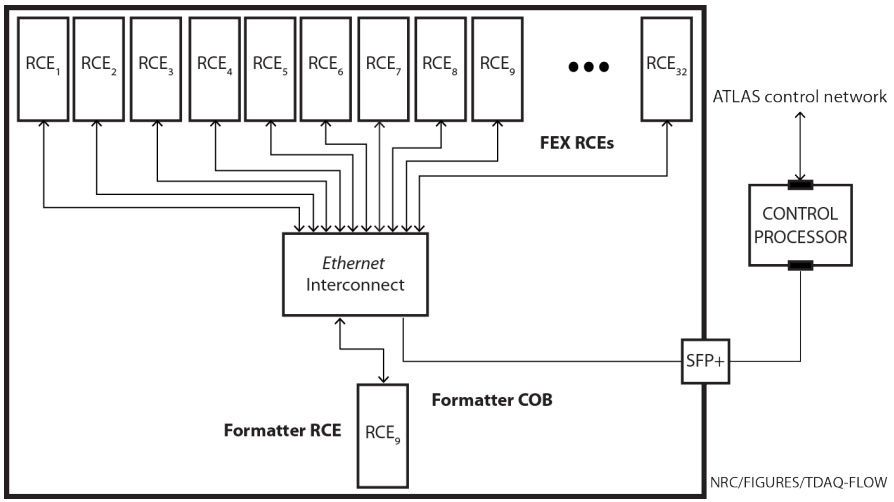
- Handles 4 sources of BUSY:
  - Input, FEX and TTC PPIs
  - Software command
- Individually maskable
- Forwards logical OR of source values to BUSY interconnect
- Provides statistics
  - Number of assertions
  - Fraction of total amount of time asserted

TDAQ Plane

SLAC

Control Processor:

- A “standard issue” Linux machine
- Takes role of RCC SBC in VME systems
  - VME replaced by ethernet
- Runs TDAQ software suite:
  - Run Control FSM
  - Configuration DB access
  - ERS for message reporting
  - OHS for histogramming
  - IS for publishing prompt data
- TDAQ proxy server to RCEs
- Could run the DHCP service to provide RCEs with their IP information
- Isolates NRC network traffic from ATLAS control network



RCEs:

- TDAQ proxy clients

Communication:

- Through TCP/IP ports and sockets

RCE Software Framework

SLAC

One time configuration of the software framework:

- Boots RTEMS from local on-board file system (FAT SD Card)
- Initializes the RCE (Most of the software work is here)
  - E.g., Connect to the Control Processor TDAQ proxies

TDAQ FSM proxy duties:

- Address requirements listed in Raul's talk
- Load a configuration
- Monitor run

Event Loop:

- Orchestrate data transfer between PPIs ("glue")
- Perform Cluster finding (FEX RCE)
- Event-build data from two chambers (Formatter RCE)
- Format data to ATLAS standard (Formatter RCE)
  - Identical event format as currently used

Development Infrastructure

Xilinx and GNU tools

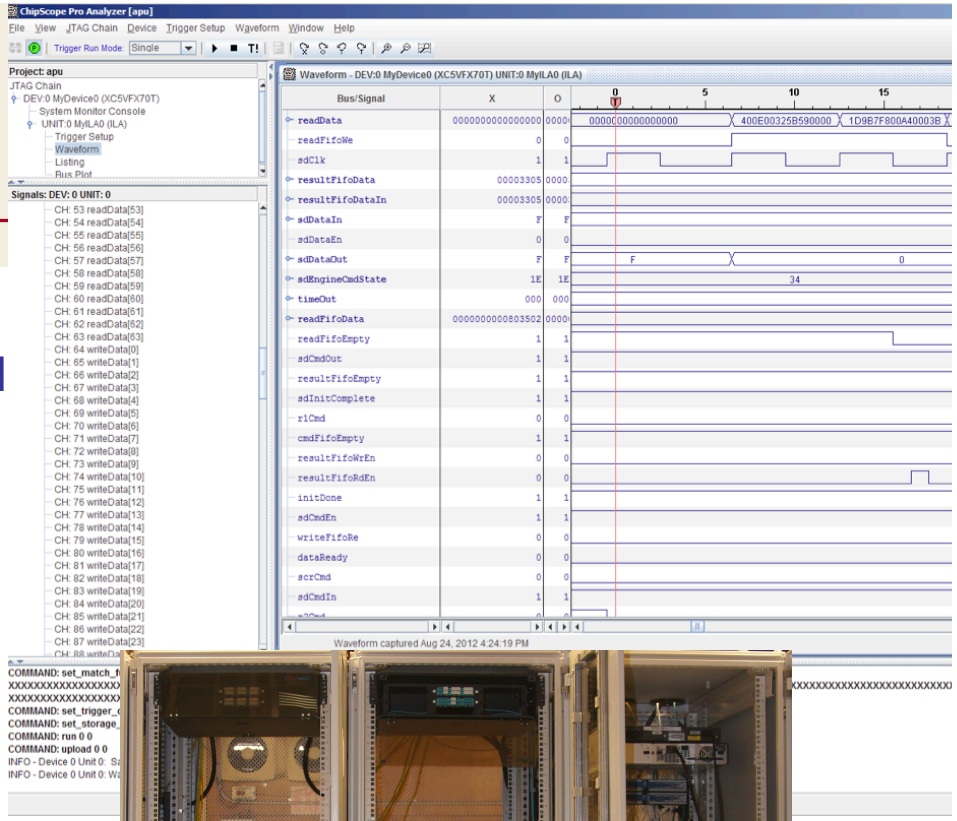
- Will put together a system level, mixed s/w and f/w simulation
- “Chipscope” tool for debugging f/w and the f/w / s/w interaction
- GDB: network and RTEMS thread aware

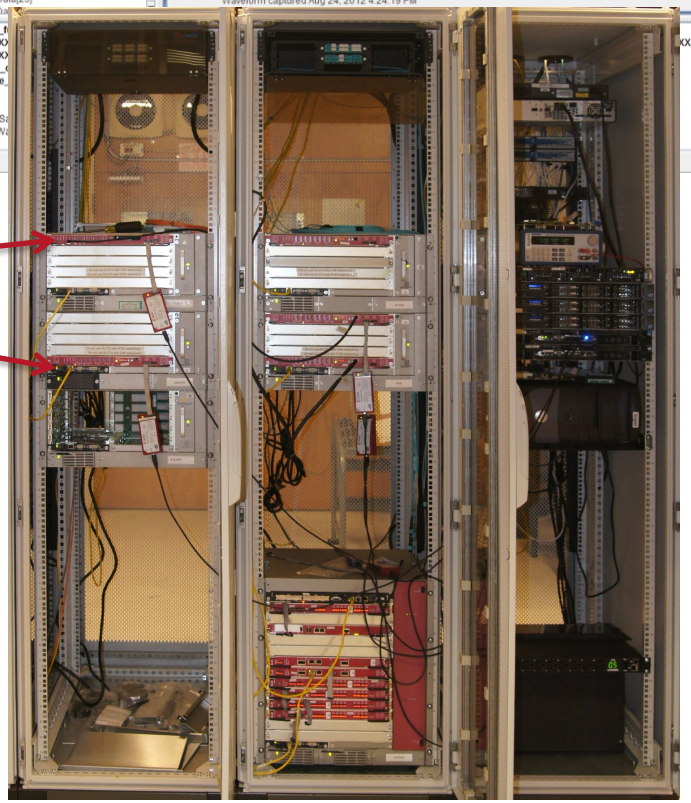
Construct test-stands

- Much can be done with a single, partially populated, COB

Resurrect TDAQ test-bed at SLAC (and instantiate one at UCI)

- To aid in development of proxies, etc.
- Decouple from TDAQ evolutionary path and Pt 1, as much as possible





Trigger Simulator

SLAC

We have LTPs available

Standard COB/RCE hardware can be used to play back trigger information

- Uses 9th RCE (DTM)
- Almost 4 GB available for the trigger pattern
- TTC_{TX} PPI is used to emit trigger messages
  - Inverse function of the TTC_{RX} PPI
- $BUSY_{DST}$ PPI is used to throttle triggers
- Allows each COB to stand alone as a separate system
  - Fine grained partitionability
  - Users can work in parallel (on separate COBs)
  - Decouples dependence on the LTP

CSC Emulator

SLAC

A COB/RCE based system can be used to simulate all or part of the CSC detector

- Hardware is identical to that used in the NRC (COB)
- Only firmware and software is different
- Here, simulate means to play back recorded data

Can use an existing RTM to drive a portion of the NRC

- Planning to build a “reverse” CSC RTM
  - Contains 8 SNAP-12 transmitters + 4 SNAP-12 receivers
  - 1 COB + “reverse CSC” RTM simulates up to 8 chambers
- Simple variation of standard CSC RTM design

4 COBs + 4 “reverse CSC” RTMs would allow testing the full system

Release Management

SLAC

- Will follow the Muon Group policy
- Will use the SVN Code repository at CERN to store both f/w and s/w
- A build system is available for compiling and installing f/w and s/w
- A standard regression test suite will be created to ensure reproducible results, reliability and performance
  - Each requirement (Raul's talk) will have an associated test
- We will maintain a rigorous deployment policy
- Will use Savannah for bug reporting
- Releases will be documented with release notes stored in the CSC twiki at CERN

Performance

SLAC

The NRC is a pipelined system consisting of 5 overlapping stages.
Its performance will be given by the slowest stage of this list:

Stage		Metric			Rate	Time
1	Input Data	4 time-slices * 192 channels * 12 bits			1.28 Gbit/sec.	**7.2 µS:** 140 KHz
2	Feature extract	Index bit array	192 ch * 5 layers / 64 bits per fetch	15 @ 3 CPI	475 MCPS	0.15 µS + 0.72 µS * <µ>/80
		TTC data	1 64 bit fetch	1 @ 3 CPI		
		Input buffer ptr	1 64 bit fetch	1 @ 3 CPI		
		Cluster finding	600 Bytes / 32 Bytes per cache-line	19 @ 18 CPI * <µ>/80		
		Output	42 Byte Enet hdr / 8 Bytes per inst.	6 @ 3 CPI		
		Output buffer ptr	1 64 bit store	1 @ 3 CPI		
		Total:		72 + 342 * <µ>/80		
3	FEX - Fmtr	600 bytes * <µ>/80			10 Gbit/sec.	0.6 µS * <µ>/80
4	Format	600 bytes * <µ>/80 * 2 chambers			1 GByte/sec.	1.2 µS * <µ>/80
5	ROL	600 bytes * <µ>/80 * 2 chambers			160 MByte/sec.	**7.5 µS** * <µ>/80

Up to a <µ> of ~80 the NRC is limited by the On-Detector Electronics to 140 KHz.
Above <µ> of ~80, the ROL takes over. Could double the output rate by doubling the ROLs.

Summary

SLAC

- Will deliver a “plug and play” solution.
  - ✓ System conforms to Event, Trigger, BUSY, TDAQ interfaces
  - ✓ Data format won’t (need to) change
- This solution is achieved through Protocol Plug-ins.
- The NRC software framework is hosted on the RCEs.
- The TDAQ software & proxies is hosted on the Control Processor.
- Algorithmically, NRC operation is identical to the current ROD Complex’
- Development and validation tools verify requirements enumerated in Raul’s talk.
- Will implement a disciplined release management plan that is consistent with ATLAS policy.
- The NRC firmware and software design meets its performance requirement to operate at an L1Accept rate of 100 KHz with a <μ> of 80.

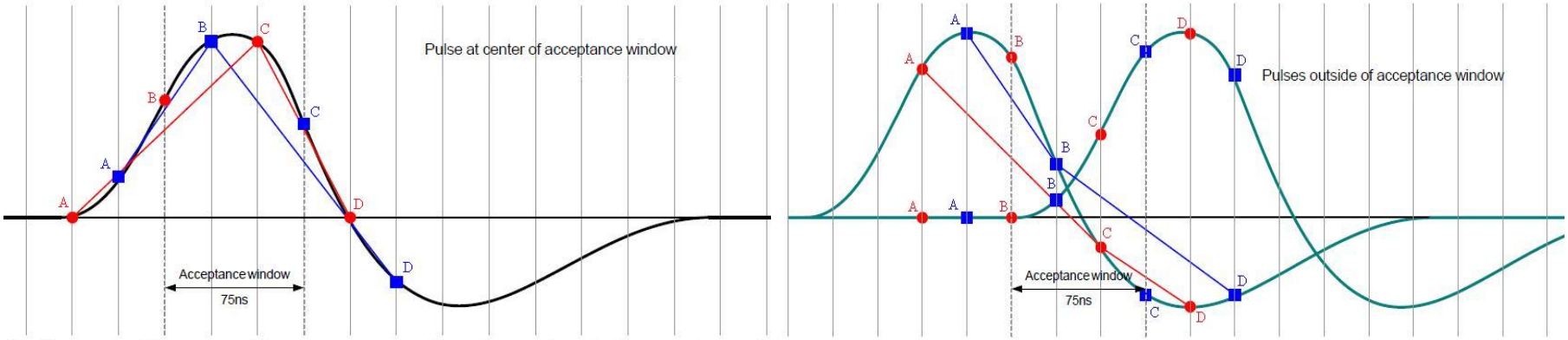
Backups

Feature Extraction

Threshold cut

Each individual channel is associated with a unique threshold value. The sample is accepted only if it is larger than the corresponding threshold.

Time cut



In-time pulse (lines every 25 ns, squares are phase B samples, circles are phase C)

Out-of-time pulses (lines every 25 ns, squares are phase B samples, circles are phase C)

The sampling is adjusted so that the nominal peaking time for in-time hits falls halfway between the B sample of the later sampling and the C sample of the earlier sampling. The nominally largest sample (B or C) must be larger than sample A and D.