ATLAS Detector Readout with FELIX

Jinlong Zhang

(Argonne National Laboratory)

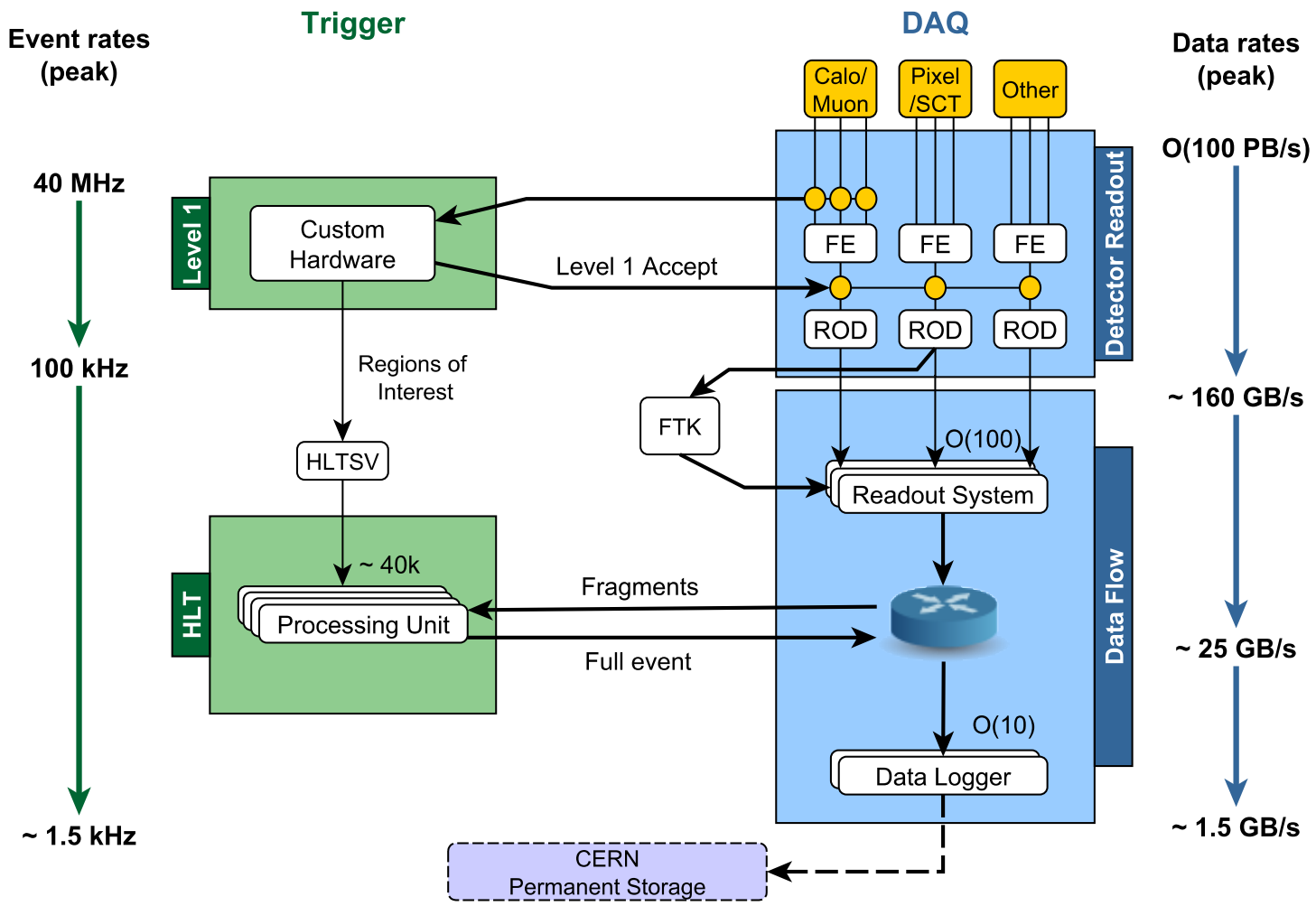
for the ATLAS Collaboration

TIPP'17, May 22-26, 2017, Beijing

Introduction

- ATLAS TDAQ system
- Front-End LInk eXchange (FELIX) in Phase-I upgrade
- FELIX in Phase-II upgrade
- Summary

Run 2 TDAQ Architecture



TDAQ Operating Parameters

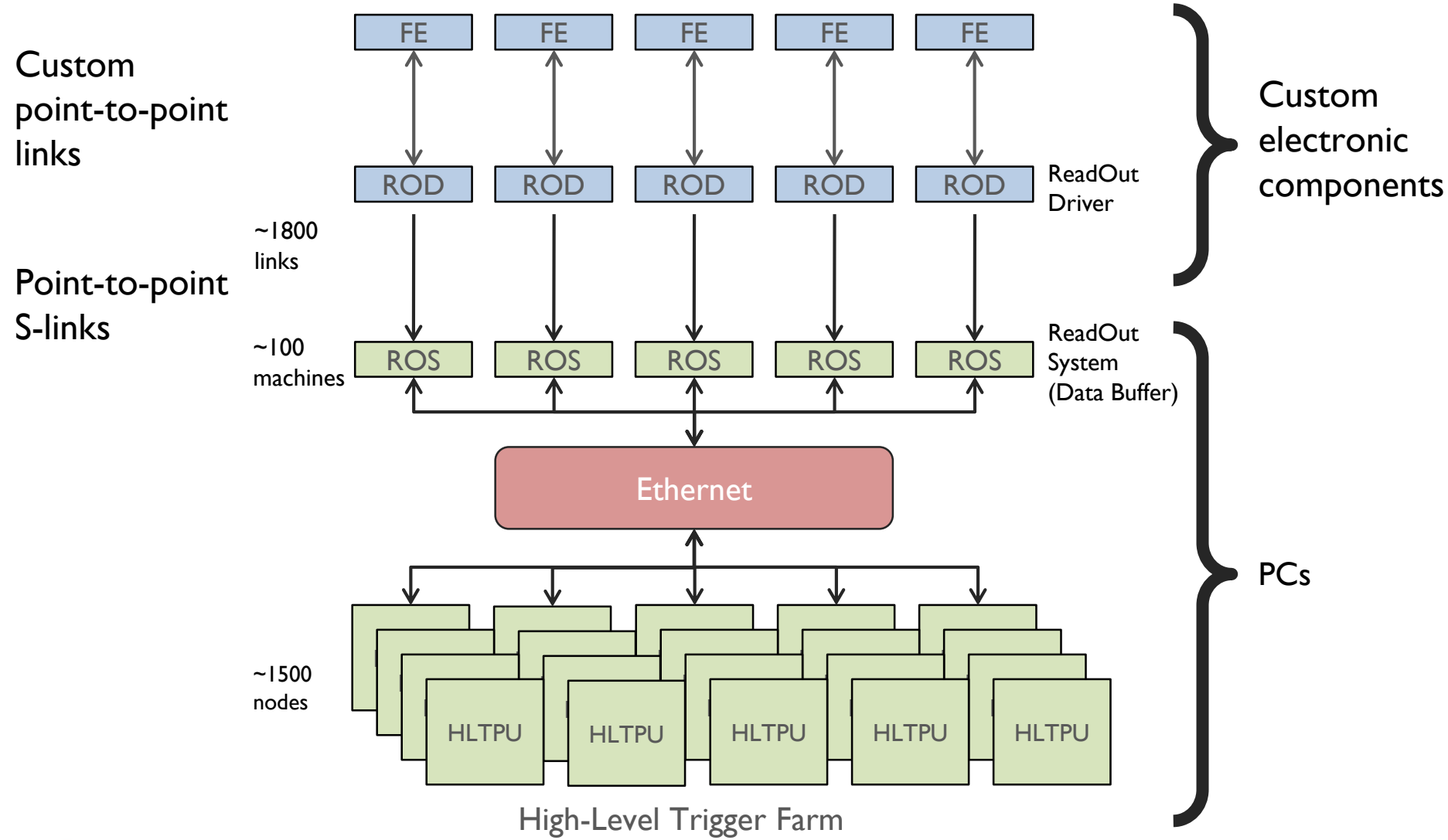
	# Trigger levels	Rates (kHz)		Event Size (MB)	Network Bandwidth (GB/s)	Storage	
						GB/s*	kHz
Run 1	3	L1 (L2+EF)	75 ~0.4	~1	10	0.5	~0.4
Run 2	2	L1 HLT	100 1	~2	50	~1	~1
Run 3	2	L1 HLT	100 1	~2	50	~1	~1
Run 4	2*	L0 HLT	1000 10	~5	~5000	~50	~10

- **Phase-I upgrades for Run 3**
  - **Level-1 Calorimeter trigger (L1Calo) with fine granularity LAr data**
  - **Level-1 Muon trigger (L1Muon) with New Small Wheel data**
  - **New readout with Front-End Link eXchange (FELIX)**

Readout Evolution Motivation

- Higher level of commonality between detectors
  - A common object providing functionalities today implemented in detector-specific back-end custom electronics (ROD)
- Increased use of COTS components
  - all ROD-like functionality (including data processing) could most likely be implemented in standard computers by Phase-II
- Performance scalability built-in
  - Programmable connectivity between detector FE and DAQ
- Capability to disentangle ROD-like functions from hardware implementation
  - Different detector granularity for monitoring, control, data handling …
  - Detector Control System (DCS) and DAQ traffic separation

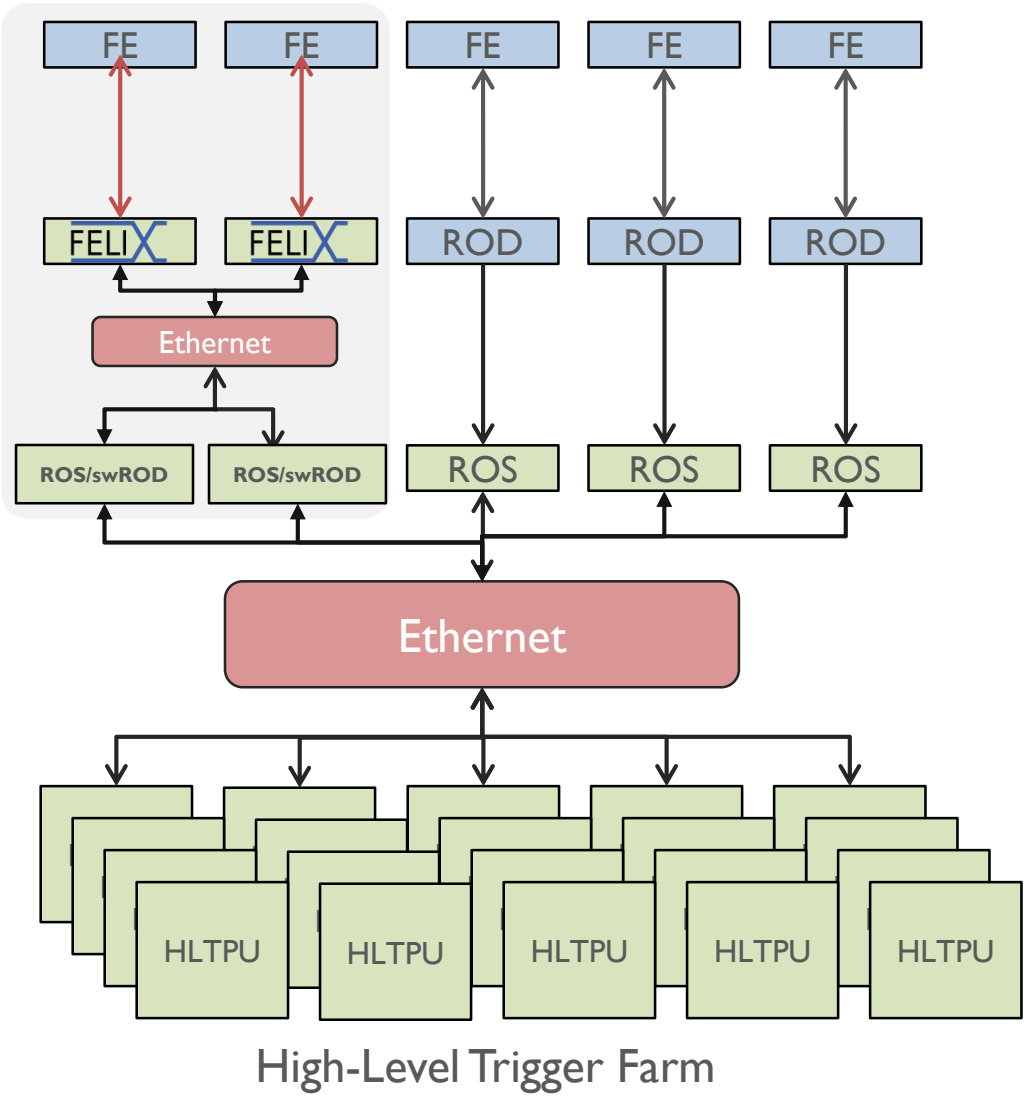
Readout (Run 2)



Readout (Run 3)

Point-to-point
(GBT, ad-hoc)

Ethernet
Infiniband, etc

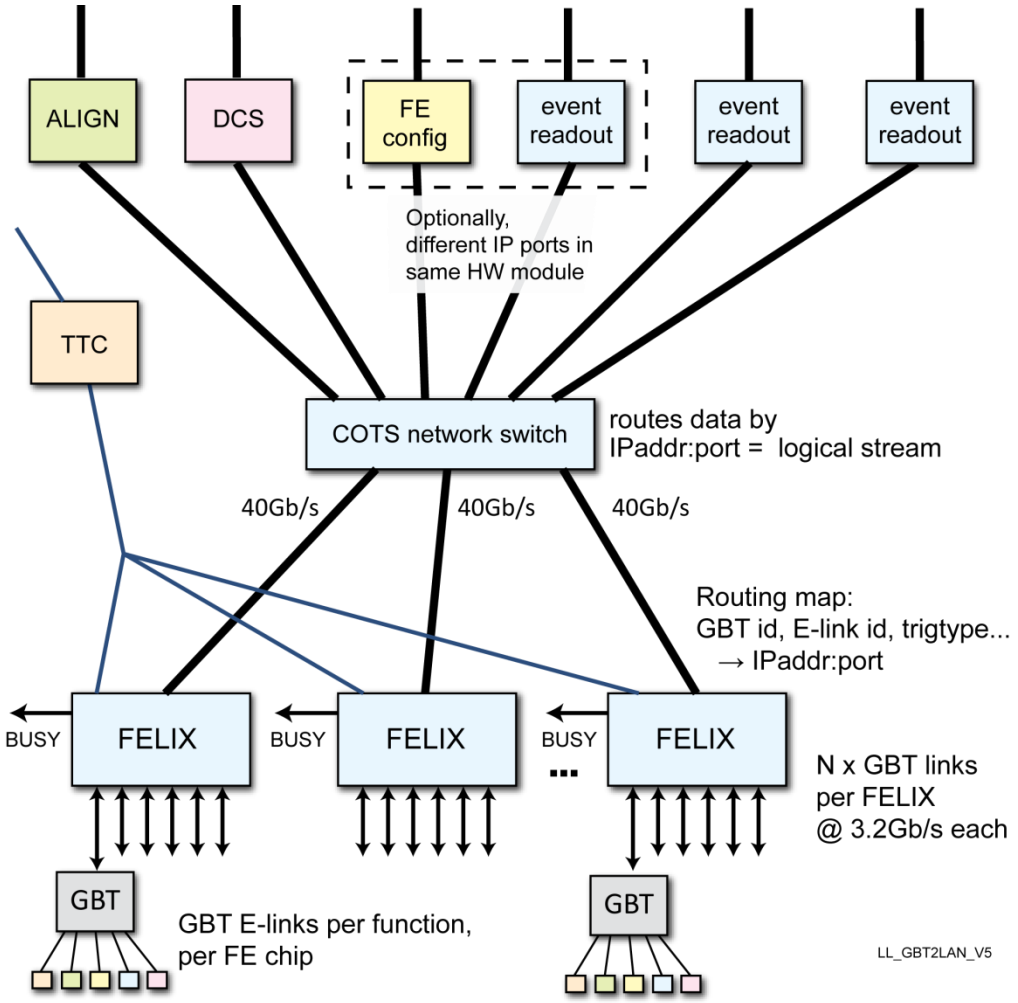


Custom electronic components

PCs

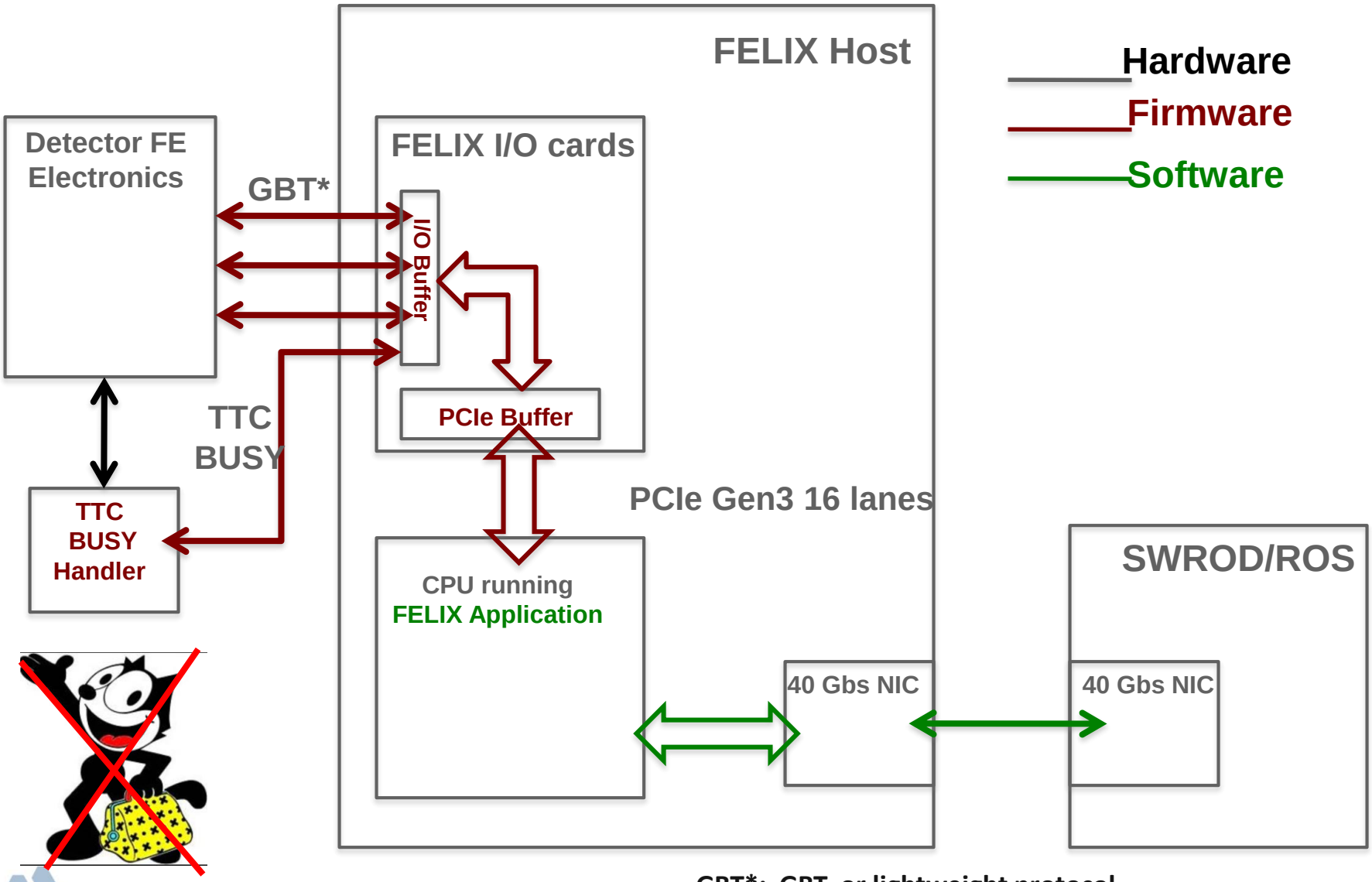
FELIX

- Enabling transition from custom hardware to COTS as early as possible in the readout path
- Using high level switch protocols of high speed and large bandwidth
- Configurable and flexible data routing and error handling, without relying on detector specific hardware
- Direct low latency path between links if needed
- Universal Timing, Trigger and Control (TTC)/BUSY handling as for Run 1&2
- Command scheduling with guaranteed timing if needed



GBT: GigaBit Transceiver (GBT), radiation hard data transmission link developed by CERN

FELIX as a System



GBT*: GBT or lightweight protocol

FELIX Development

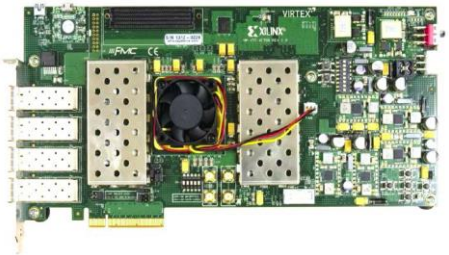
TTCfx
Custom FMC accepting TTC input
Outputing TTC clock and CH A/B info

Data emulator cards (GLIB, KC705, ZC706, HTG-710) and GBT devices (GBTx chip card and VLDB)

Hitech Global HTG-710
24 10Gbps links(2 CXP
Virtex-7 X690T
PCIe Gen 3 x 8 lanes

Mellanox Dual-port 40 Gb
Ethernet or Infiniband

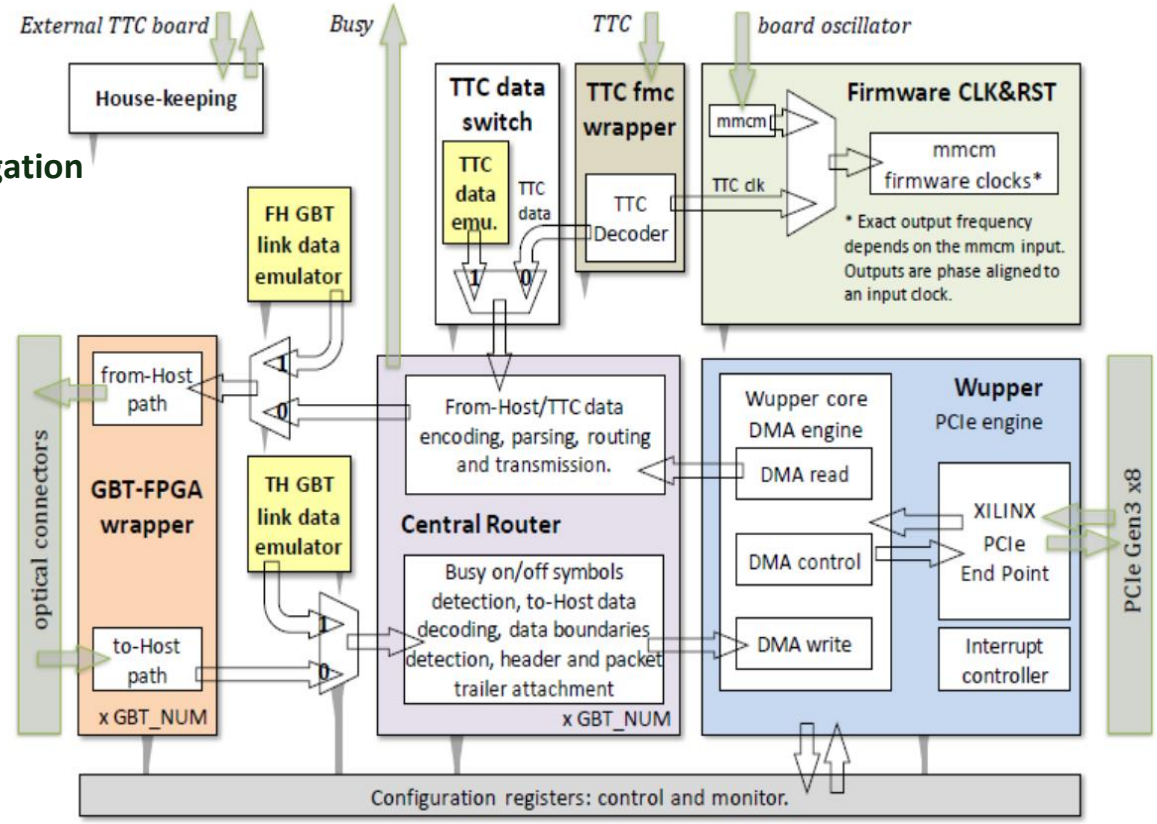
Xilinx VC-709
4 10 Gbps links(4 SFP+)
Virtex-7 X690T
PCIe Gen 3 x 8 lanes



BNL-711
48 14Gbps links(4 miniPOD RX & Tx)
Kintex Ultrascale
PCIe Gen 3 x 16 lanes
Two DDR4 SODIMM up to 16GB
Onboard: Clock conditioner, TTC receiver, BUSY lemo output

Firmware

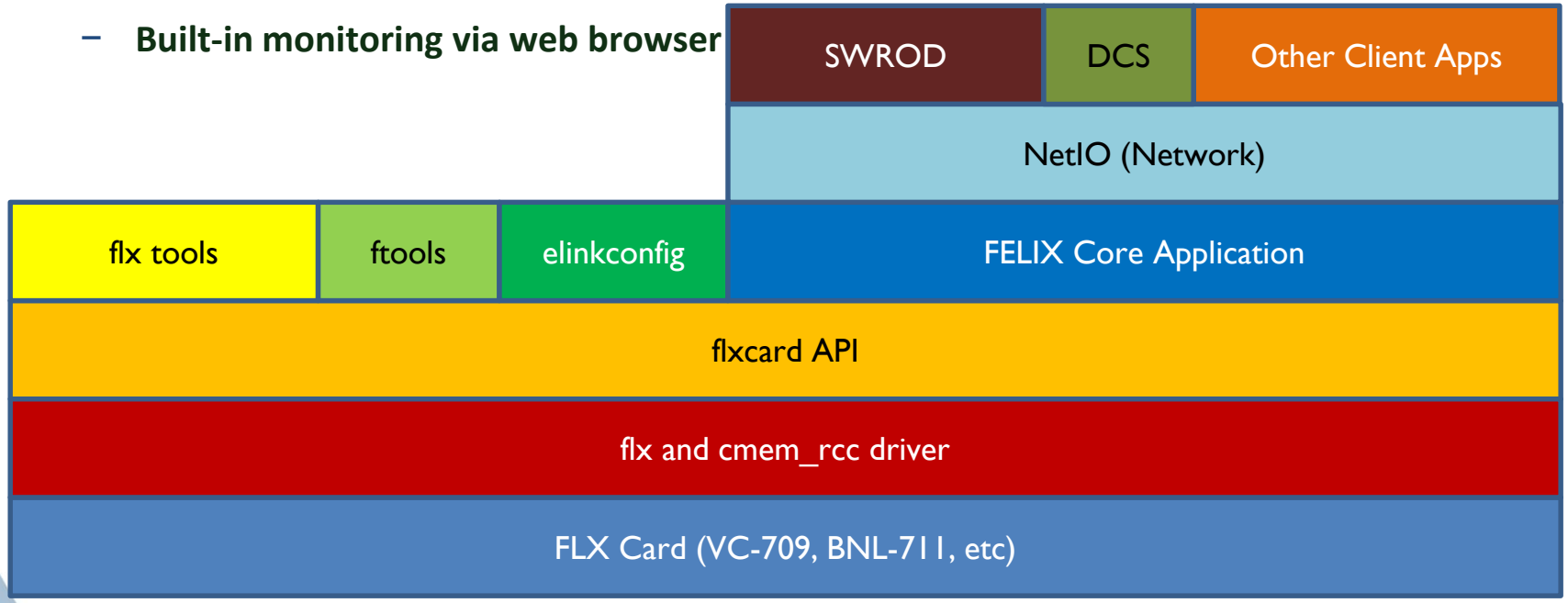
- I/O protocol
  - GBT (4.8 Gbps)
  - Ful mode lightweight protocol for high bandwidth between FPGA to FPGA (9.6 Gbps)
- TTC
  - Fan out
  - BUSY aggregation and propagation
- Central Router
  - Path from FE to host
  - Path from host to FE
- PCIe
  - Wupper PCIe engine (OpenCores project)
- Top level
  - 24-Ch GBT
  - 12-Ch full mode
  - 48-Ch TTC fanout



Block diagram for one of the two PCIe endpoints

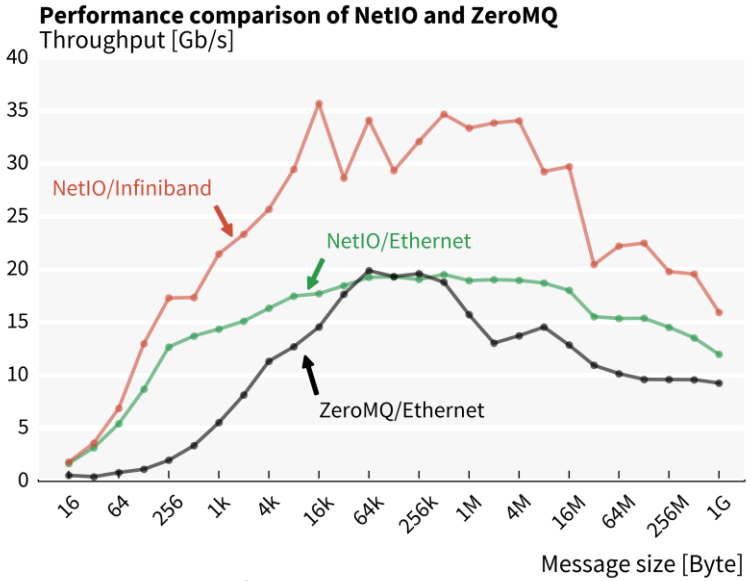
Software

- Drivers
- API
  - Low level object oriented library
- Core application
  - Reads and writes data in block form
  - Outputs and Inputs data over NetIO
  - Built-in monitoring via web browser
- Configuration tools
  - Register mapping
  - Graphic Configurator
  - … …
- Standalone testing tools

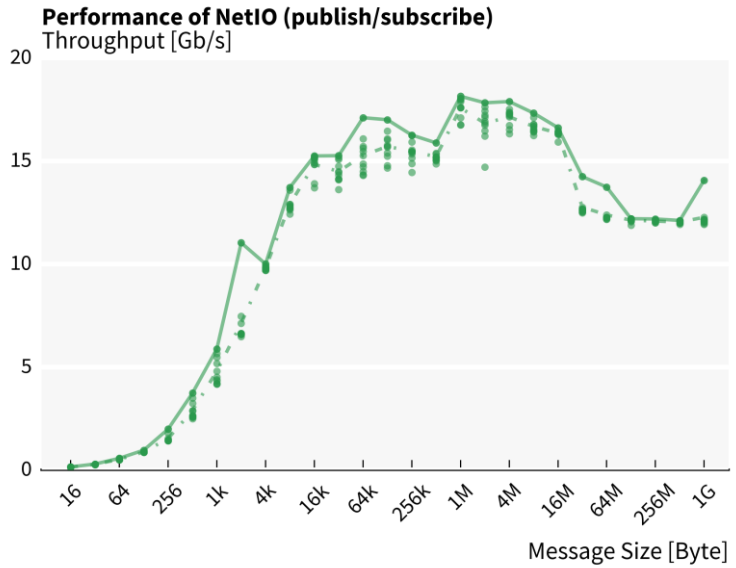


NetIO

- NetIO (generic message-base networking library) developed
  - Providing the FELIX network stack
  - Different communication patterns
    - Low latency or high throughput
    - Point-to-point or publish/subscribe
  - Different network back-ends
    - POSIX sockets (TCP/IP over Ethernet)
    - Libfabric (Infiniband and others)
- Performance measured against reference ZeroMQ
  - NetIO over Ethernet outperforming ZeroMQ for smaller and larger than ~100 kB size
  - NetIO over Infiniband outperforming both NetIO over Ethernet and ZeroMQ



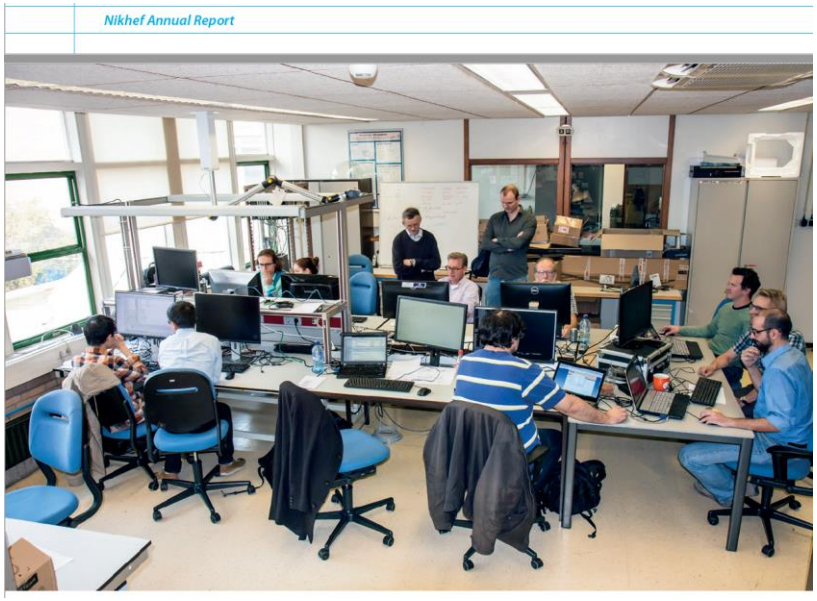
Typical ATLAS fragment size ~1 KB



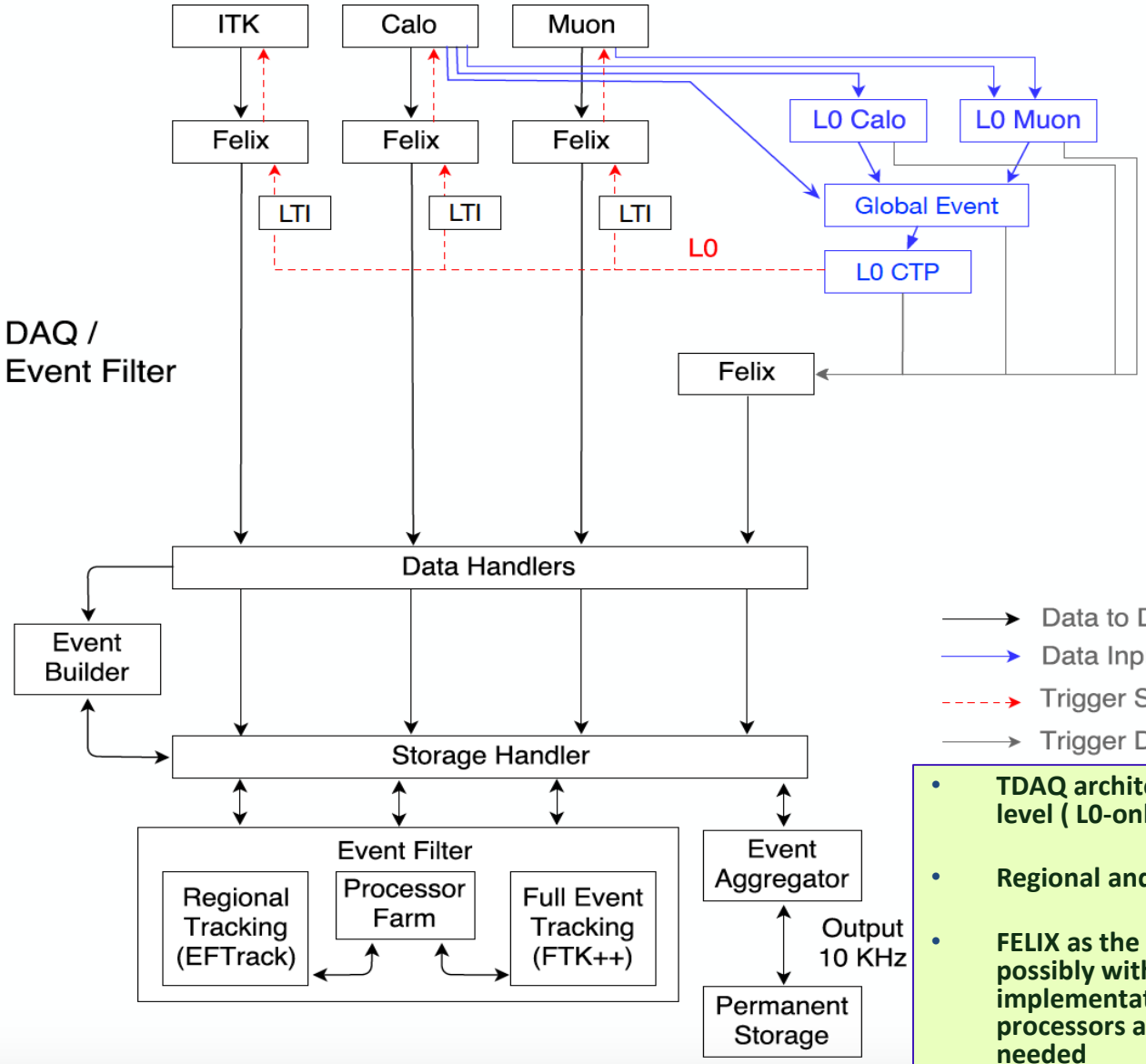
FELIX in Progress

- Full chain functionality demonstrated and several iterations of integration
- Supporting a variety of detector test setups
- Being adopted by other experiments (proto-DUNE and more coming)

- Prototype I/O card available and being tested
- All major firmware components available, debugging and scaling up
- Stable software releases and performance benchmarking



Run 4 TDAQ Architecture



Readout Parameters	
Link from detectors	~20000
FELIX I/O card	~530
FELIX PC servers	~270
FELIX NICs (100 Gbps)	~530

- **TDAQ architecture baseline with one hardware trigger level (L0-only)**
- **Regional and full event tracking at Event Filter (EF) level**
- **FELIX as the readout for all detector subsystems, possibly with new hardware/firmware/software implementation, and with low latency link to trigger processors and detector specific firmware/software if needed**

Summary

- **Current ATLAS DAQ system performing well while upgrades progressing as planned**
  - **Phase-I projects on schedule**
  - **Phase-II upgrade Technical Design Report in Q4 2017**
- **Increased use of commodity hardware**
  - **Transit as early as possible from custom rad-hard links to commodity network (FELIX)**
  - **Take advantage of arising technologies**