

# FELIX: The new Detector Readout System for the ATLAS Experiment



Marco Trovato on behalf of the ATLAS Collaboration

Argonne National Laboratory



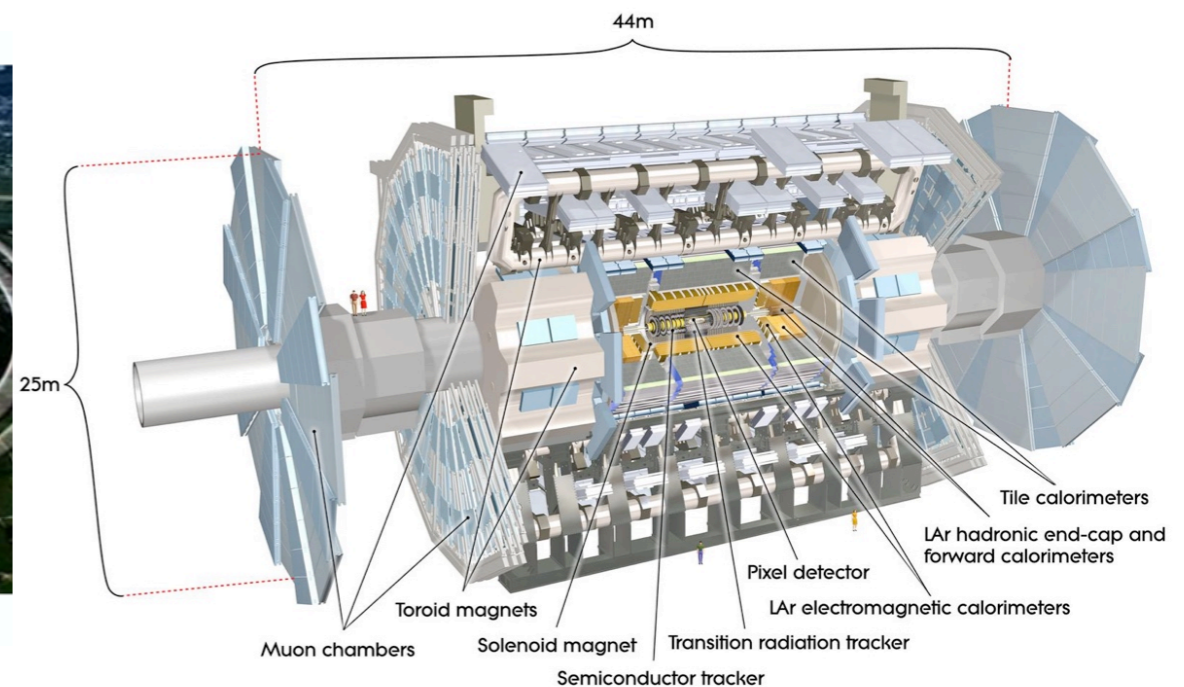
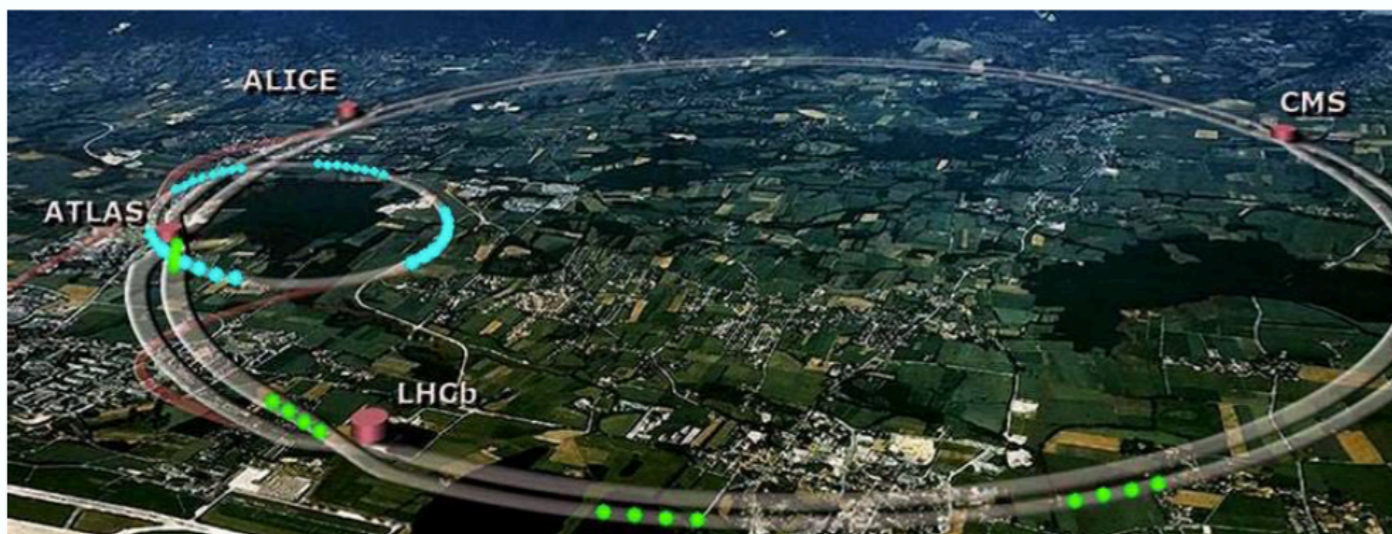
U.S. DEPARTMENT OF  
**ENERGY**

Argonne National Laboratory is a  
U.S. Department of Energy laboratory  
managed by UChicago Argonne, LLC.

01.Oct.2019, INFN-Pisa

# Experimental Context: ATLAS @ LHC

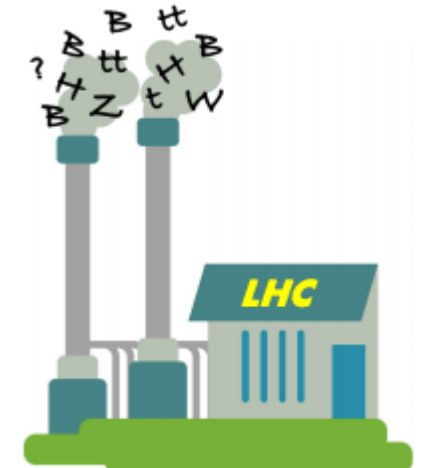
- The Large Hadron Collider (LHC) is the most powerful accelerator in the world
- LHC is 27 Km long collider located beneath the France-Swiss border near Geneva
- Four major detectors on the LHC: ATLAS, ALICE, CMS, LHCb
- ATLAS is 46 meters long, 25 meters in diameter and 7,000 tonnes in weight



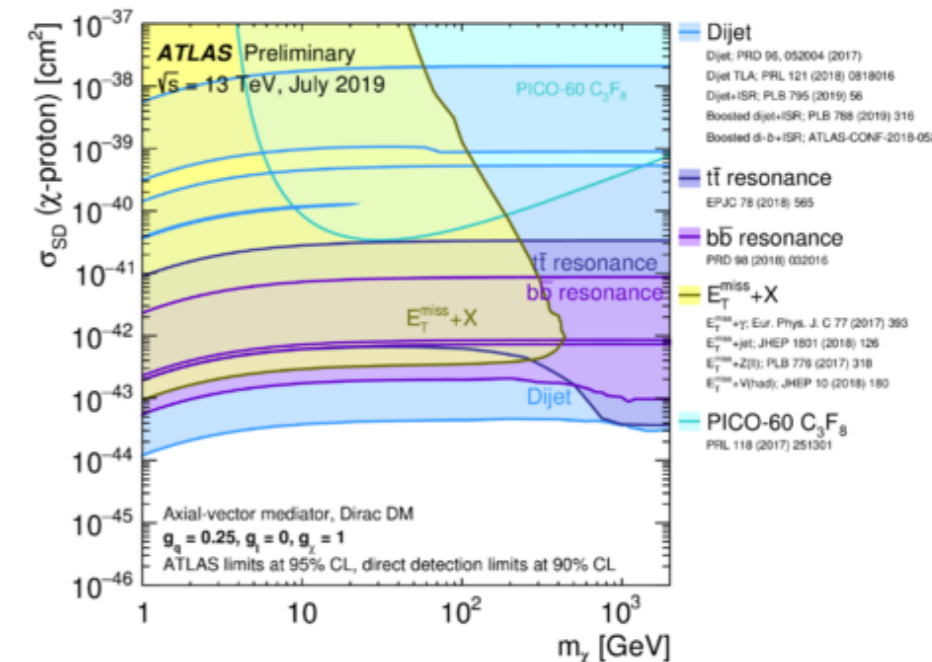
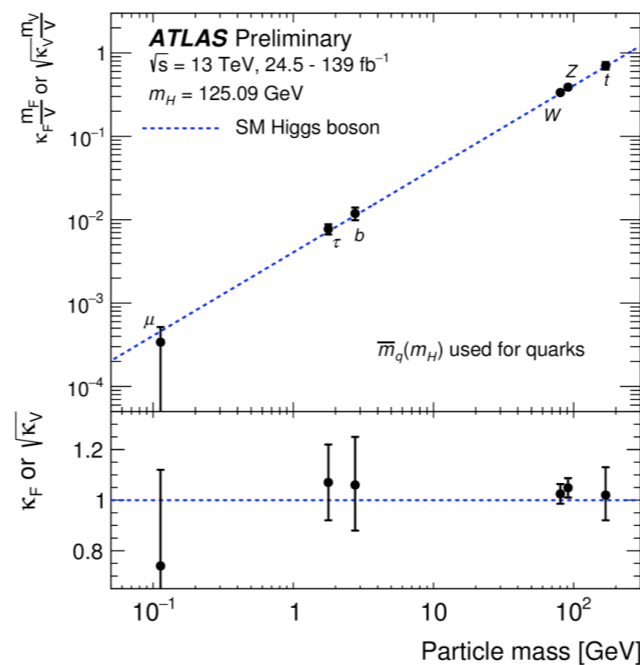
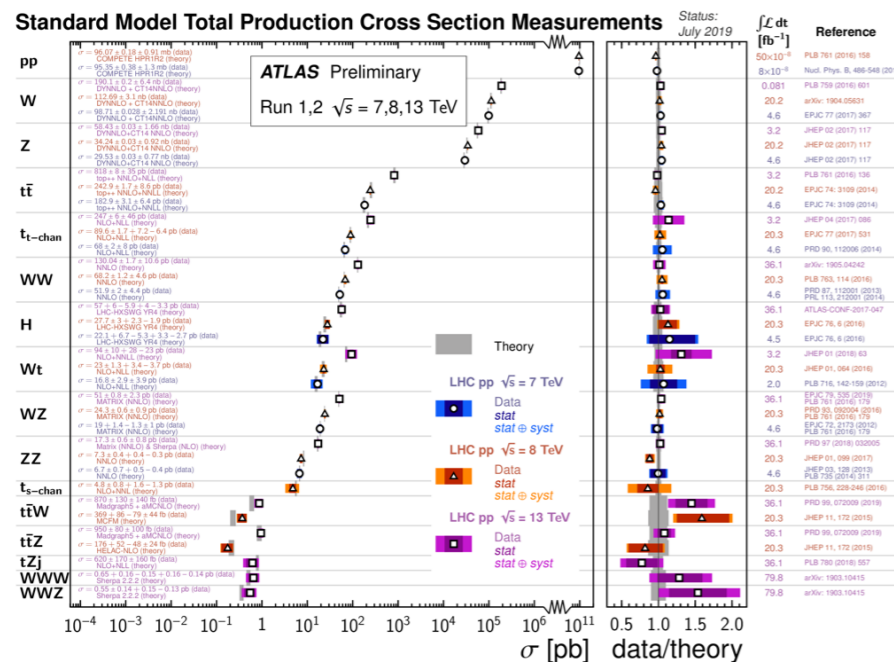
# ATLAS Results Run2 (2015-2018)

## LHC is an EVERYTHING factory

Particle	Produced in $139 \text{ fb}^{-1}$ at $\sqrt{s} = 13 \text{ TeV}$
Higgs boson	7.7 million
Top quark	275 million
Z boson	2.8 billion ( $\rightarrow \ell\ell$ , 290 million)
W boson	12 billion ( $\rightarrow \ell\nu$ , 3.7 billion)
Bottom quark	$\sim 40$ trillion (significantly reduced by acceptance)

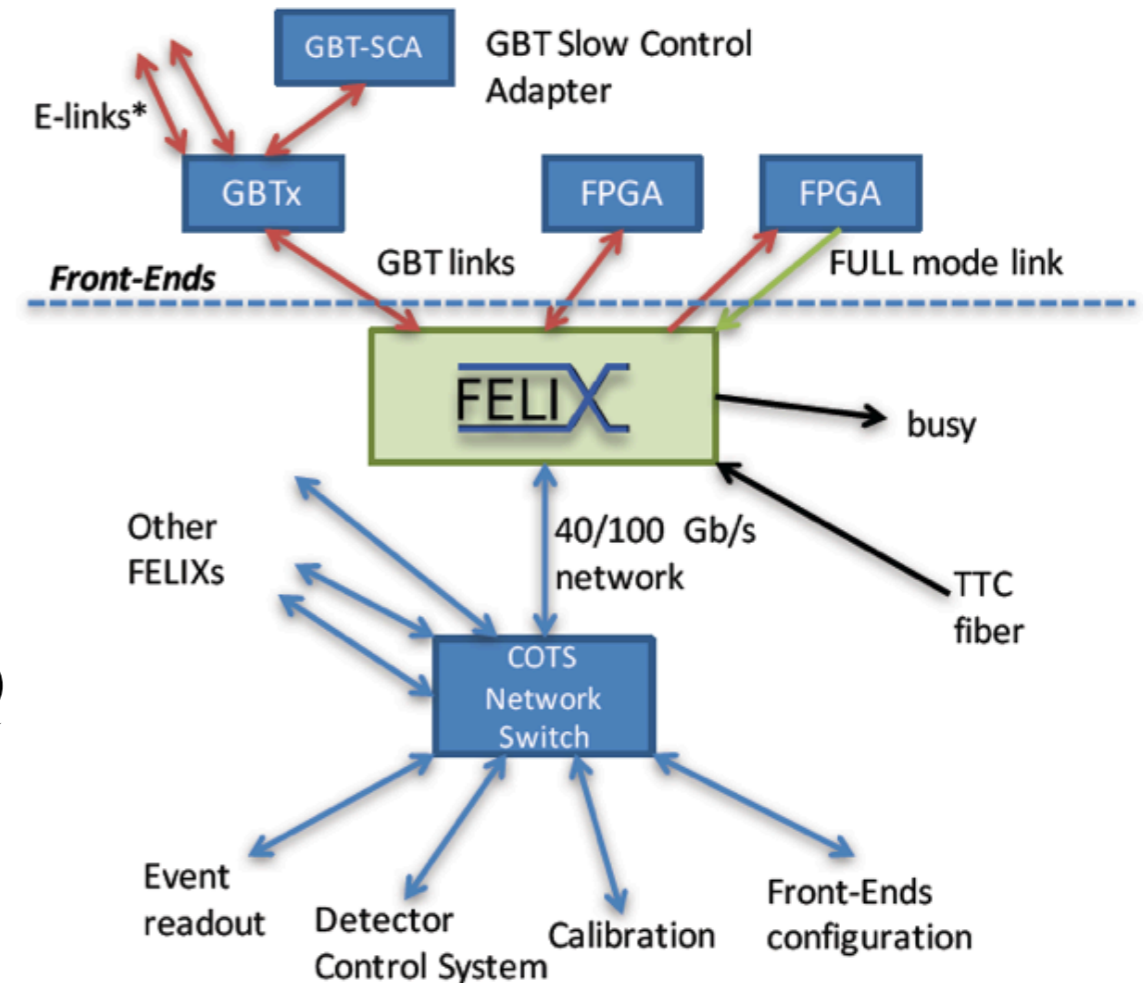


## The broad physics potential is UNDOUBTED



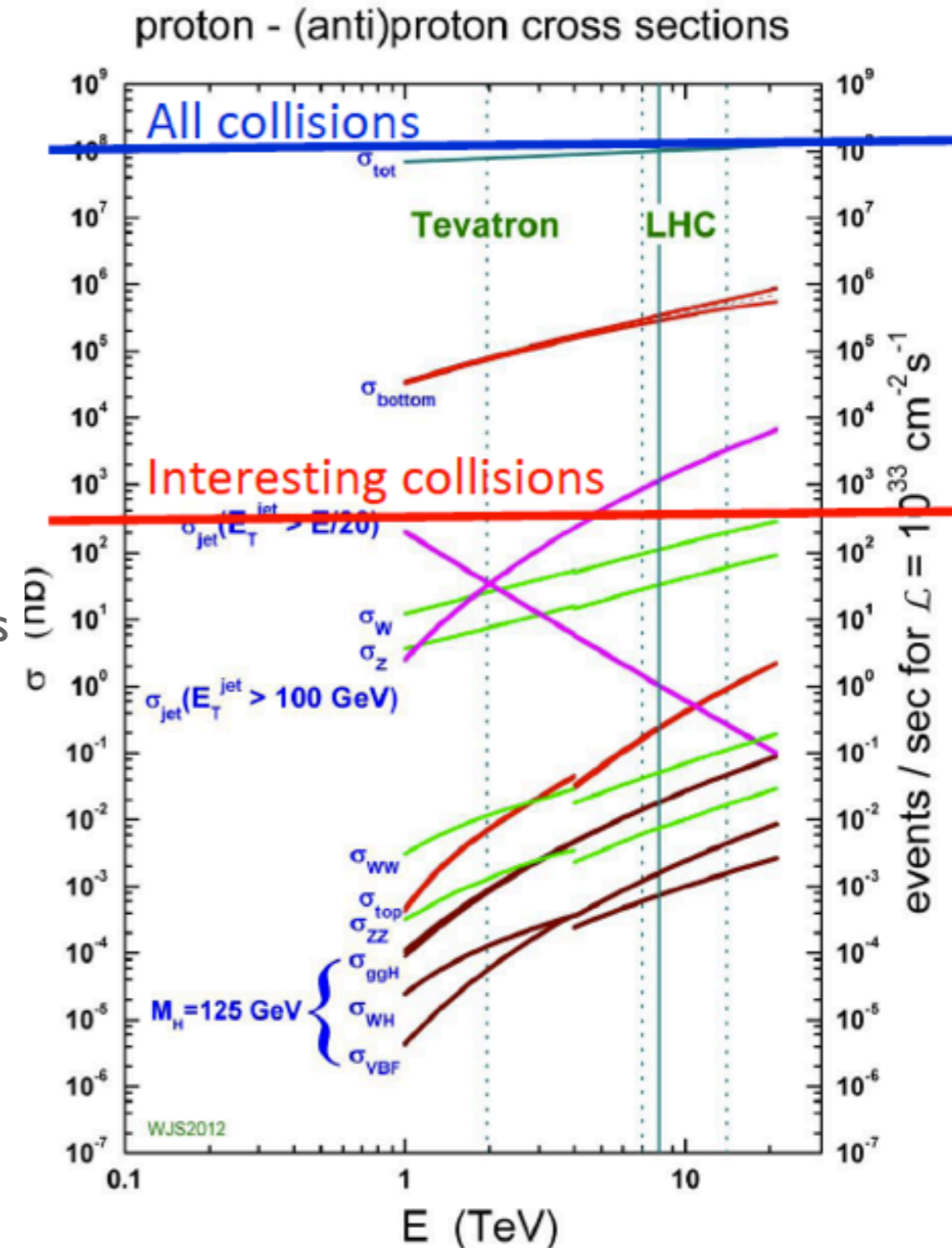
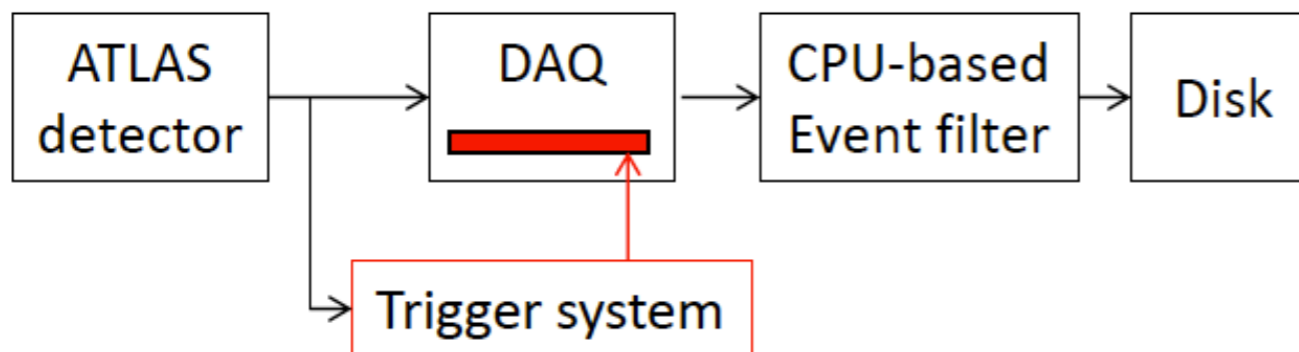
# Outline

- ATLAS trigger and data acquisition (TDAQ) system: current and next
- Front End Link Exchange (FELIX):
  - ▶ functionality
  - ▶ hardware components
  - ▶ firmware and data formats
  - ▶ software
  - ▶ integration and performances
  - ▶ utilization in ATLAS
- ANL involvement in Phase2 TDAQ and pixel readout
- Tests and results
- Conclusions and outlook



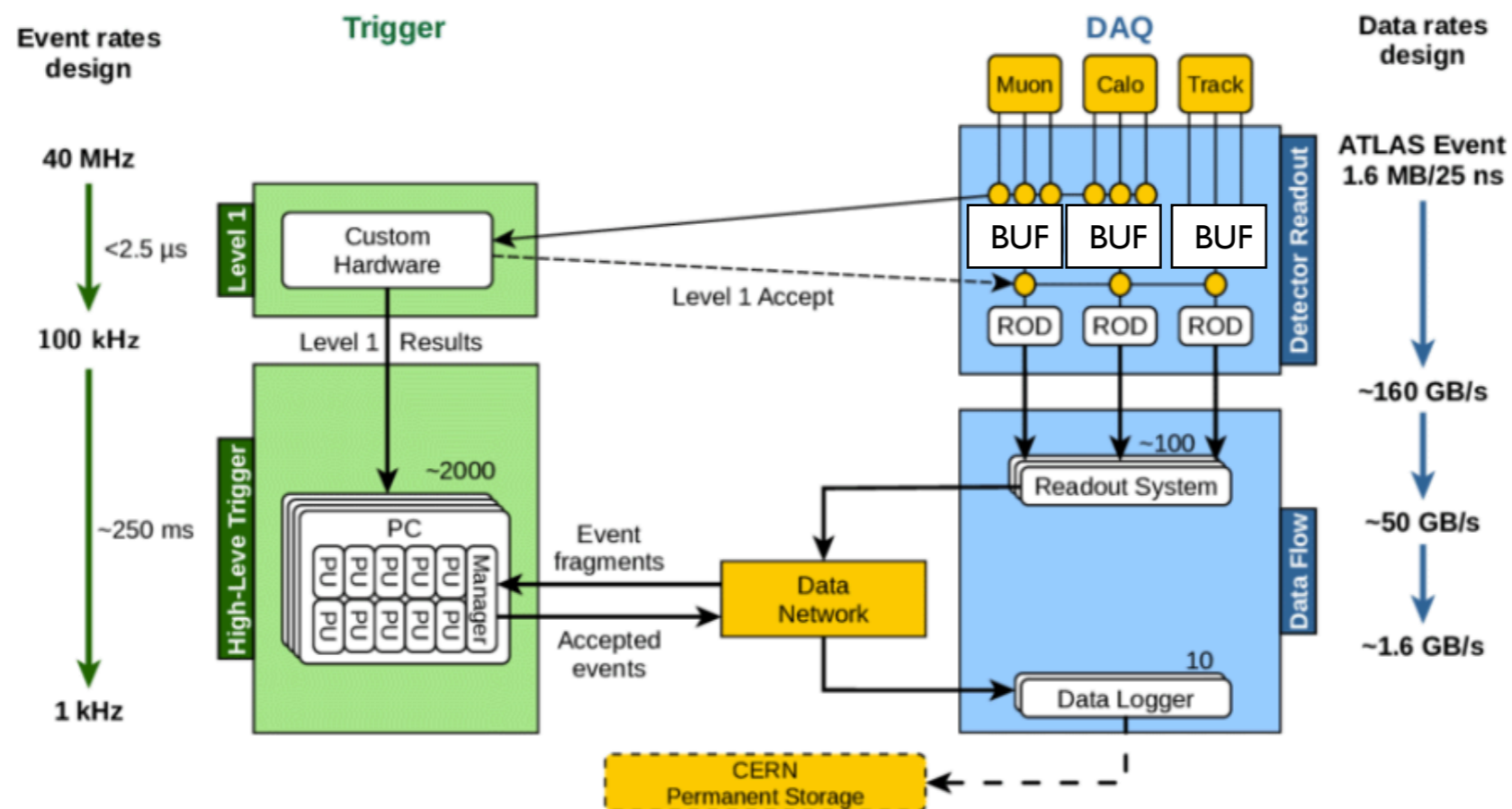
# Trigger and Data Acquisition System

- ~1 out of 100,000 collisions is interesting
- The Trigger/DAQ system filters the entire data volume to keep data from interesting events
  - ▶ The trigger system performing real time selection is efficient
  - ▶ The DAQ system buffer data while the trigger is analyzing it



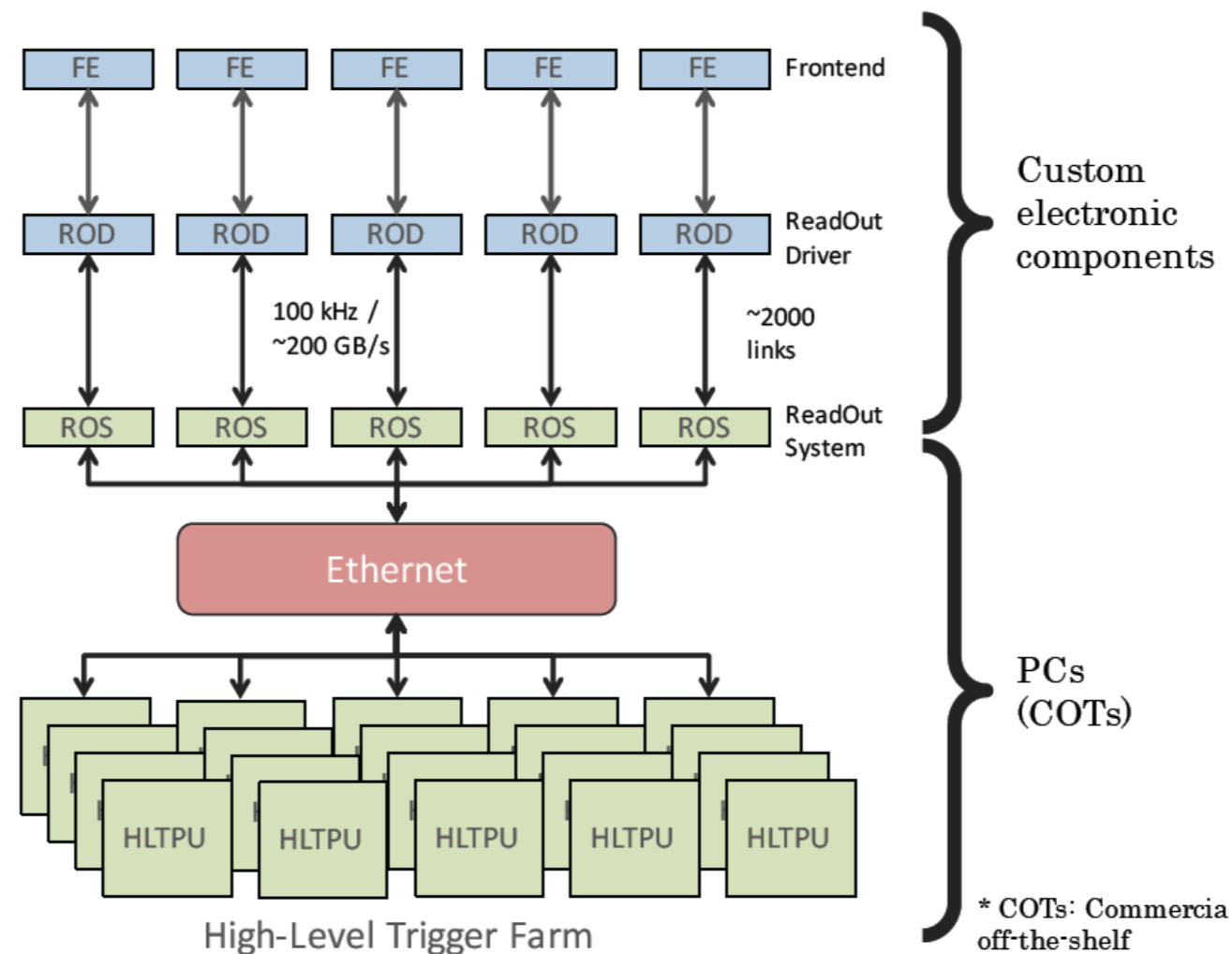
# ATLAS Trigger/DAQ in Run2

- With respect to Run 1:
  - ▶ level 1 rate increased to 100 kHz
  - ▶ HLT rate to 1-1.5 kHz
  - ▶ Because of higher trigger rates and larger instantaneous luminosities ( $\sim 2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ) event throughput to permanent storage increased to 1.5 GB/s

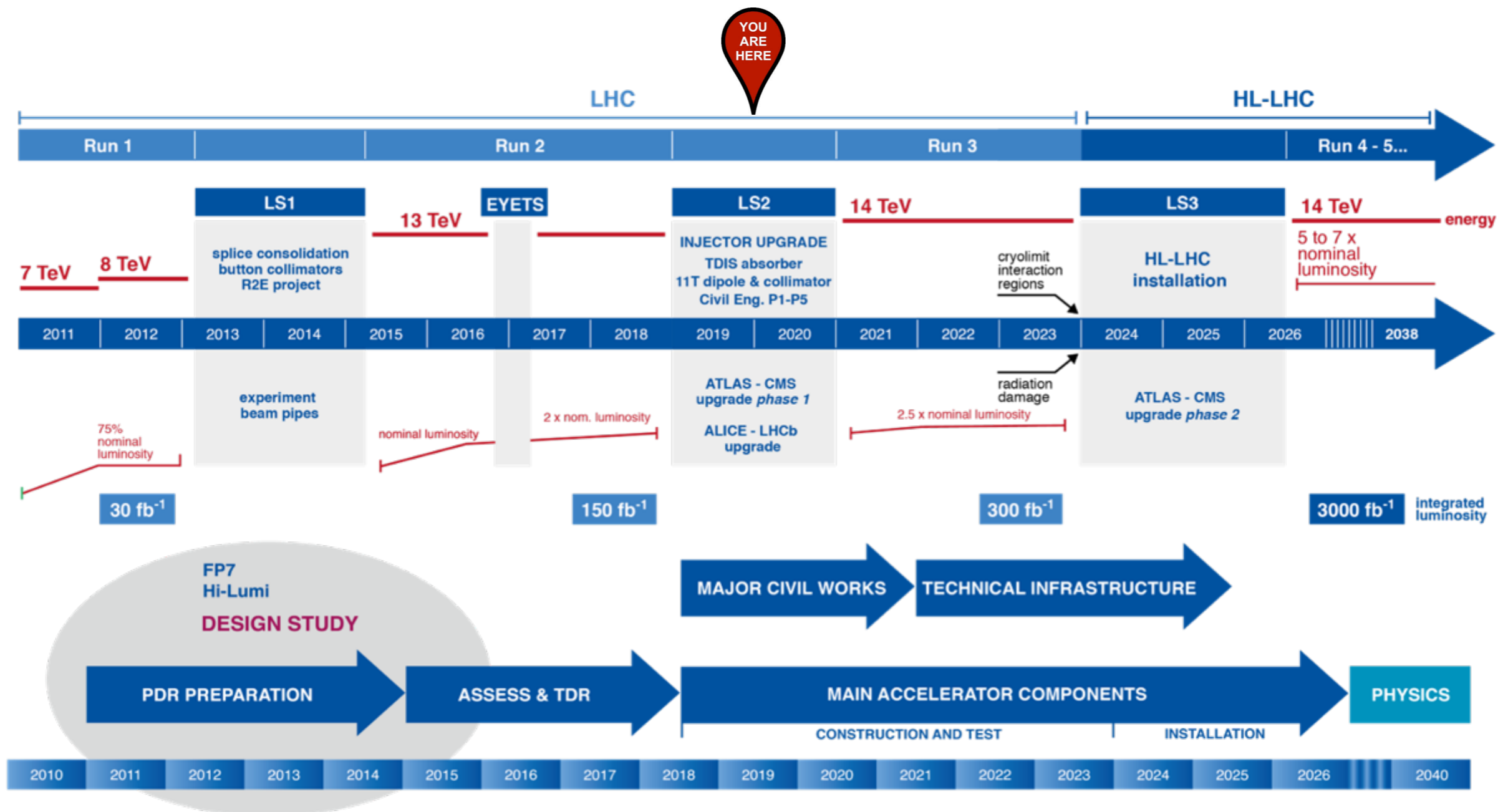


# ATLAS Trigger/DAQ in Run2

- Custom HW/protocols are used for the front-end (FE) readout
- Data is buffered in the FE electronics while L1 trigger decides
- Trigger and LHC clock are sent to both FE and off-detector Read-Out Drivers (ROD)
- Read-out systems (ROS) receives data from RODs
- High-Level trigger (HLT) finalized trigger selection in two steps



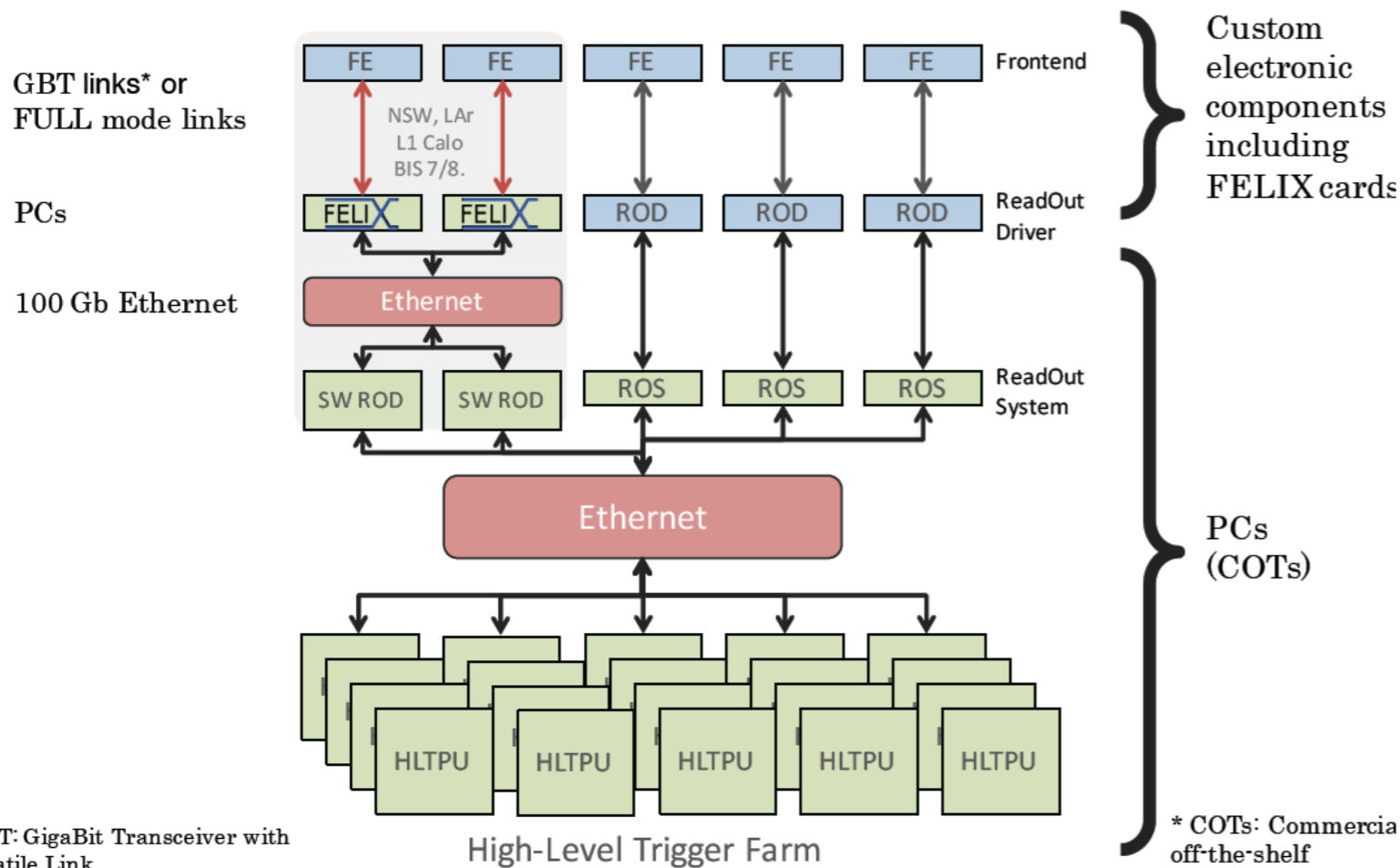
# What's next: LHC Schedule





# Upgrade for Run3

- **FELIX** is a modern readout system
  - ▶ it relies on commercial technology and it is much easier to support and upgrade than the custom system
- **FELIX** will readout the new muon and calorimeter detector FEs
  - ▶ Readout and configuration of on-detector electronics
    - ▶ communication relies on the GigaBit Transceiver (GBT) with Versatile Link, developed at CERN
  - ▶ Distribution of trigger information and LHC clock to the on-detector electronics



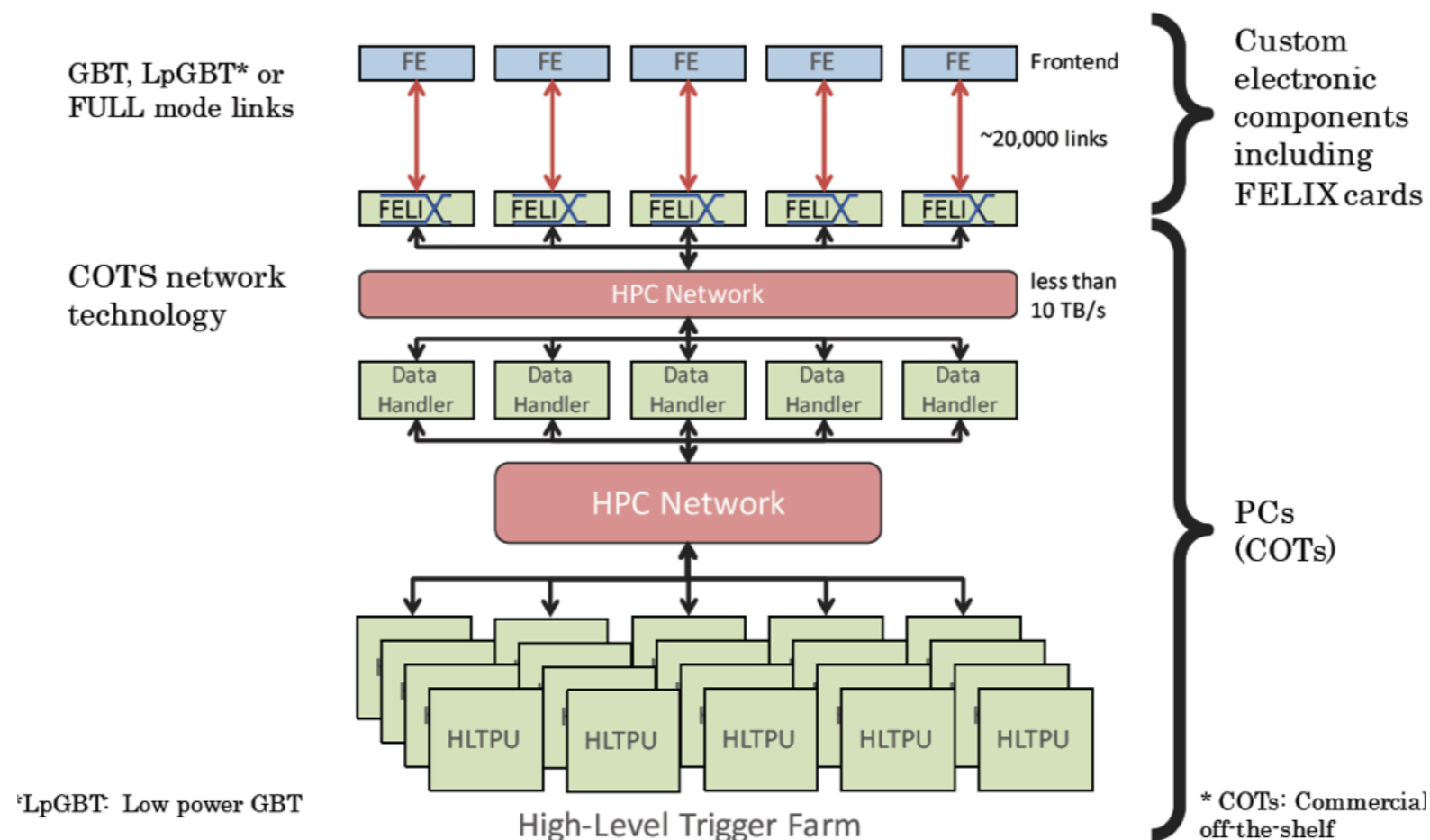
Design Parameter ( $L \sim 3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , PU=80)

Parameter	Run 3 Value
Level-1 trigger latency	$\approx 1.9 \mu\text{s}$
Maximum Level-1 readout latency	$2.5 \mu\text{s}$
Level-1 trigger rate	100 kHz
Detector system readout deadtime ( $1.9 \mu\text{s}$ latency)	< 2%
Number of detector readout data sources	$\sim 100$
Bandwidth of detector ROLs	1.28-9.6 Gb/s
Average readout event size at peak luminosity	$\approx 2.9 \text{ MB}$
ROLs bandwidth at peak luminosity	$\sim 290 \text{ GB/s}$
Maximum readout fraction of the ROS	50%
Average data collection bandwidth at peak luminosity	$\approx 25 \text{ GB/s}$
Average output rate at peak luminosity	1.5 kHz
Number of data loggers	$\sim 10$
Average output event size (raw) at peak luminosity	$\approx 3.5 \text{ MB}$
Average output bandwidth at peak luminosity	$\approx 3.2 \text{ GB/s}$

GBT: GigaBit Transceiver with Versatile Link

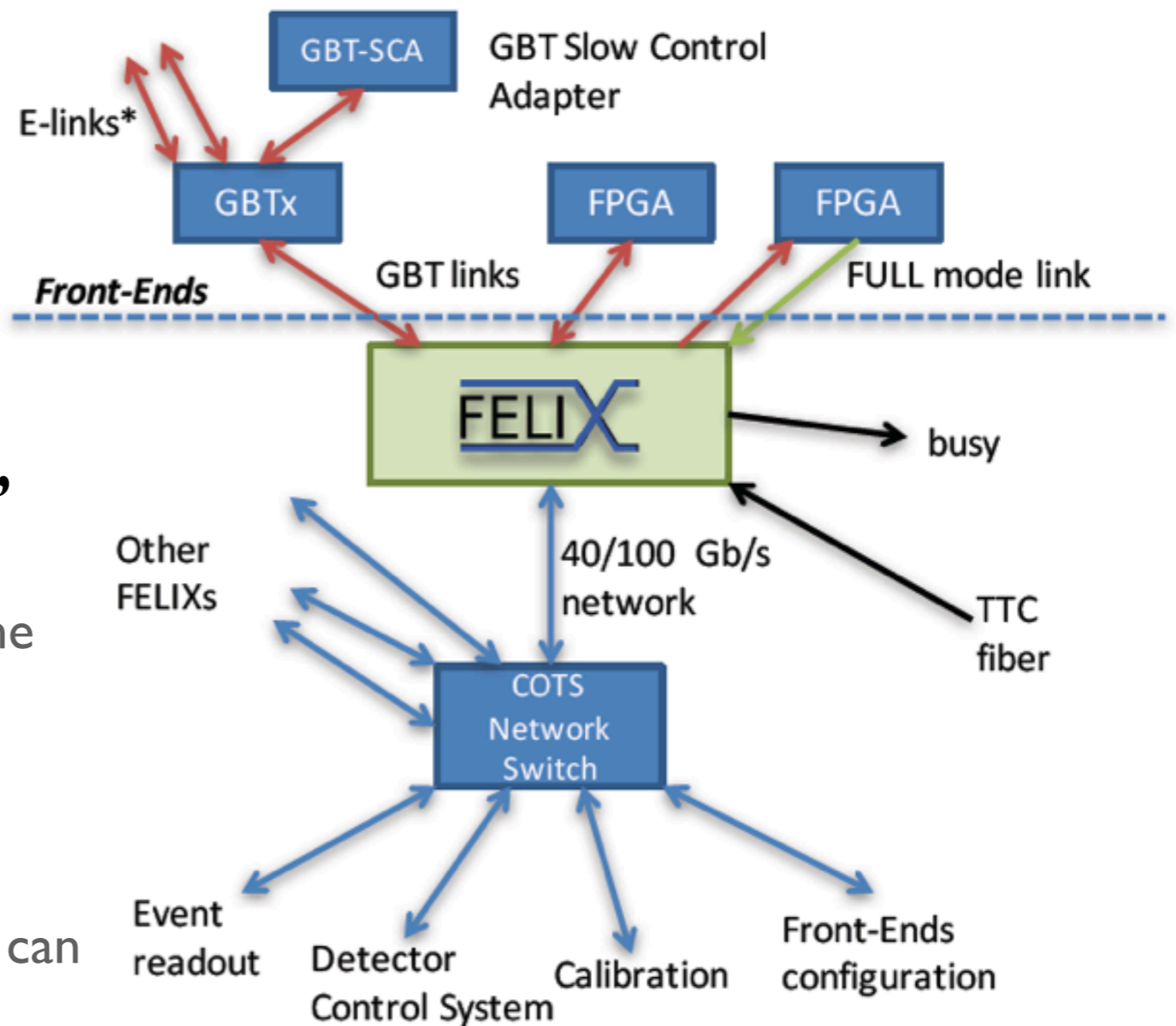
# Upgrade for HL-LHC

- Phase I DAQ cannot cope with HL-LHC harsh conditions
  - ▶ LI Trigger rate of 1 MHz to preserve physics performances
  - ▶ 5 MB event size mainly due to the increase in pile-up
  - ▶ >20x increase of incoming bandwidth
- FELIX will readout the entire ATLAS detector
  - ▶ taking advantage of the in the technology progress available “off-the-shelf”
  - ▶ readout system designed to be generic wrt to different FEs, protocols, bandwidths
  - ▶ will rely on the more performant low power GBT radiation tolerant protocol (LpGBT)



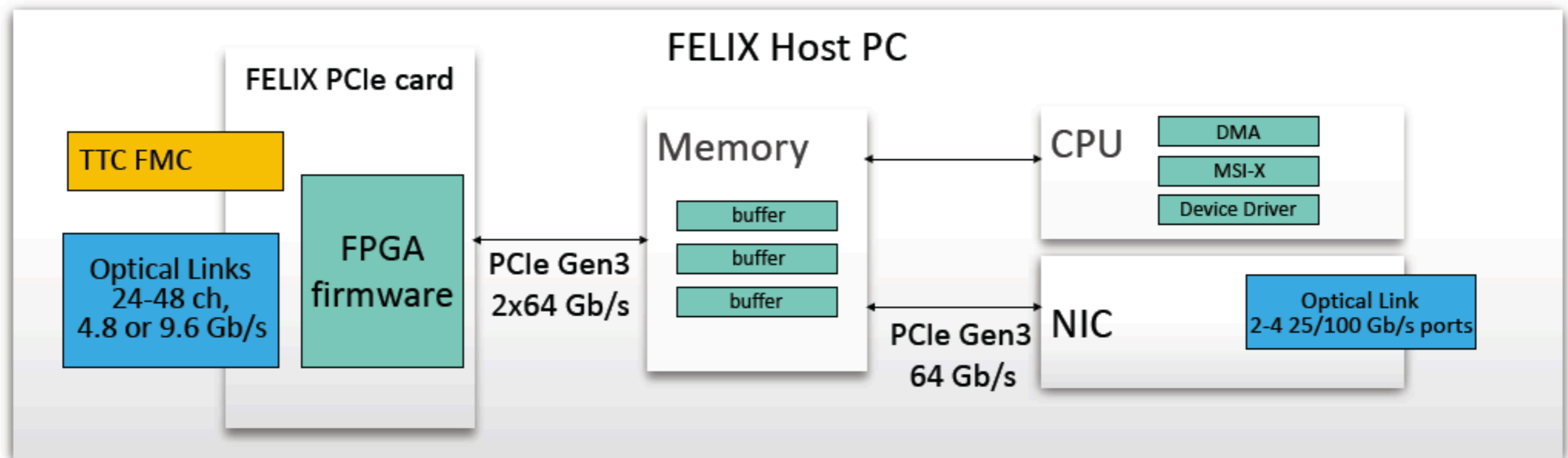
# FELIX Functionality

- FELIX is a router between FE serial links and commercial network
  - ▶ separates data transport from data processing
  - ▶ routes detector control, configuration, calibration, monitoring and detector event data
- FELIX is fed by TTC (Timing, Trigger and Control)
  - ▶ TTC is a CERN protocol to distribute the 40.08 MHz LHC clock and LI trigger info
- FELIX features configurable E-links in GBT mode
  - ▶ E-link is a variable-width logical link that can be used to separate different streams on a single physical link
- FELIX is detector independent



# FELIX Block Diagram

PC Hosting a Peripheral Component Interconnect express (PCIe) card and Network Interface Card (NIC)



# FELIX Hardware Components

- **FLX-712 (Final prototype)**

- ▶ Xilinx Kintek Ultrascale XCKU115
- ▶ 48 optical links (MiniPODs)
- ▶ TTC input ADN2814
- ▶ PCIe Gen3 x16 (2x8 with switch)
- ▶ Si5345 jitter cleaner

- **Timing mezzanine Card**

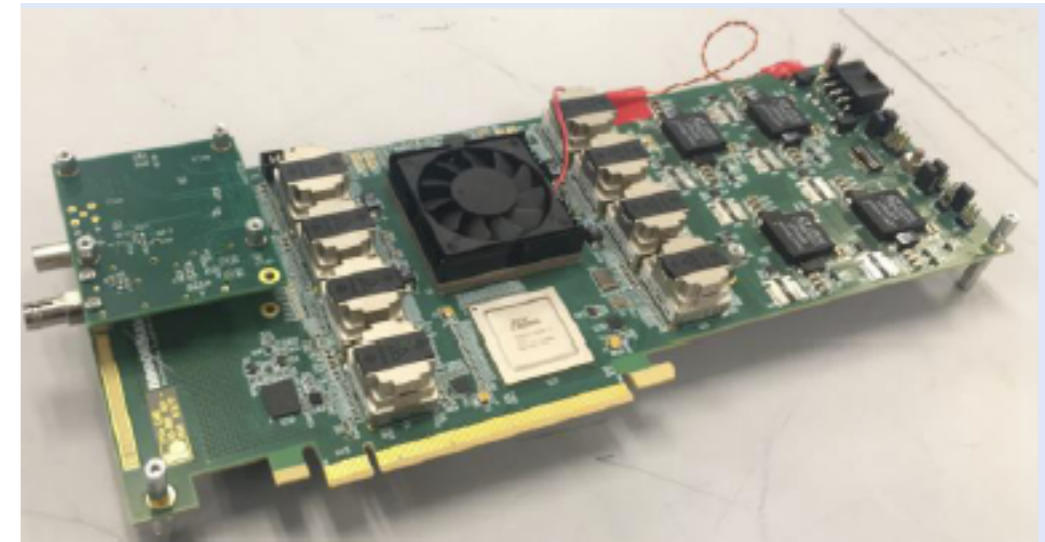
- ▶ Supports TTC, TTC-PON, White Rabbit
- ▶ TTC configuration is for Run3 Timing mezzanine Card
- ▶ Supports TTC, TTC-PON, White Rabbit

- **Current motherboard for development**

- ▶ SuperMicro X10SRA-F
- ▶ Broadwell CPU, e.g: E5-1650V4, 3.6 GHz
- ▶ PCIe Gen3 slots

- **Current NIC for testing**

- ▶ Mellanox ConnectX-3
- ▶ 2x FDR/QDR Infiniband
- ▶ 2x 10/40 GbE



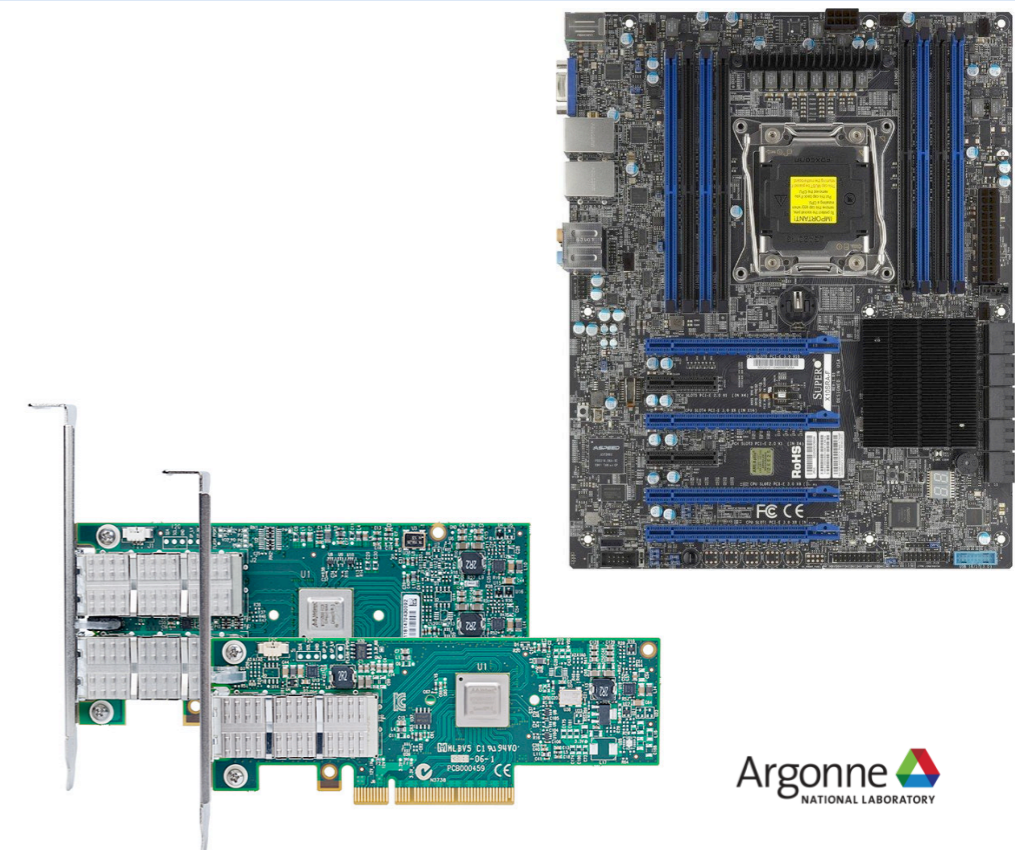
TTC



TTC-PON



White Rabbit



# FELIX Hardware Components

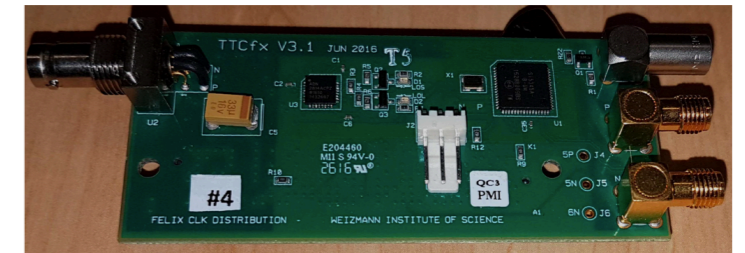
- Mini FELIX (for development)

- ▶ Xilinx VC-709 Evaluation Board
- ▶ Virtex7 X690T FPGA
- ▶ 4 optical links (SFP+)
- ▶ PCIe Gen3 x8
- ▶ Si5345 jitter cleaner



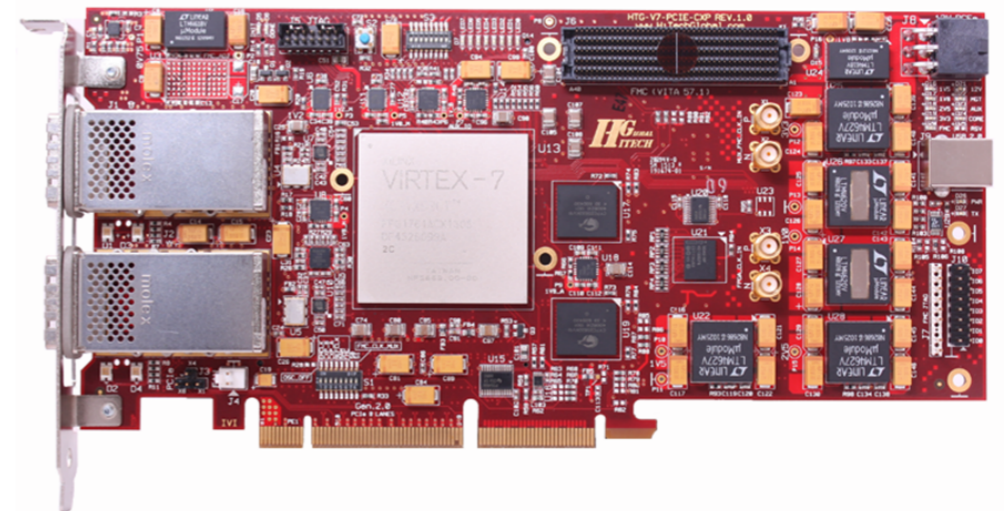
- TTCfx (v3) Mezzanine Card

- ▶ TTC input and busy output
- ▶ ADN2814 for TTC clock recovery
- ▶ Si5345 jitter cleaner



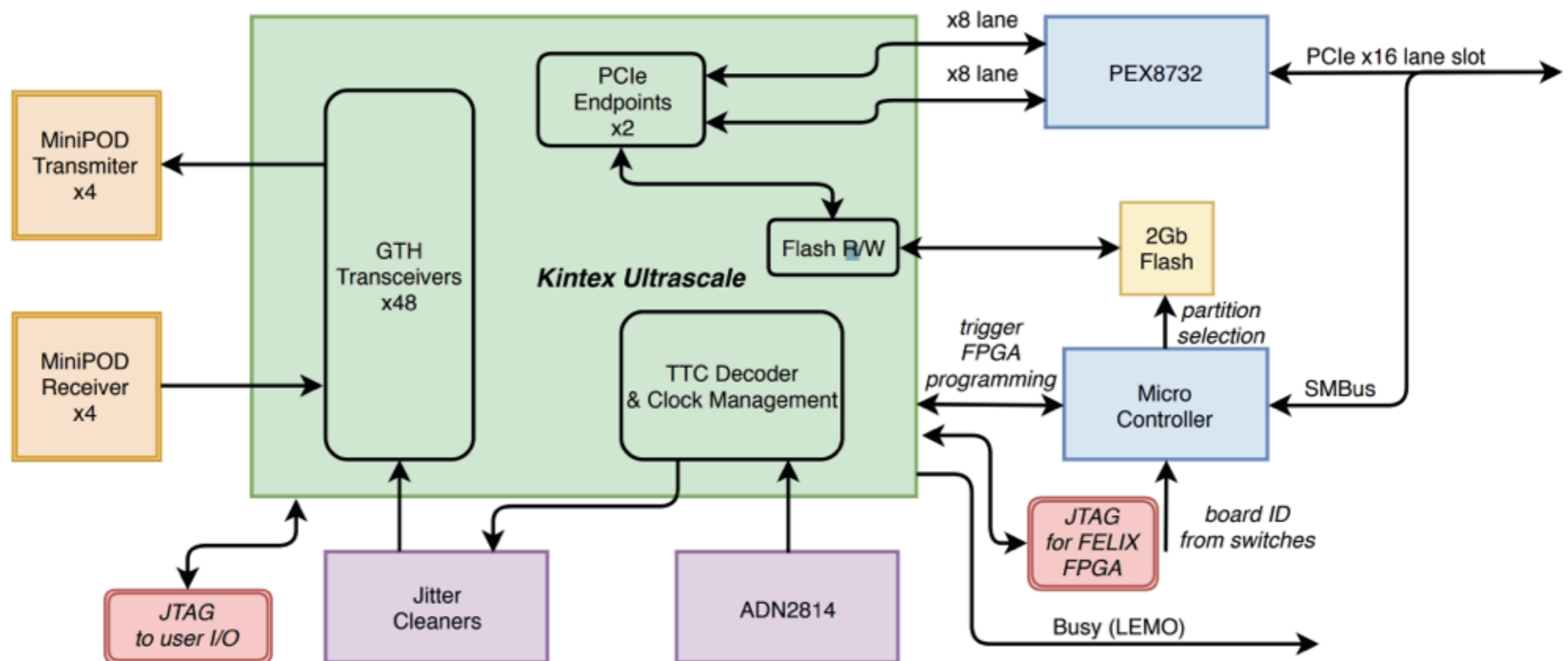
- FLX-710 for emulation of detector FEs

- ▶ High Tech Global HTG-710
- ▶ Virtex7 X680T FPGA
- ▶ 2 CXP connectors: 24 channels
- ▶ PCIe Gen3 x8



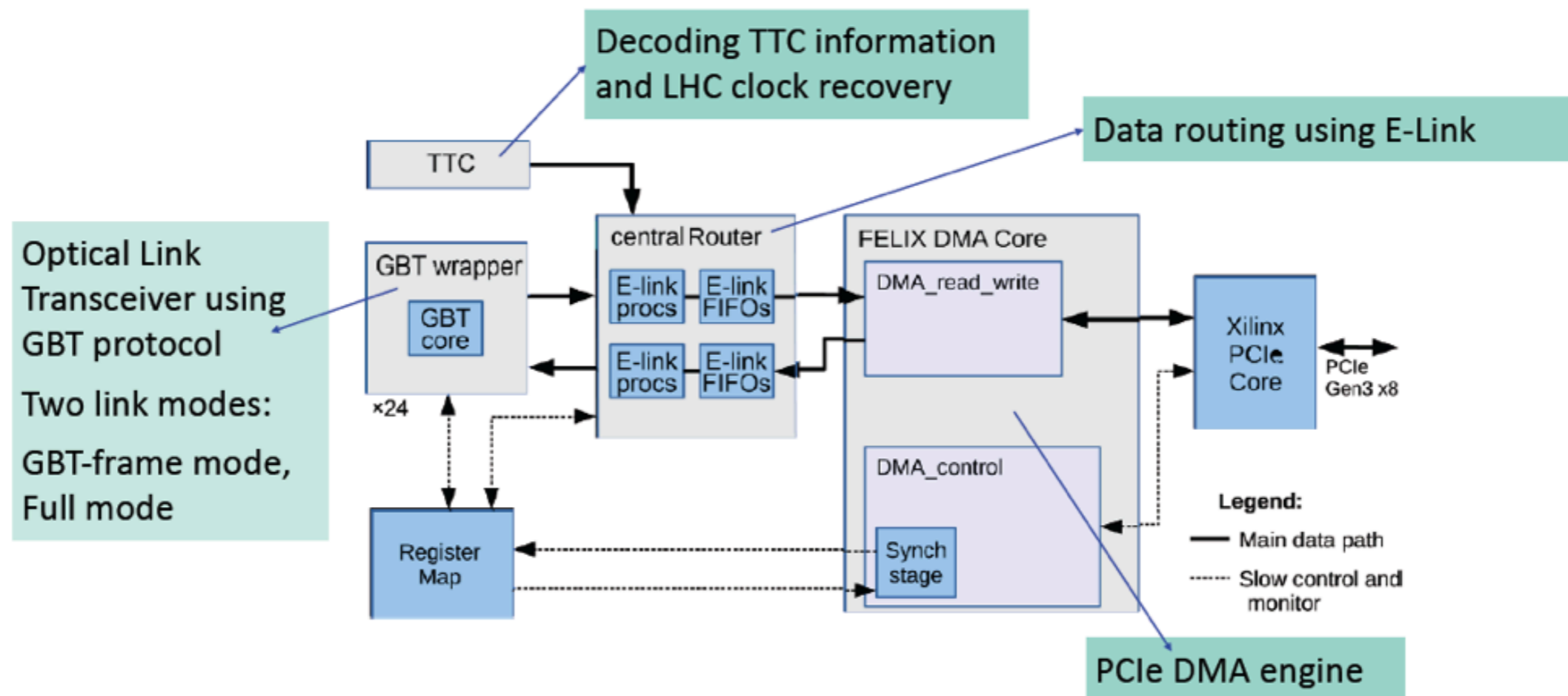
# FLX-712 PCIe Features

- FPGA: Kintex UltraScale XCKU115
- 8 MiniPODs to support 48 bidirectional optical links
- 16-lane PCIe Gen3 (two 8-lane Endpoints with a switch)
- Flash and Micro-controller to support firmware update
- Two 0-delay jitter cleaners: Si5345 or LMK03200 (backup)
- Timing mezzanine to interface to the TTC system



# FELIX firmware design

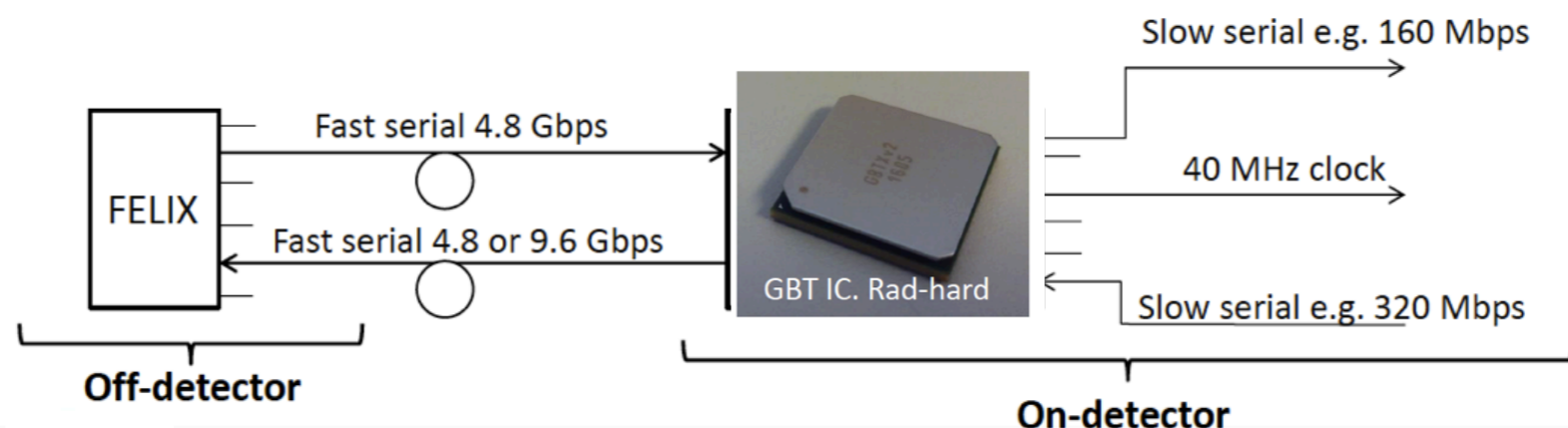
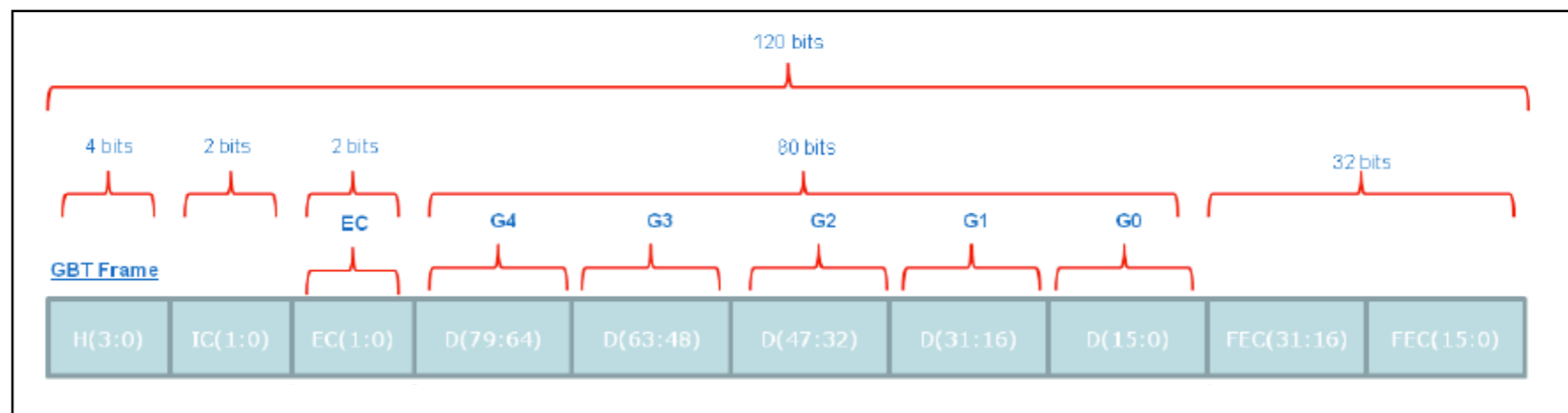
- Up to 24 GBT or full-mode links
- Sophisticated central router block handles e-links
  - e-link: data multiplexing/demultiplexing protocol designed for ATLAS
- Fixed latency transmission to FE
- Maximum PCIe throughput to the host PC (2x64 Gbps)
  - PCIe block relies on Direct Memory Access (DMA)



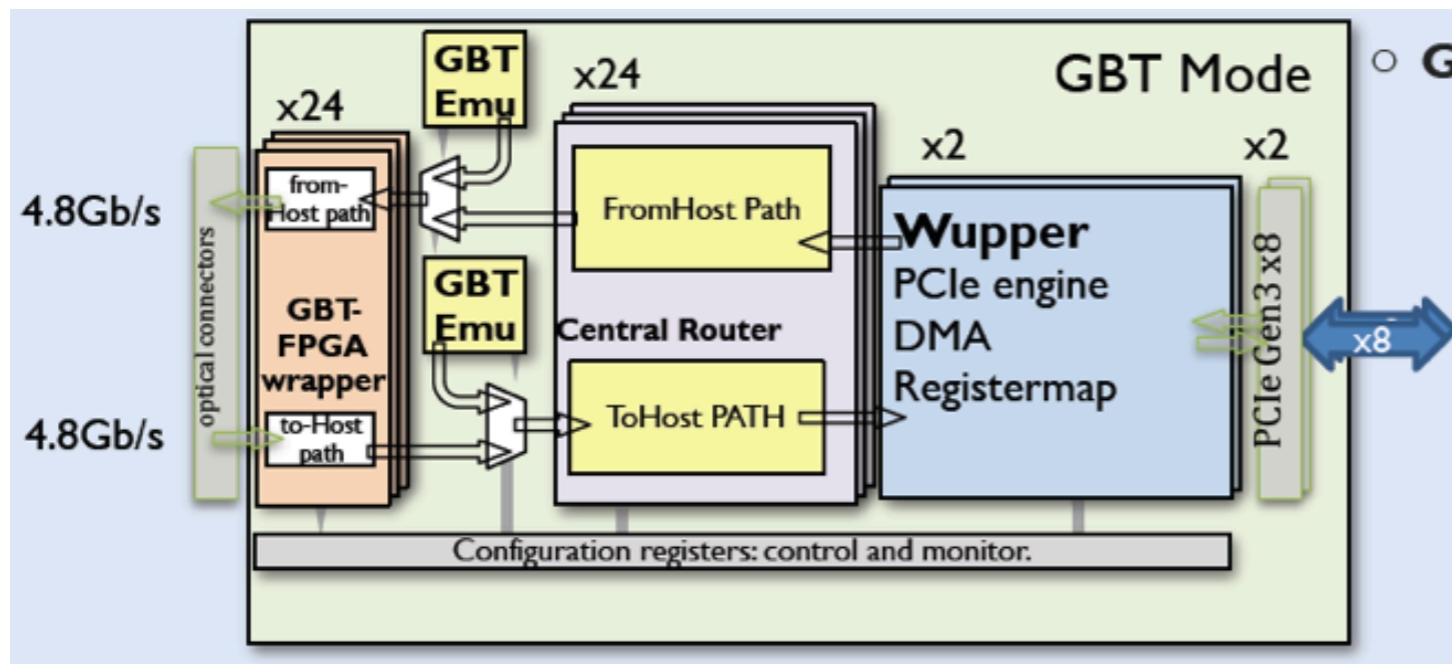


# Clarification: GBT Protocol

- GBT protocol multiplexes slower data streams (e-links) into a single (GBT) frame
  - ▶ e-links number and data rates are programmable
  - ▶ data scrambled before serialization
- GBT frame is fully synchronous with the LHC clock
  - ▶ e-links number and data rates are programmable by configuring the I/O ports
- GBTx can communicate to  $\leq 40$  FEs within 25 ns
- High fidelity clock (40, 80, 160, 320 MHz) provided

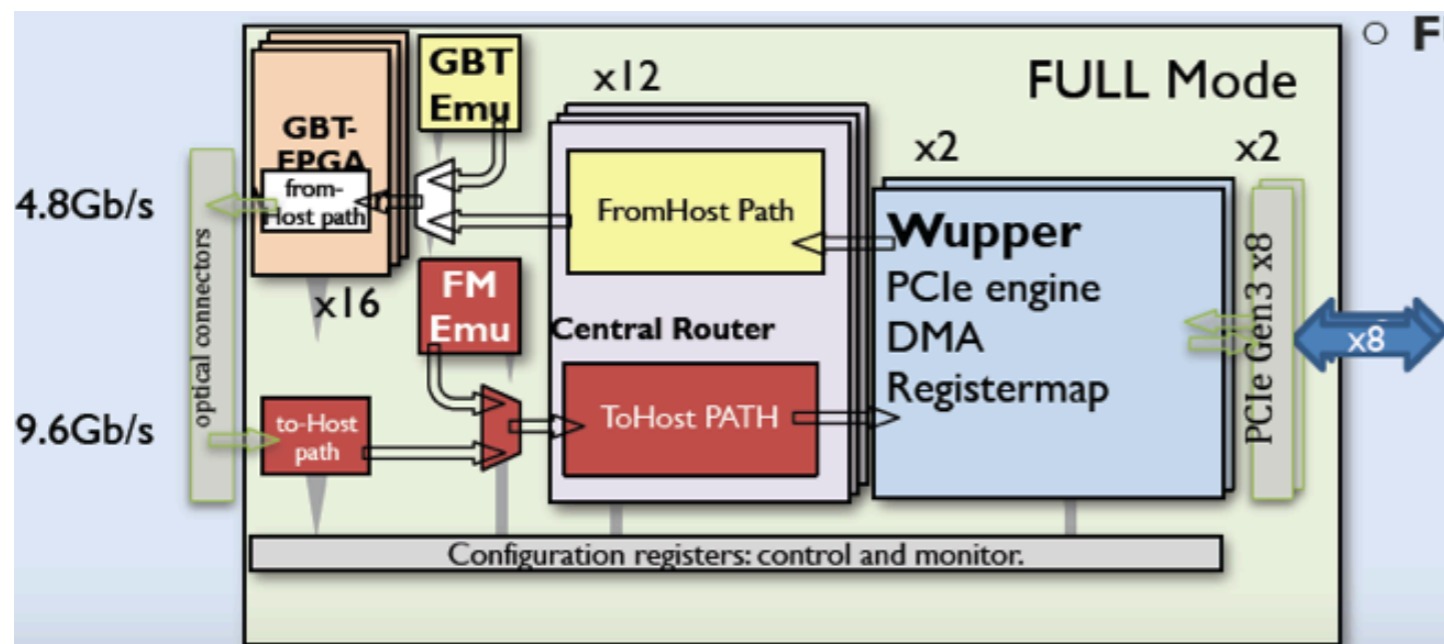


# FELIX modes of operation



## • GBT mode

- ▶ Up/Down Link Line rate: 4.8 Gb/s
- ▶ Up to 24 bidirectional optical links
- ▶ 3.2 Gb/s payload with FEC or 4.48 Gb/s payload (wide)
- ▶ Routes TTC information
- ▶ Optical link divided in E-links
- ▶ Communicate with GBTx and GBT-SCA

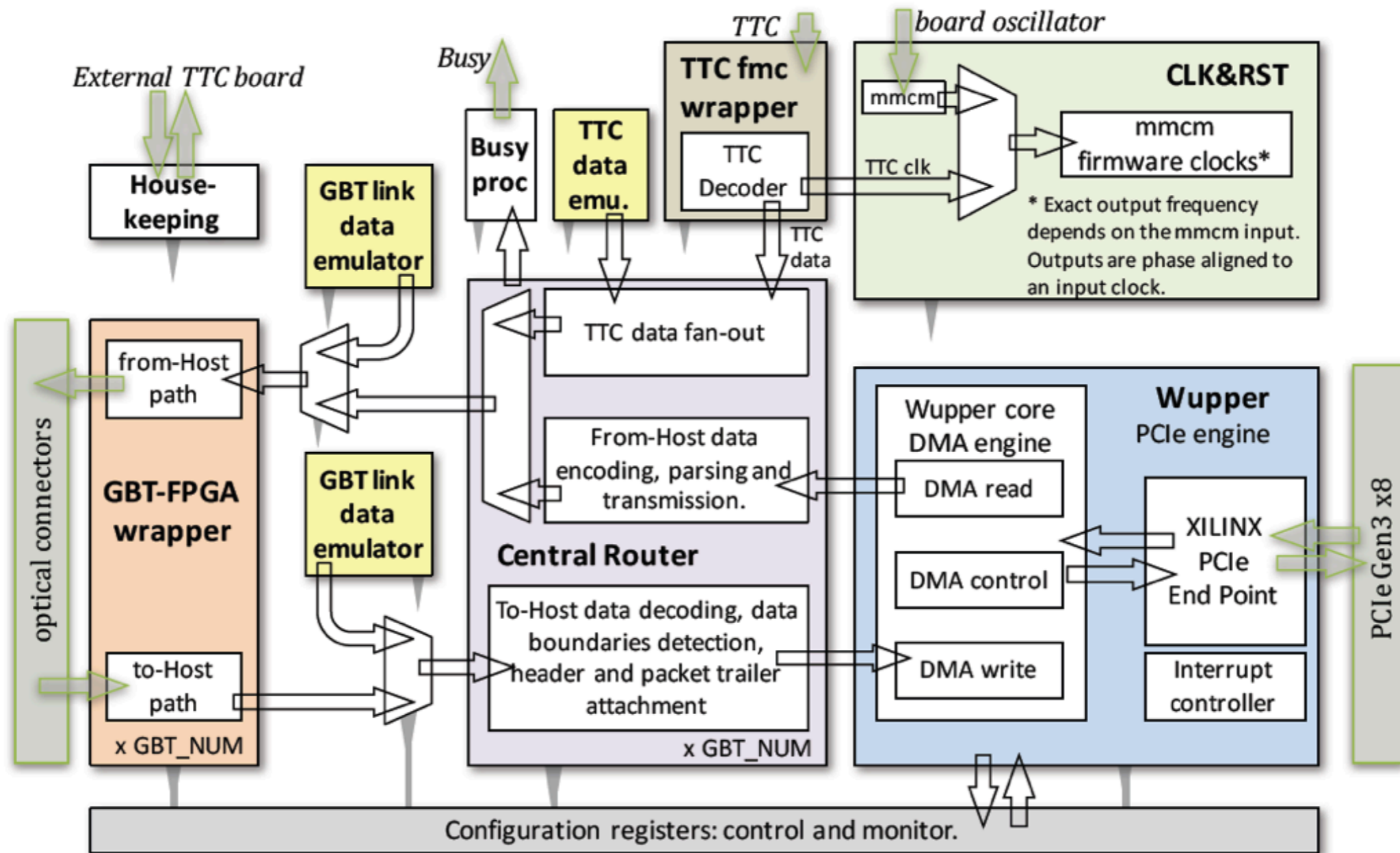


## • Full mode

- ▶ Up Link line rate: 9.6 Gb/s
- ▶ Down Link line rate: 4.8 Gb/s
  - ▶ down link operates as in GBT mode
- ▶ Up to 24 bidirectional optical links
- ▶ 7.68 Gb/s payload: 8b/10b encoding
- ▶ Routes TTC information
- ▶ BUSY-ON and OFF
- ▶ GBT links to FE

# FELIX Firmware Block Diagram

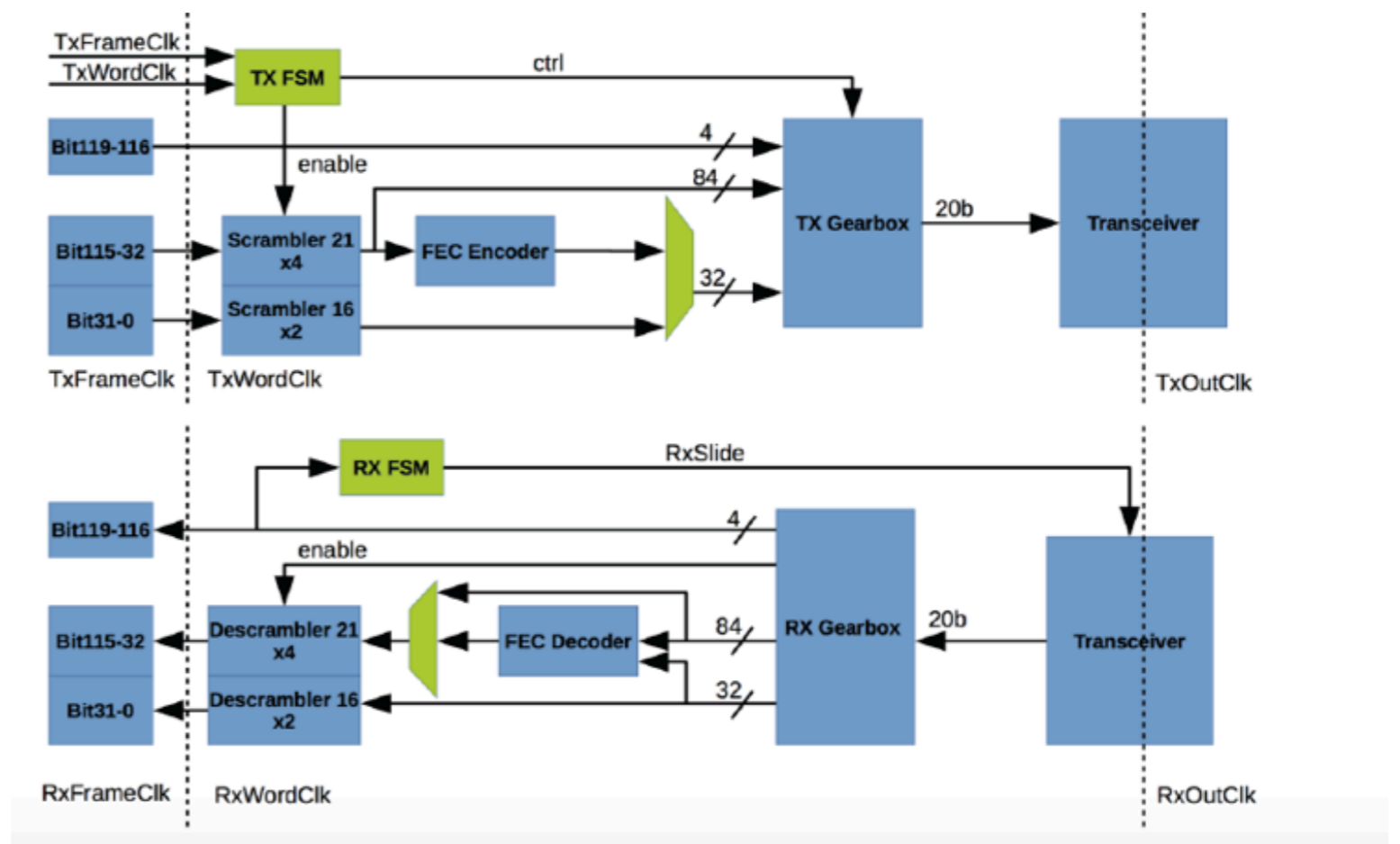
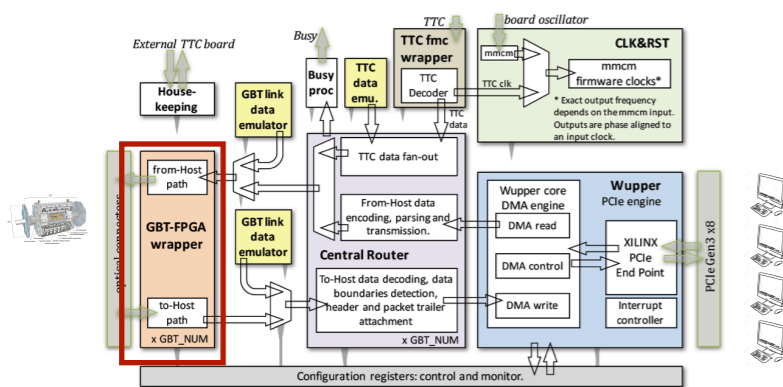
(N.B: Showing GBT mode only, Full Mode shown backup)



# GBT Wrapper

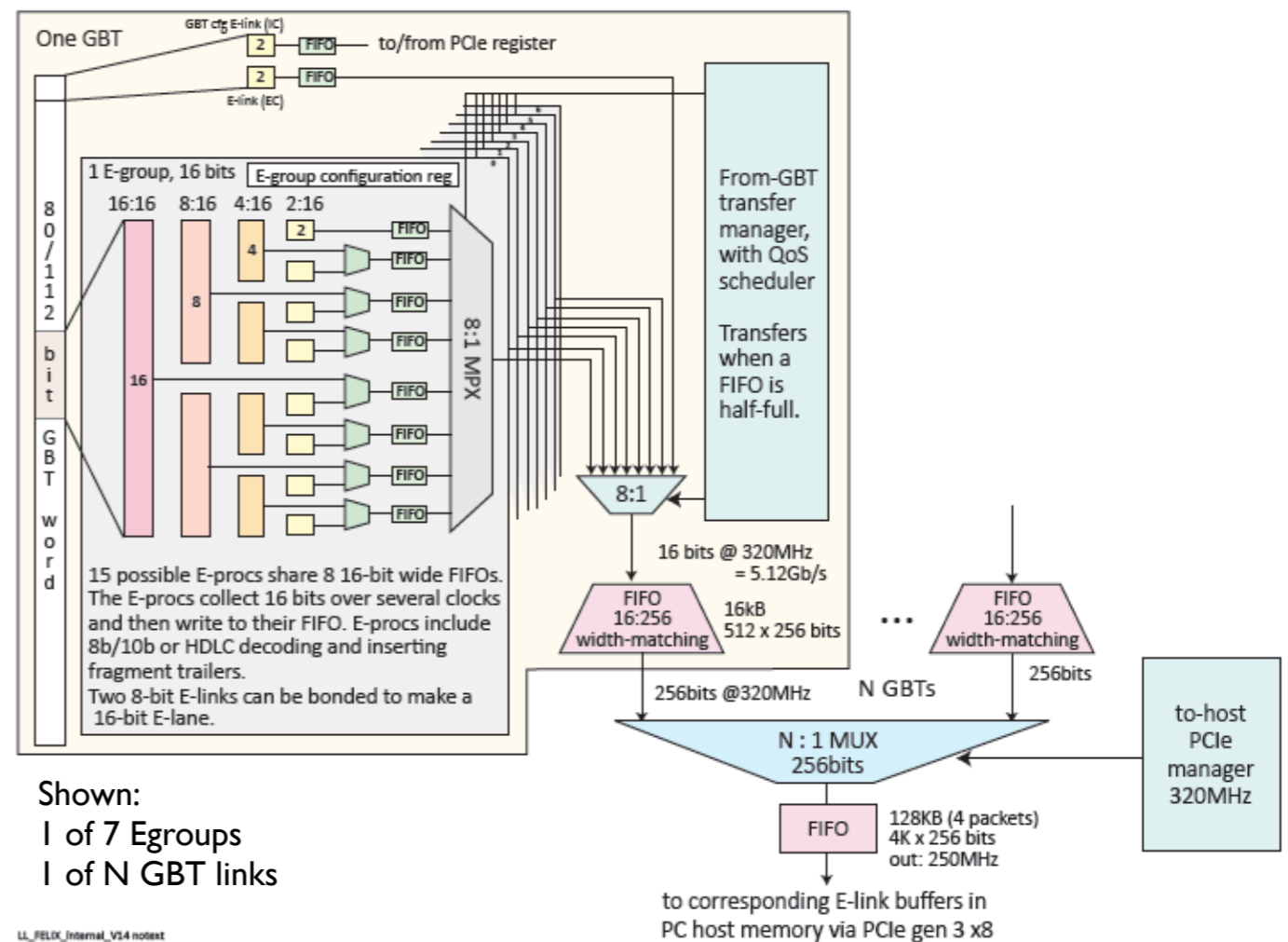
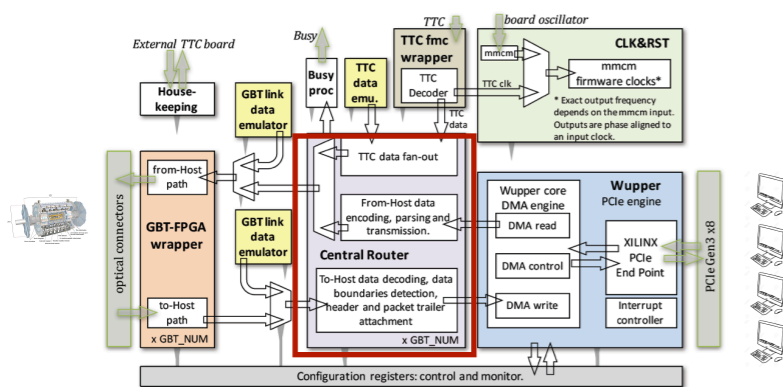
- FELIX GBT Wrapper builds on top of CERN GBT-FPGA (015 JINST 10 C03021). Modifications are:

- ▶ Separated GBT firmware from transceiver block
- ▶ GBT mode configurable at run time between default mode (w/ FEC, 3.2 Gbps payload) or wide-bus mode (w/o FEC, 4.48 Gbps payload)
- ▶ Lower ~fixed latency (TX: 27.8-32 ns; RX-default mode 56. ns, RX-wide mode 43.9 ns)
  - ▶ GBT encoding/decoding are in 240 MHz domain
- ▶ JINST 12 P07011



# Central router

- Handles data stream from/to GBT channels to/from host PC
- Routes TTC information
- Dedicated manager for E-links in GBT channels
- Main manager towards PCIe engine

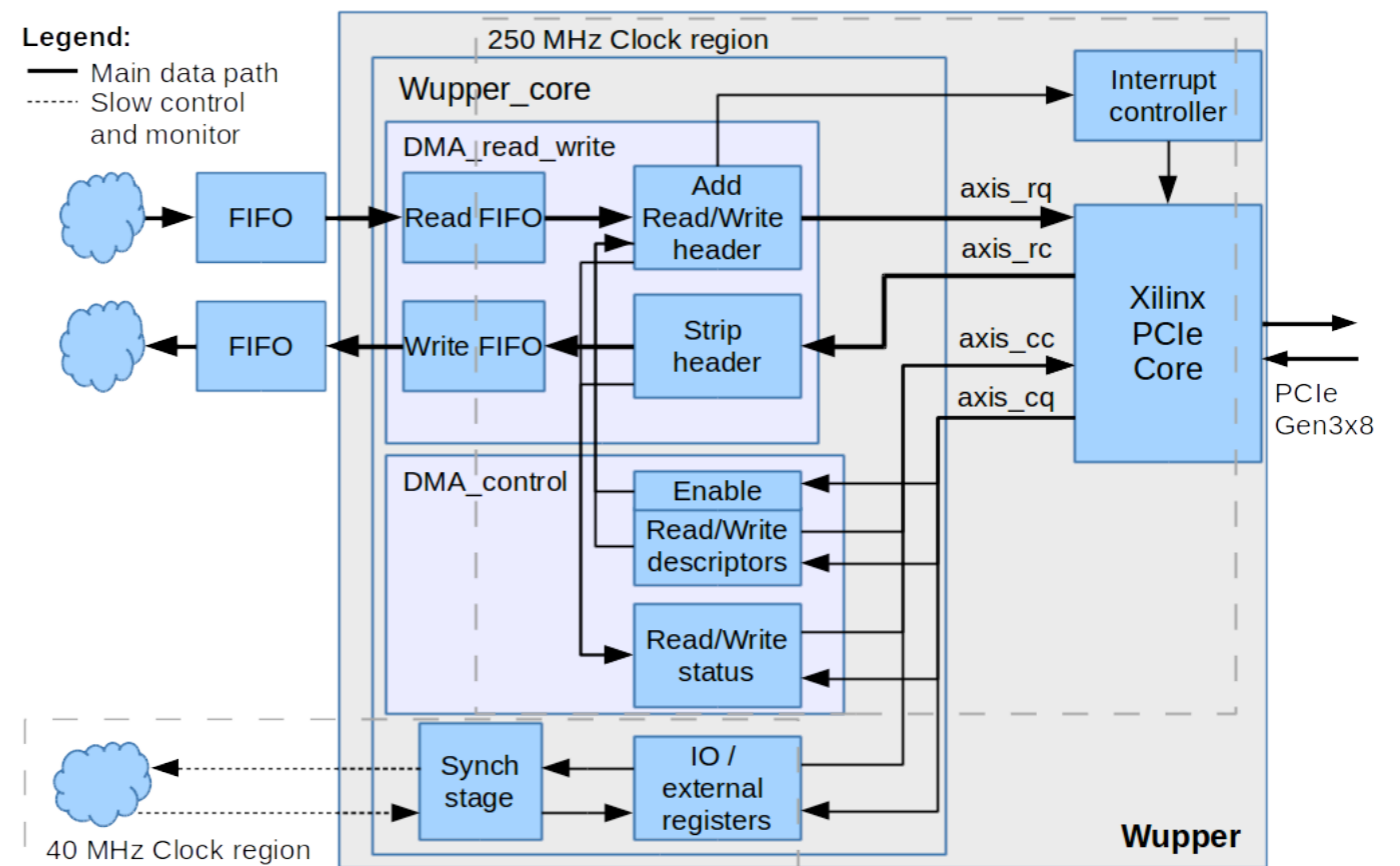
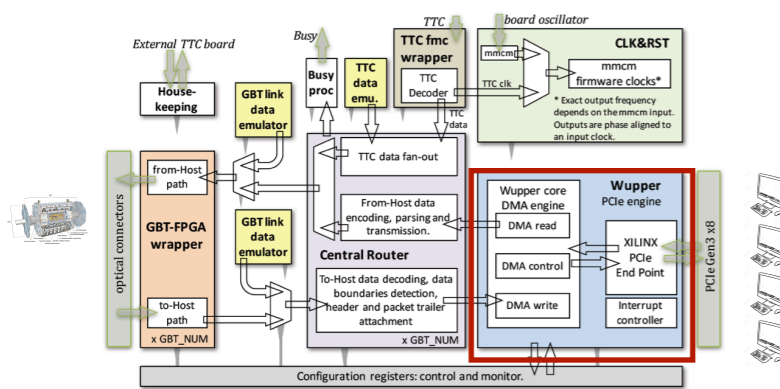


Shown:  
 1 of 7 Egroups  
 1 of N GBT links

LL\_FELIX\_internal\_V14 notext

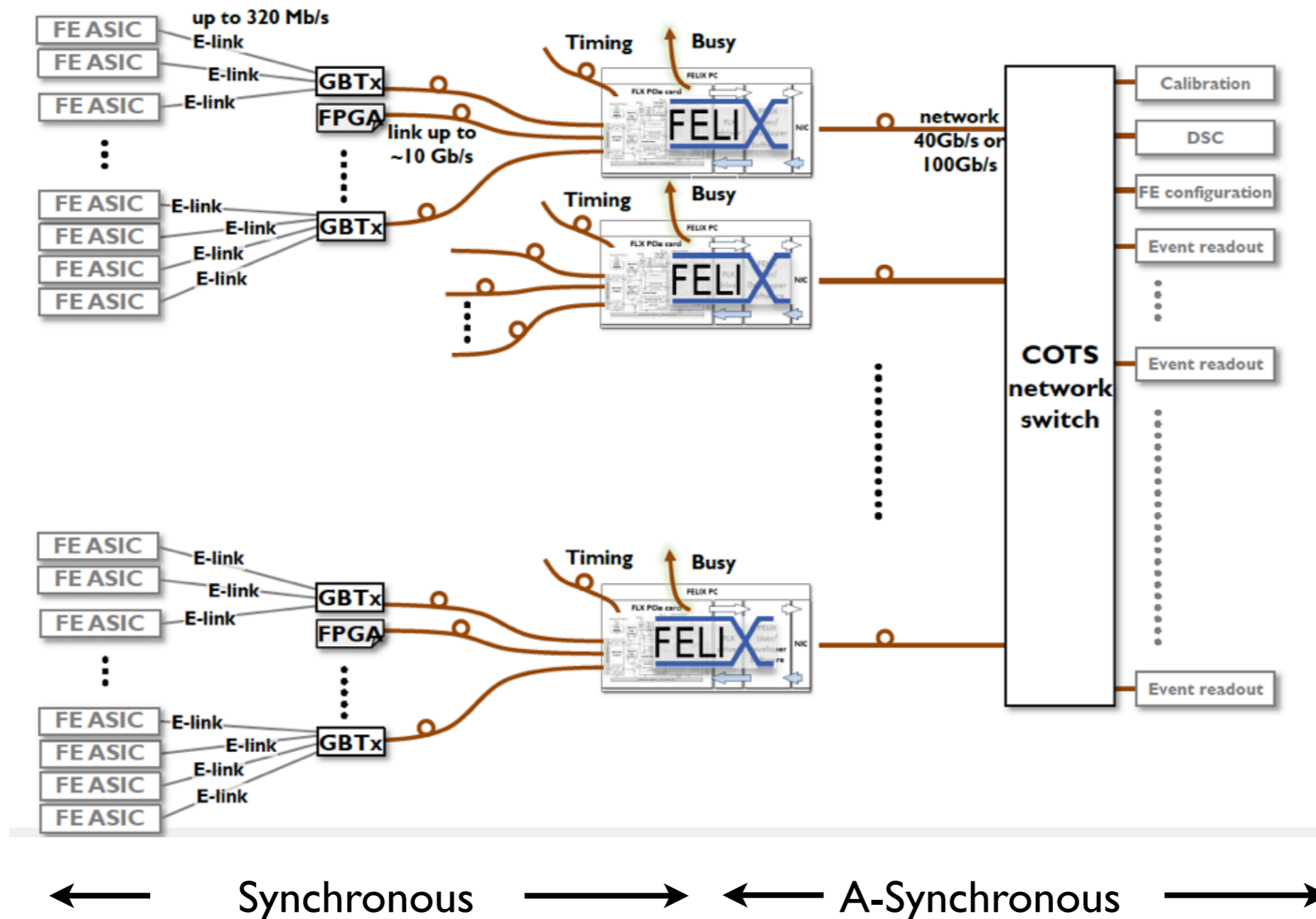
# PCIe Engine

- PCIe Engine with DMA interface to the Xilinx Virtex-7/Kintek Ultrascale PCIe Gen3 Integrated Block for PCI Express
  - Xilinx AXI Stream Interface ([UG761](#))
- MSI-X compatible interrupt controller
- Applications access the engine via simple FIFOs
- Register map synchronized to a lower clock speed
- Published as Open Source (LGPL) on [OpenCores](#)



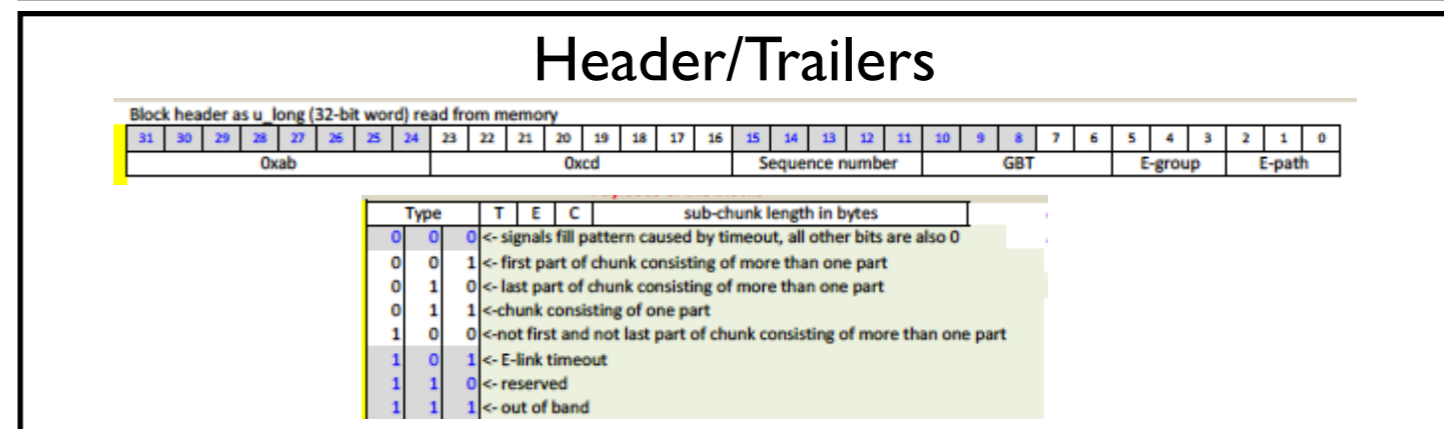
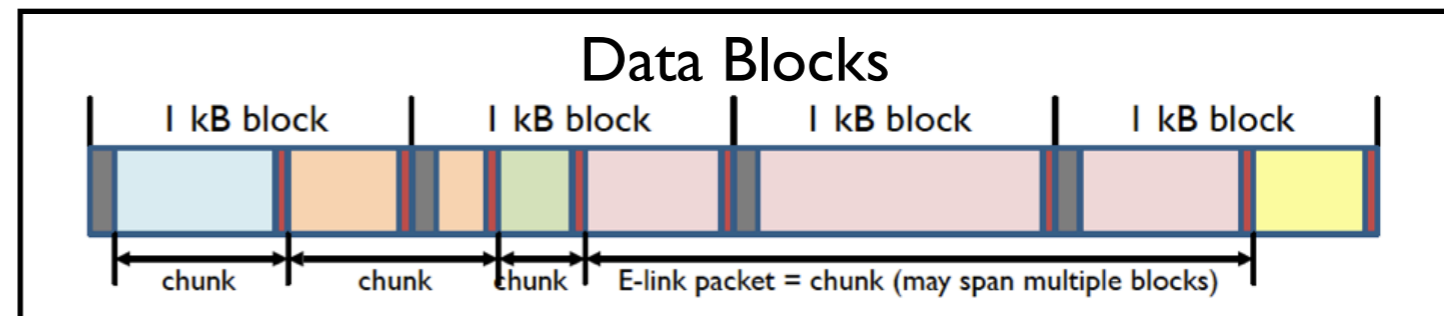
# FELIX Data Overview

- FE to FELIX via custom data links: fully synchronous systems
- Commodity networks: multiple a-synchronous streams



# Data Format

- Data per (e-)link buffered in the FPGA and transferred under the DMA control
- Blocks transferred into a contiguous areas (circular buffer) in the PC memory
  - ▶ DMA runs continuously: 12 GB/s throughput for the 16-lane interface of the FLX-712
- Chunks is the data of arbitrary size from the FE E-link
- Fixed block size of 1 kB chosen
- Blocks are multiplexed into a single stream
- Blocks feature:
  - Headers: 32b
    - E-link ID
    - Block sequence
    - Start of Block Symbol
  - Fragment trailers: 16b
    - Type (i.e: first/last/middle) : 3b
    - Flags (e.g: errors, truncation): 3b
    - Fragment length : 10b





# FELIX Software Package

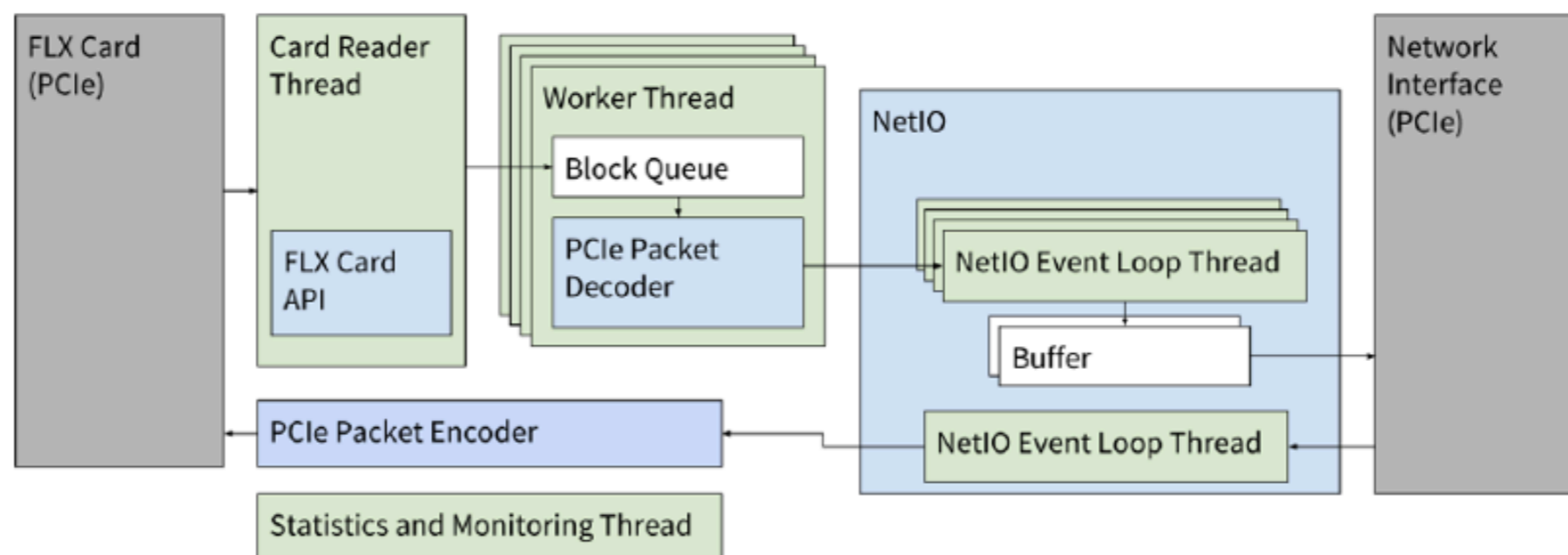
- “FLX” driver: provides virtual access to the FLX card register
  - ▶ conventional for PCIe cards
- flxcard: for monitoring, testing, configuring
- f-tools: testing/debugging FELIX functionalities
- E-link configuration tool
- FELIX Core: process for FLX card to/from PC communication



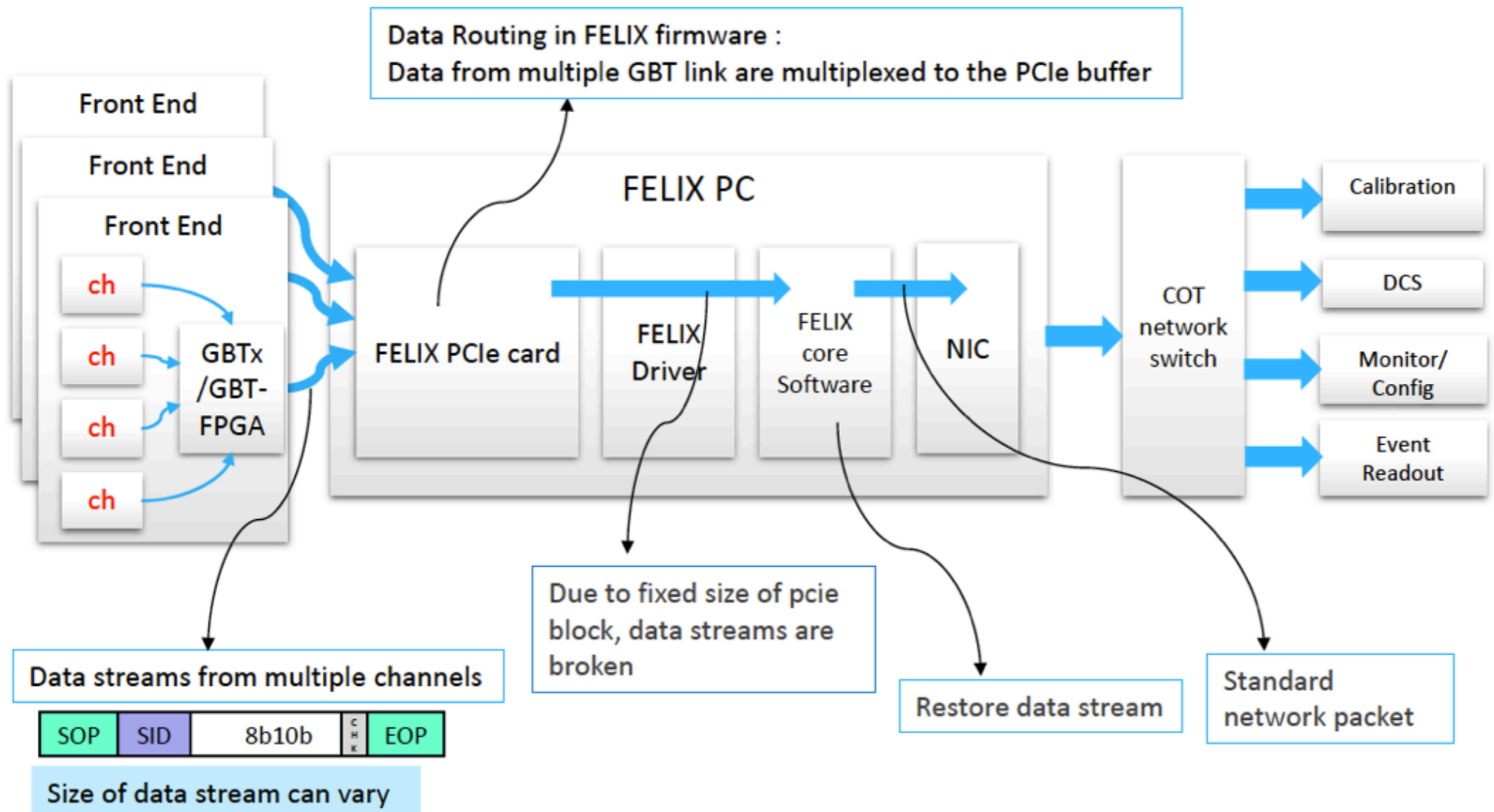
# FELIX Core

- Felix Core Application: data to/from the PCIe DMA buffer from/to the NIC

- ▶ ToHost: data from FLX card is routed to NetIO clients via DMA into the PC memory
- ▶ FromHost: data from NETIO clients (SW RODs) is copied to FLX card via DMA
- ▶ NetIO is used for the data exchange between FLX-card and clients
  - ▶ a publish-subscribe model allows clients to subscribe to a given subset of ports
  - ▶ POSIX and Infiniband supported
  - ▶ more details in CHEP 2016 talk
- ▶ API for an easier interface with status/control via register, data transfer, interrupt handling
- ▶ Constant monitoring: gather statistics, performances, status info (e.g: link status, hardware failure)
  - ▶ monitoring info made available on an internal web server

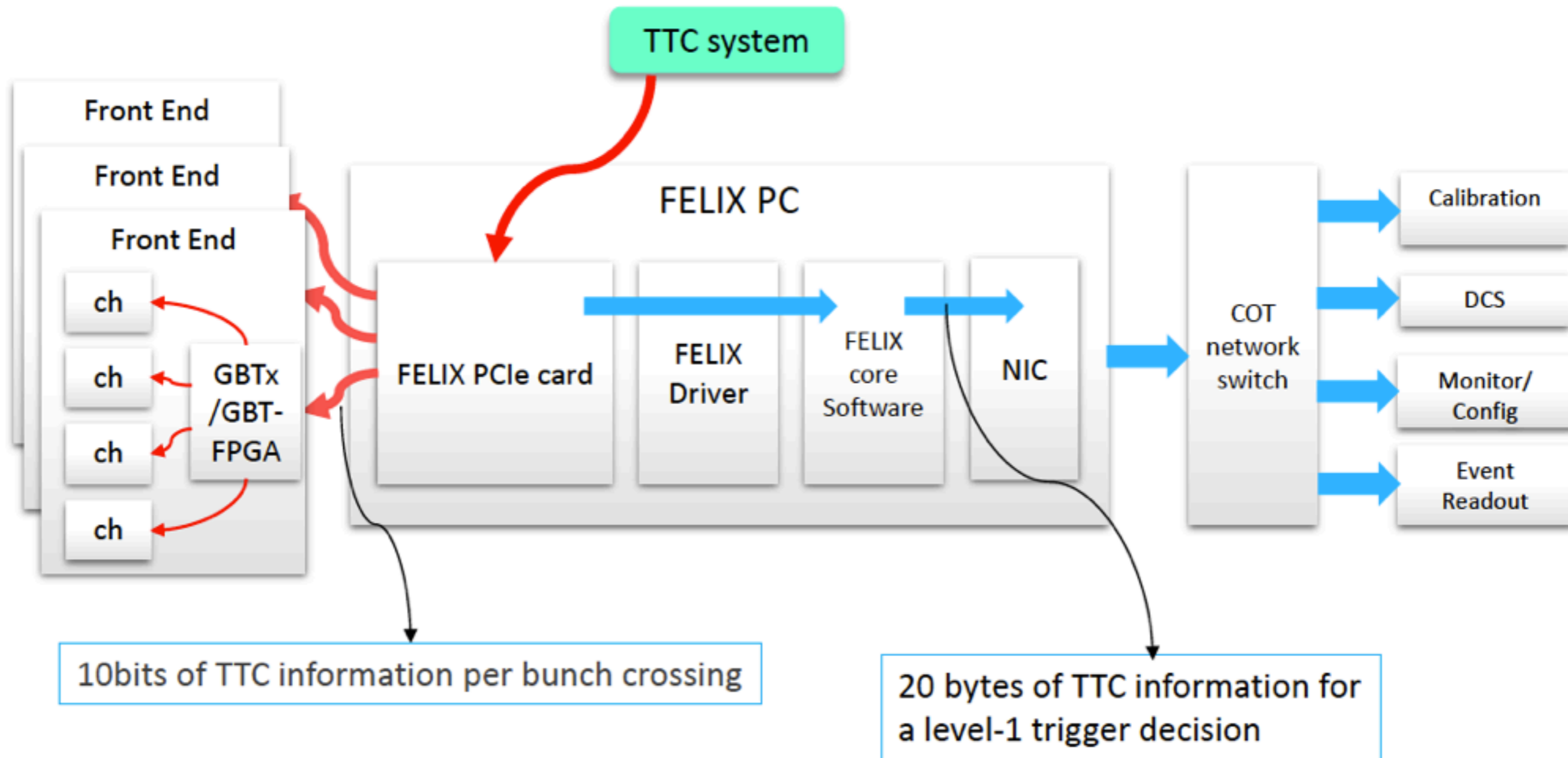


# FELIX data flow: FE->Host (toHost)



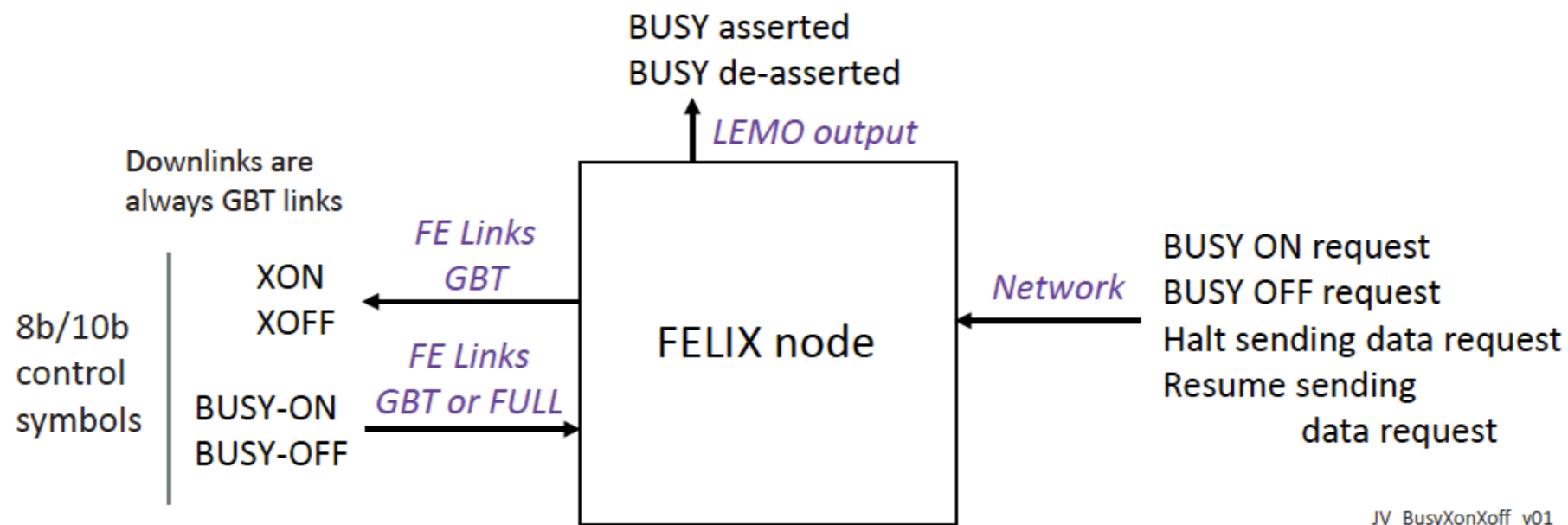


# FELIX TTC flow: to FE and to Host



# FELIX Busy and XON/XOFF

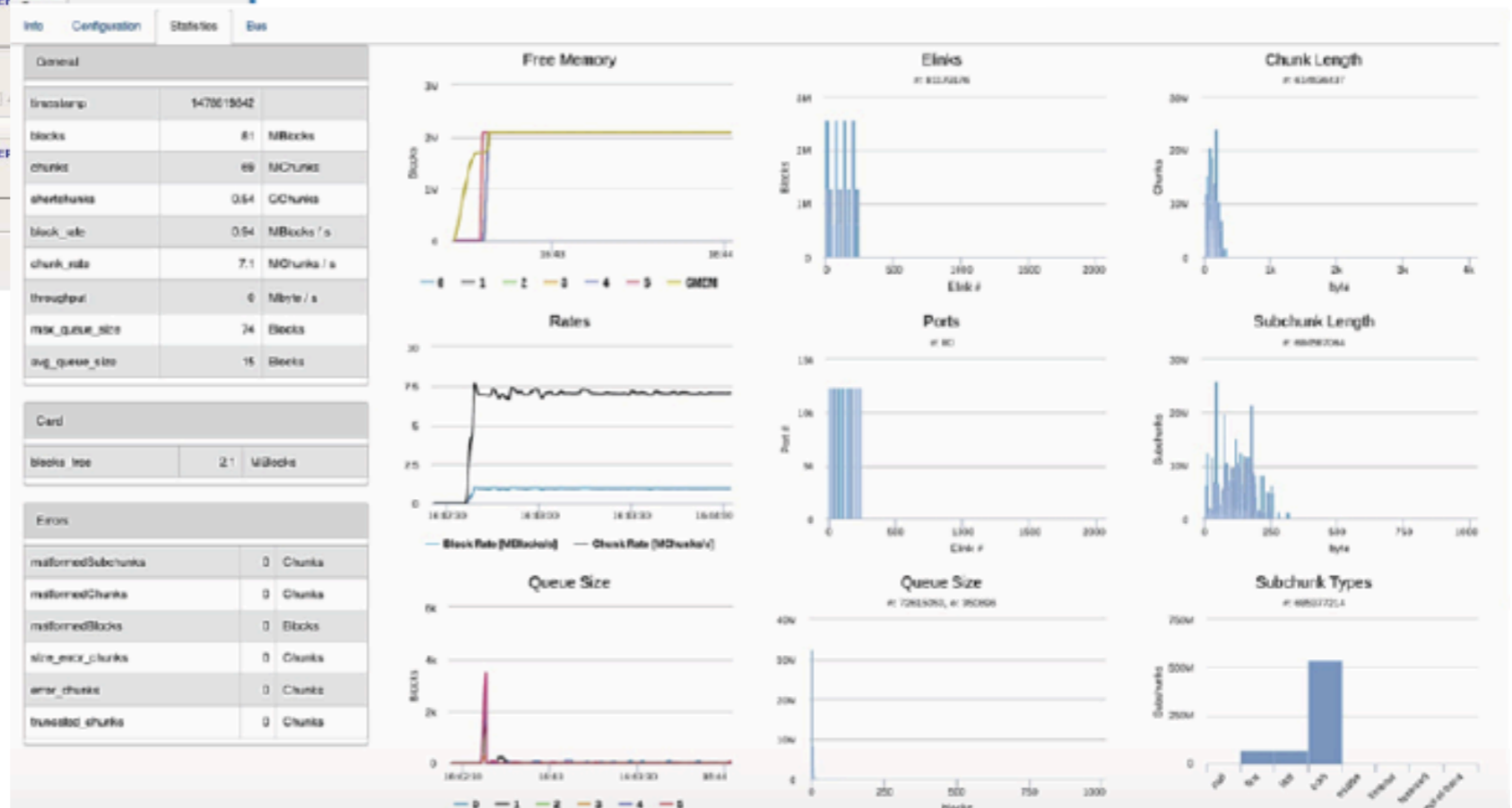
- Busy signals sent through LEMO to the central trigger processor to stop generating LI accepts
  - ▶ data flow stopped once FE buffers drain
  - ▶ used only for emergency situations when buffers are about to overflow
  - ▶ sources of busy: busy-on/off cmdnd from FEs, software/firmware signaling internal overflows
- XON/XOFF signals used to throttle transmission of data from the FE to FELIX to prevent bus overflow
  - ▶ assuming that the data source has sufficient buffer, dataflow can be paused w/o data loss
  - ▶ fullmode uplinks support flow controls, GBT uplinks do not
- Busy has been tested and validated



JV\_BusyXonXoff\_v01

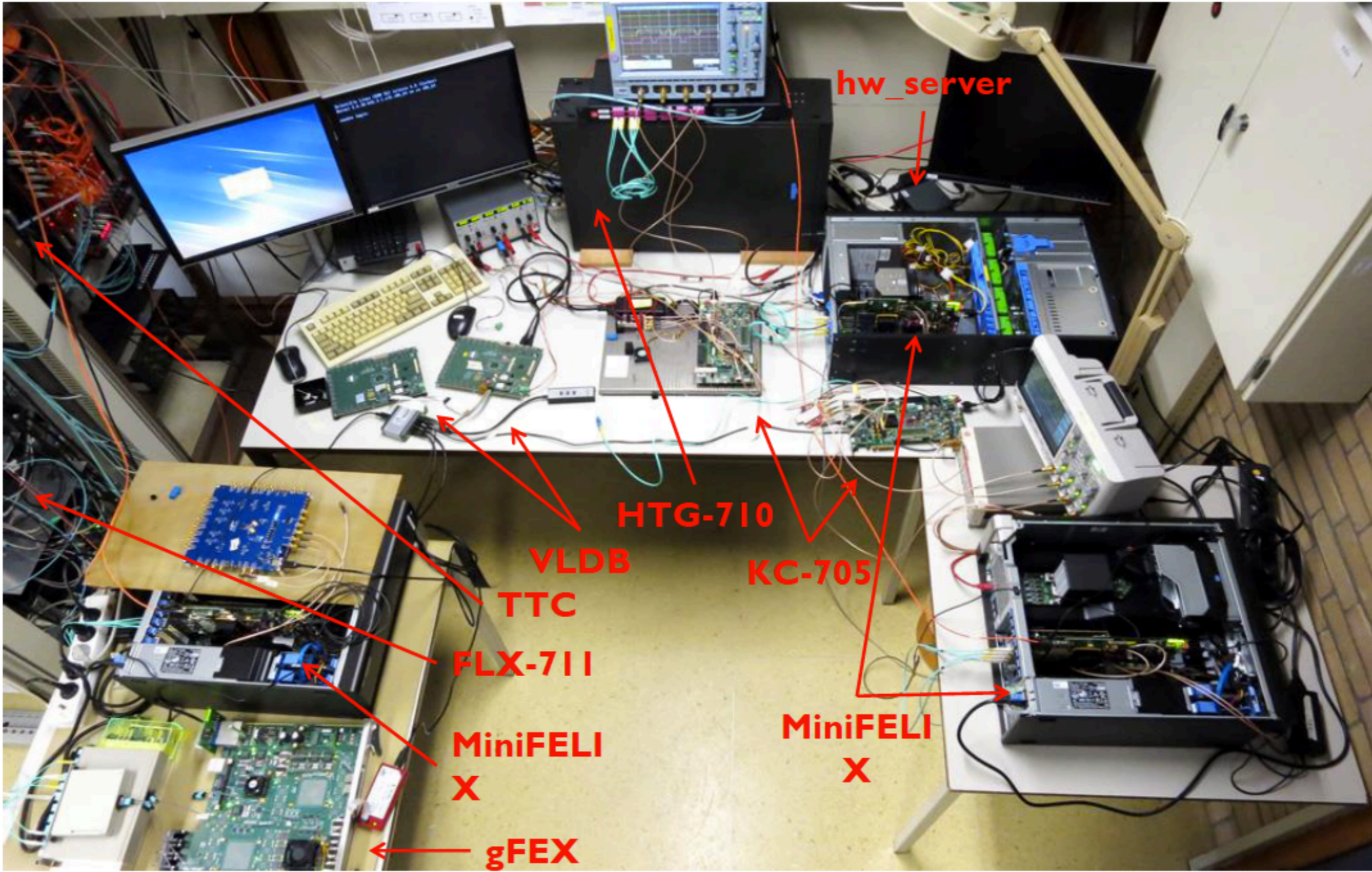
# Configuration and Monitoring

FELIX  
Configuration GUI



FELIX Application  
monitoring web  
server

# Integration of FELIX with other systems

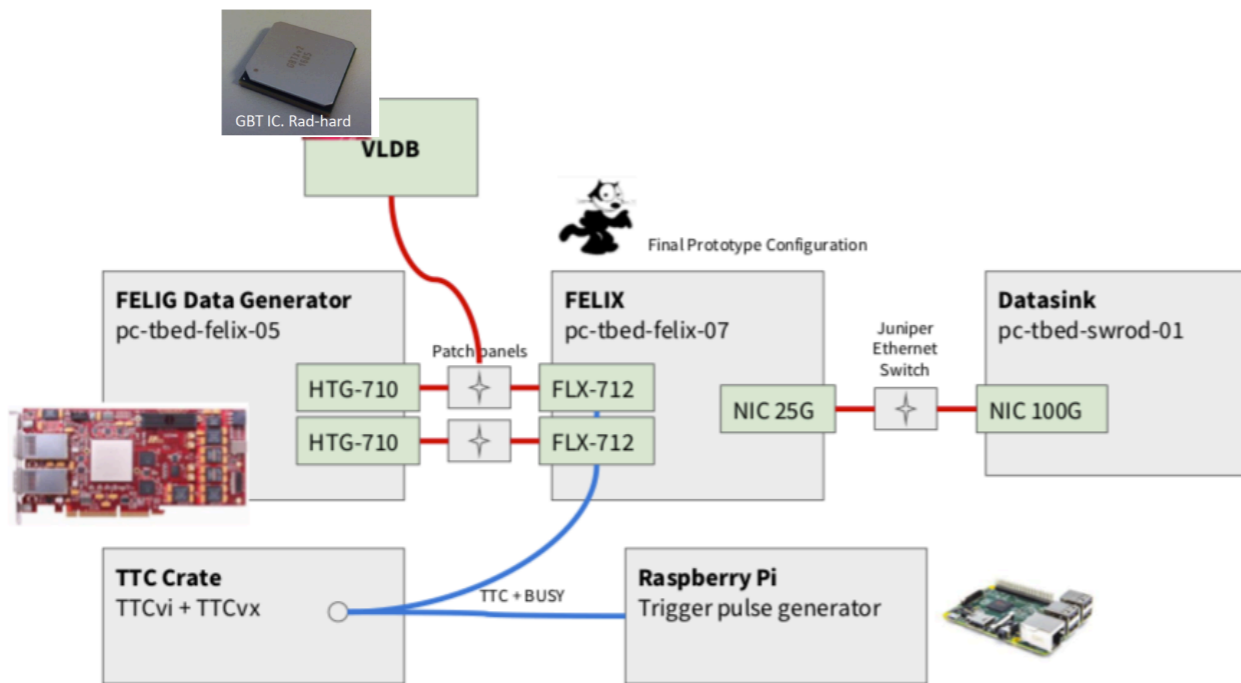
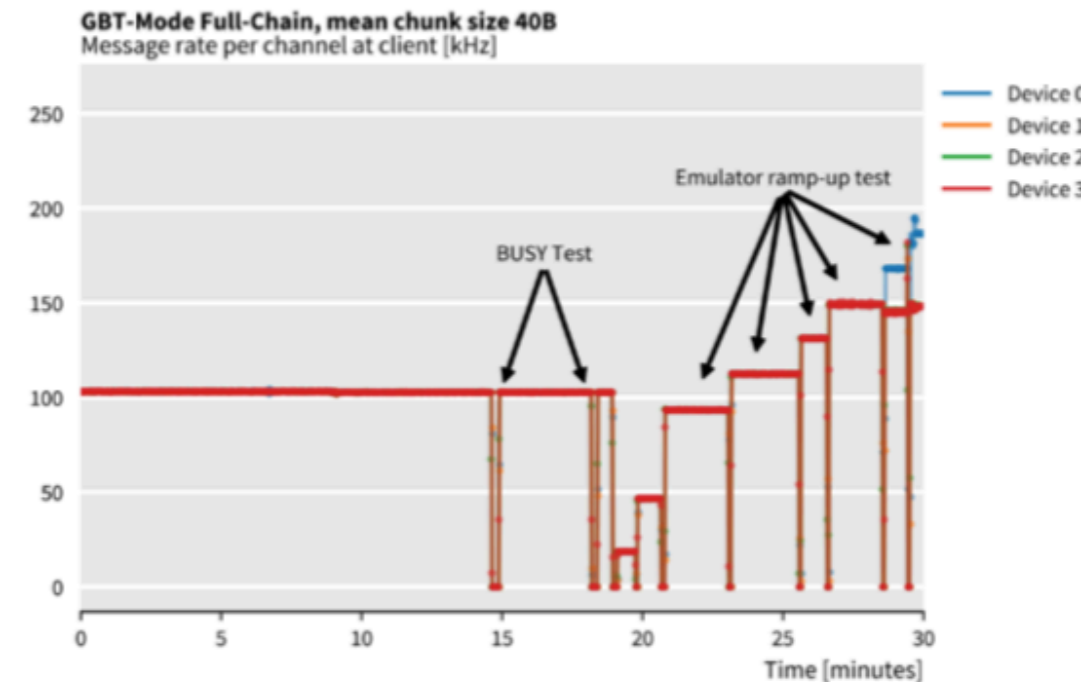
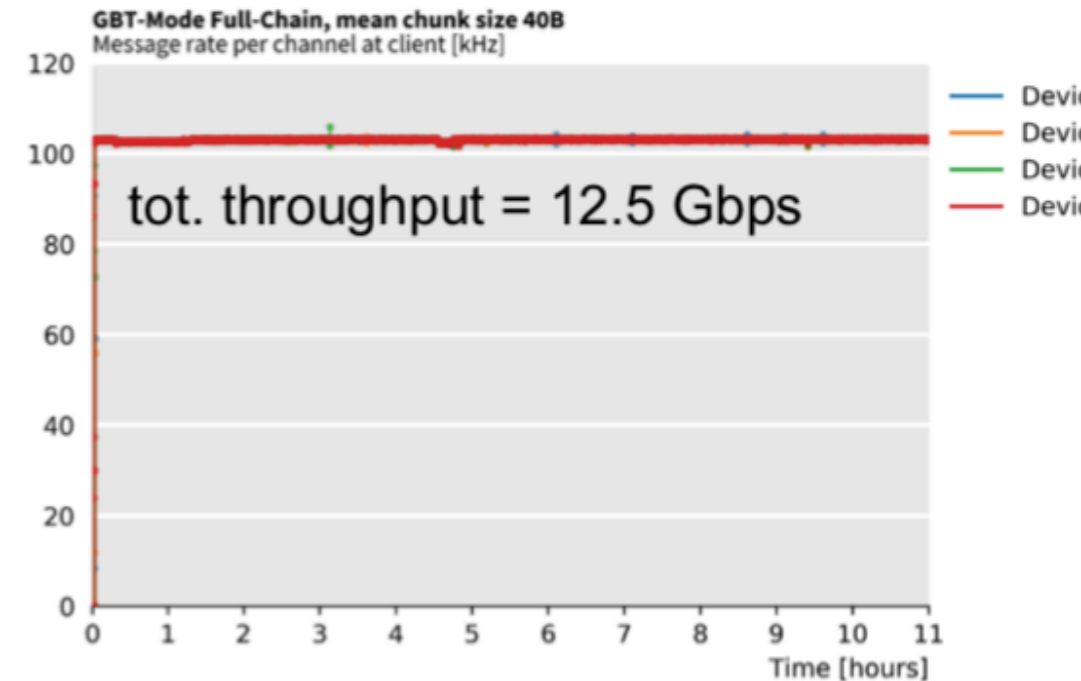




# Integration and Performances

- **Full-chain test: GBT mode**

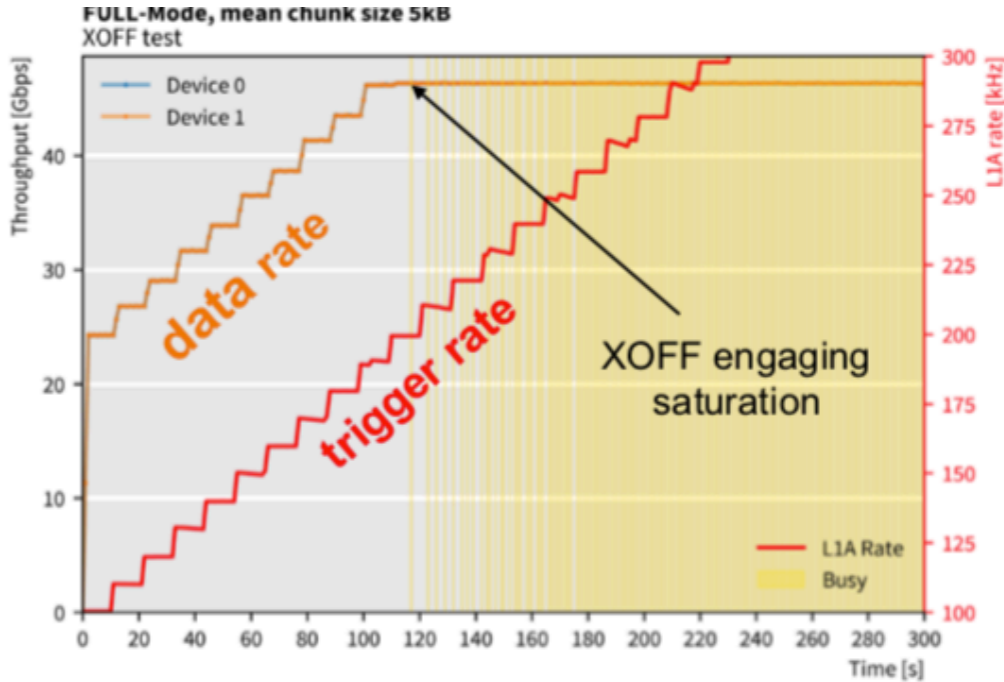
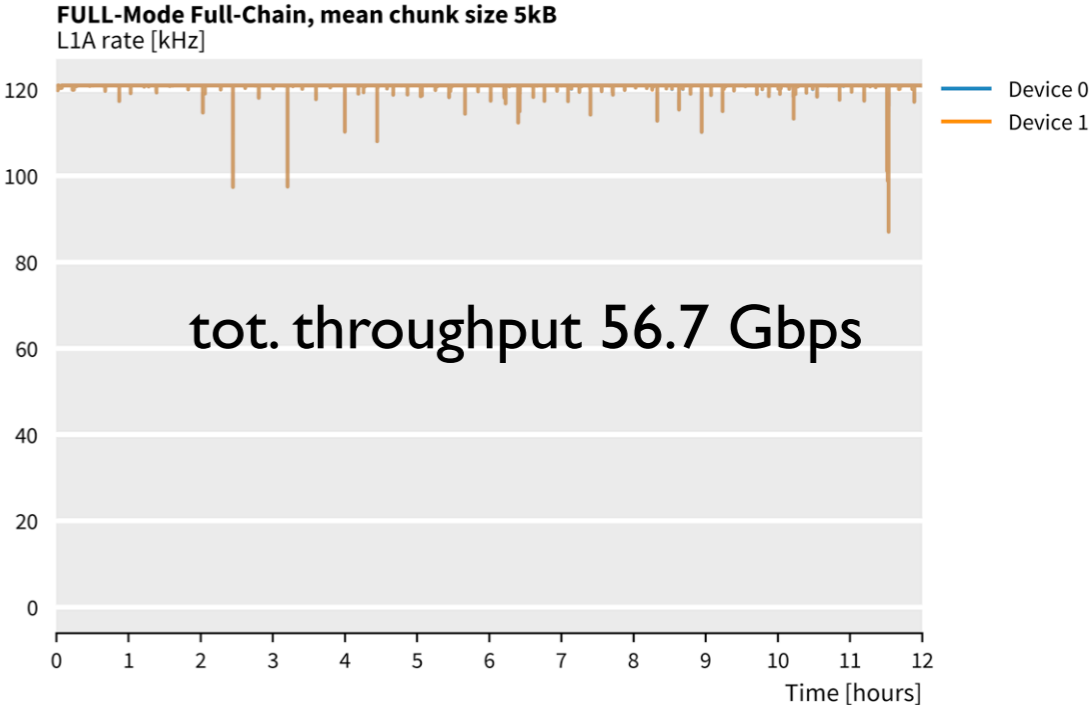
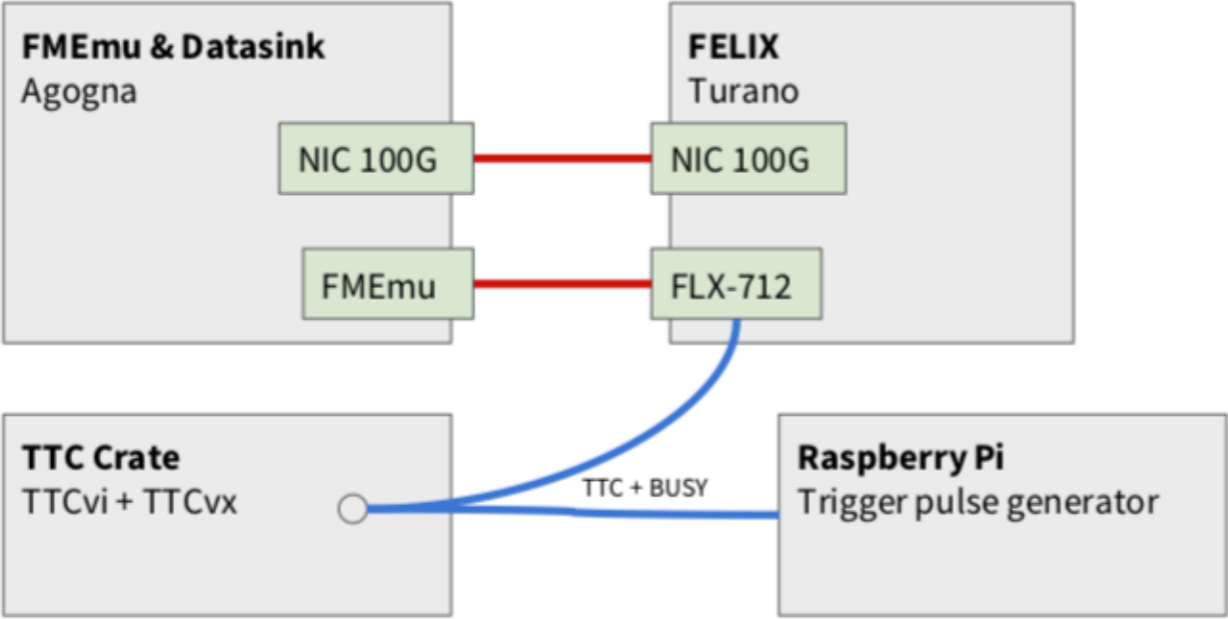
- full chain: emulator + FELIX + SW Rod
- long run in worst scenario: 432 e-links, 40B chunk, 100 kHz
- busy + detector control system (DCS) data testing



# Integration and Performances

- **Full-chain test: Full mode**

- full chain: emulator + FELIX + SW Rod
- long run in worst scenario: 12 links, ~5kB chunk, 120 kHz
- busy + traffic control (XOFF) testing



# Utilization in ATLAS

## • Liquid Argon Calorimeter

- ▶ LTDB (LAr Trigger Digitizer Board): integration testing ongoing with 40+ channels to monitor the FE and operate the TTC distribution (LAr Trig Dig Board), with MiniFELIX in FULL mode (LAr Digital Proc Blade)
- ▶ LI calorimeter trigger
- gFEX (Global Feature Extractor): connection established for 12 FULL mode links, long term stability ongoing
- ROD, Hub for eFEX (Electron Feature Extractor) and jFEX (Jet Feature Extractor): users in the process of setting up their test facilities
- TREX (Tile Rear Extension) modules: users in the process of setting up their test facilities

## • Muon Spectrometer

- ▶ New Small Wheels (NSW): sTGC (Small-strip Thin Gap Chamber) and MicroMegas (Micro Mesh Gaseous Structure) detector for
- ▶ muon tracking: integration of the FELIX system in the NSW Vertical Slice including the complete DCS (Detector Control System)
- ▶ chain, now targeting performance and long term stability
- ▶ BIS78 (Barrel Inner Small MDT (sector 7/8) ): users in the process of setting up their test facilities with FELIX

## • Tile Calorimeter

- ▶ Test system for Phase-II readout, initial communication established with the Tile PPr board in GBT mode
- ▶ Stepping toward FULL mode communication

## • Pixel sensors readout (for the control and readout of the ITk inner tracker)

- ▶ Test system for Phase-II ITk HV-CMOS pixel sensor R&D and Pixel demonstrator readout
- ▶ A FELIX system has been used to readout a telescope during recent HV-CMOS beam tests at CERN
- ▶ A vertical slice test stand for Pixel demonstrator readout with FELIX has been set up at CERN

# Utilization elsewhere

- Interest to use FELIX from other experiments
- The current most important group is the Proto-DUNE

## Collaboration

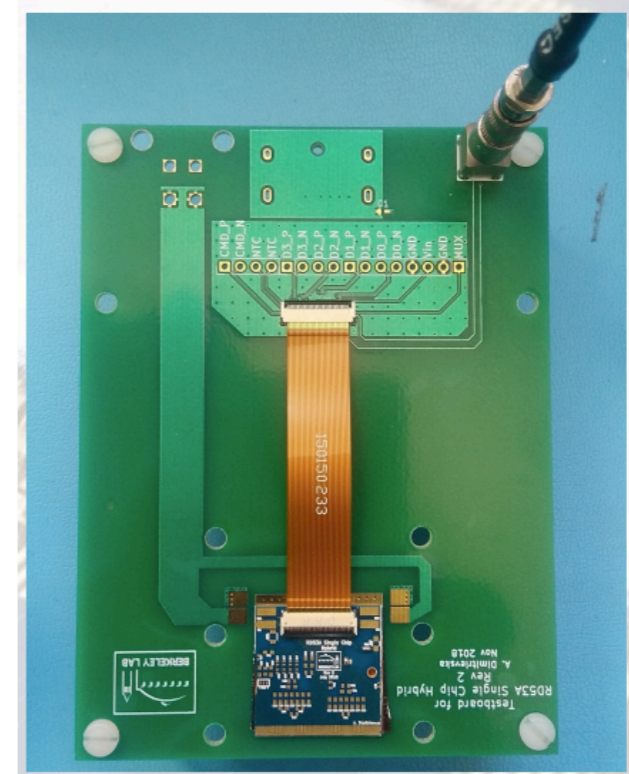
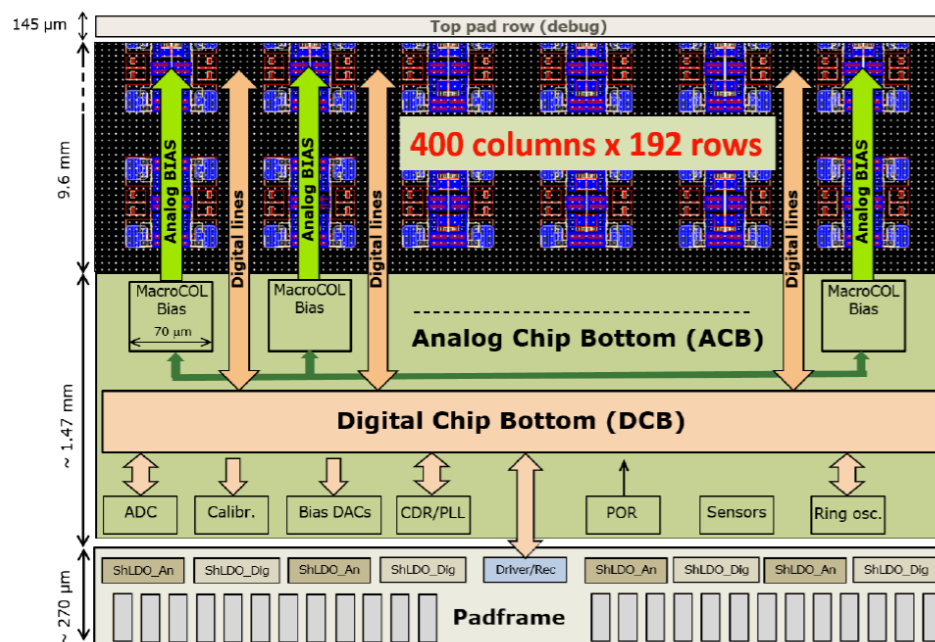
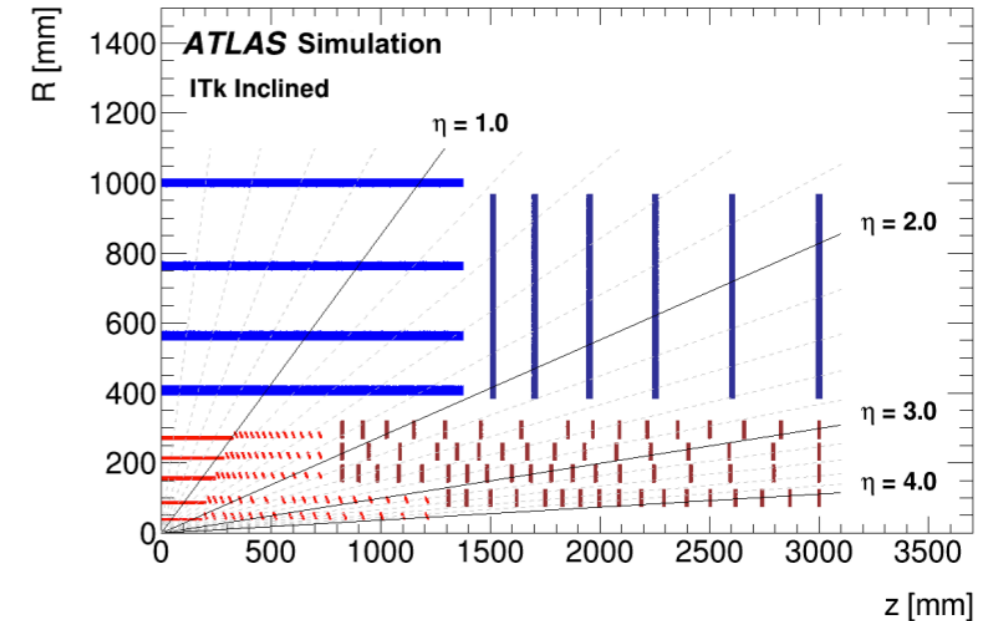
- ▶ FELIX emulator: software development ongoing using the FULL mode complete chain kit provided by the FELIX team (FULL Mode Generator + FULL Mode FELIX)
- ▶ WIB (Warm Interface Board): is being correctly read out by a FELIX system and long term stability testing is now being target

# ANL Involvement in Phase2 TDAQ/Pixel Detector



# Pixel Modules for the Phase 2 ITK

- Extended acceptance for the inner tracker: up to  $\eta=4$ 
  - ▶ pixel detector in the inner layers, strip detector in the outer layers. 13 hits max
- (Hybrid) Pixel Module design similar to the present one
  - ▶ hybrid pixels: passive silicon sensor, FE readout-chip (CMOS), and a flexible PCB
  - ▶ reduced size:  $50 \times 50$  or  $25 \times 100 \mu\text{m}^2$
  - ▶ higher radiation tolerance, lower threshold, power consumption
  - ▶ increased bandwidth: 5.12 Gbps

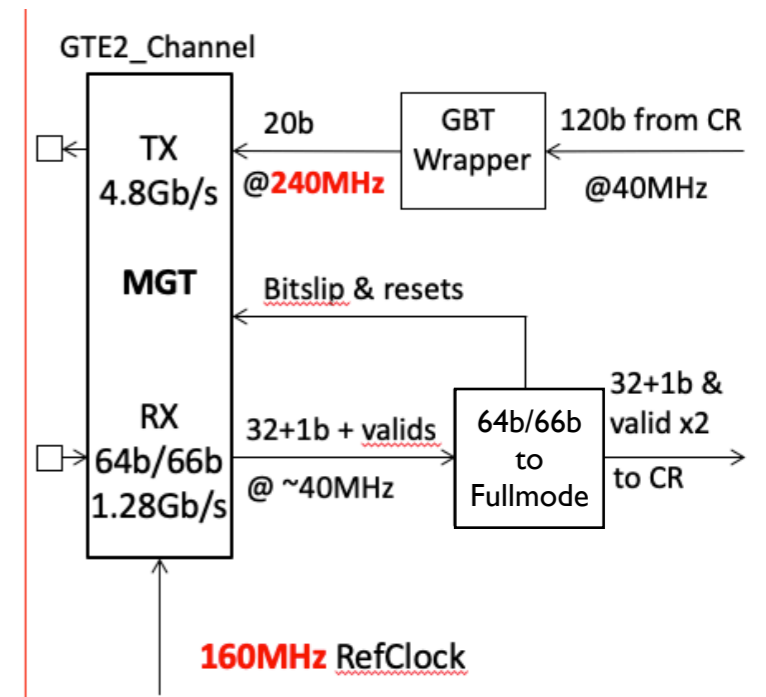
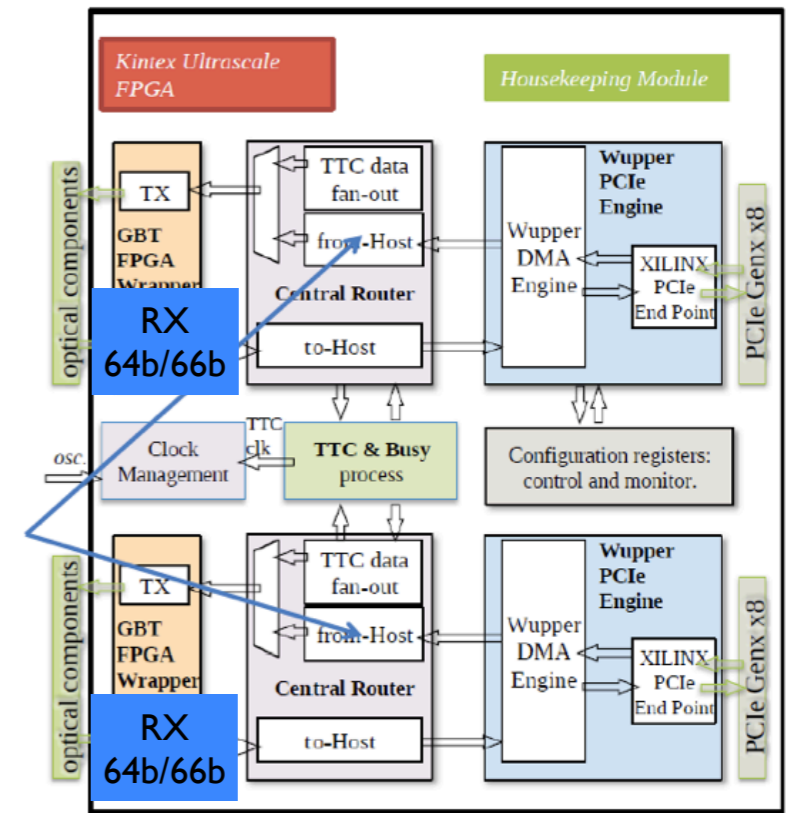


# Argonne involvement for Phase 2

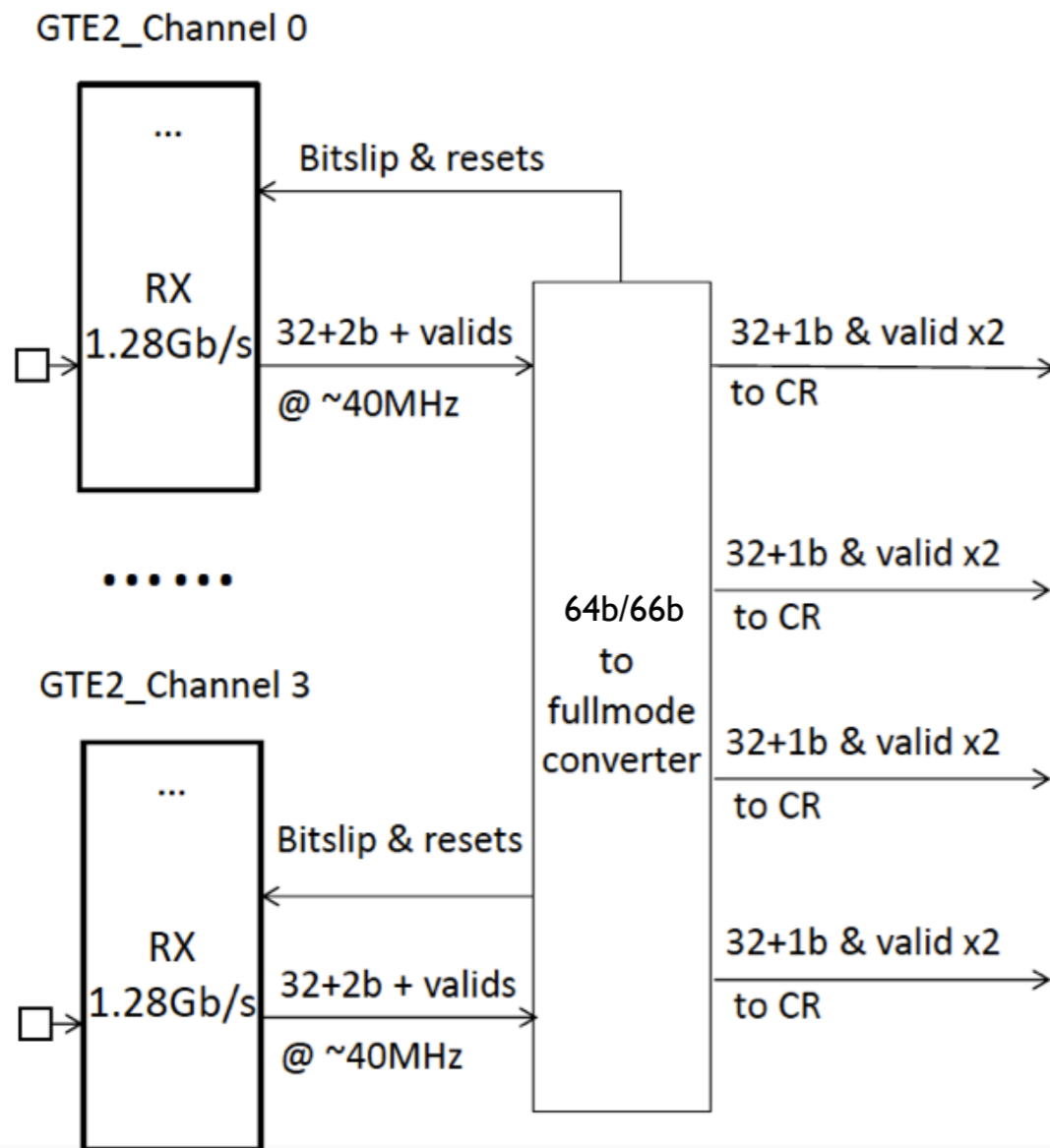
- Official Full mode FELIX F/W is modified to support 64b/66b @ 1.28 Gbps and a custom 160 Mbps serial stream

- ▶ FELIX software is agnostic of the changes
- ▶ MGT RefClock @ 160 MHz (rather than 240 MHz) to accommodate the change from 9.6 Gbps (Fullmode) to 1.28 Gbps (64b/66b)
  - ▶ software changed to configure the Si5345 jitter cleaner
- ▶ 4-bit e-link processor (EPROC, Central router) modified to comply with RD53 requirements
  - ▶ 16b words (e.g: trigger, calibration injections) are encoded into the bitstream

RD53A protocol encoder (EPROC) for the 160 Mbps stream



# FELIX F/W for RD53A readout

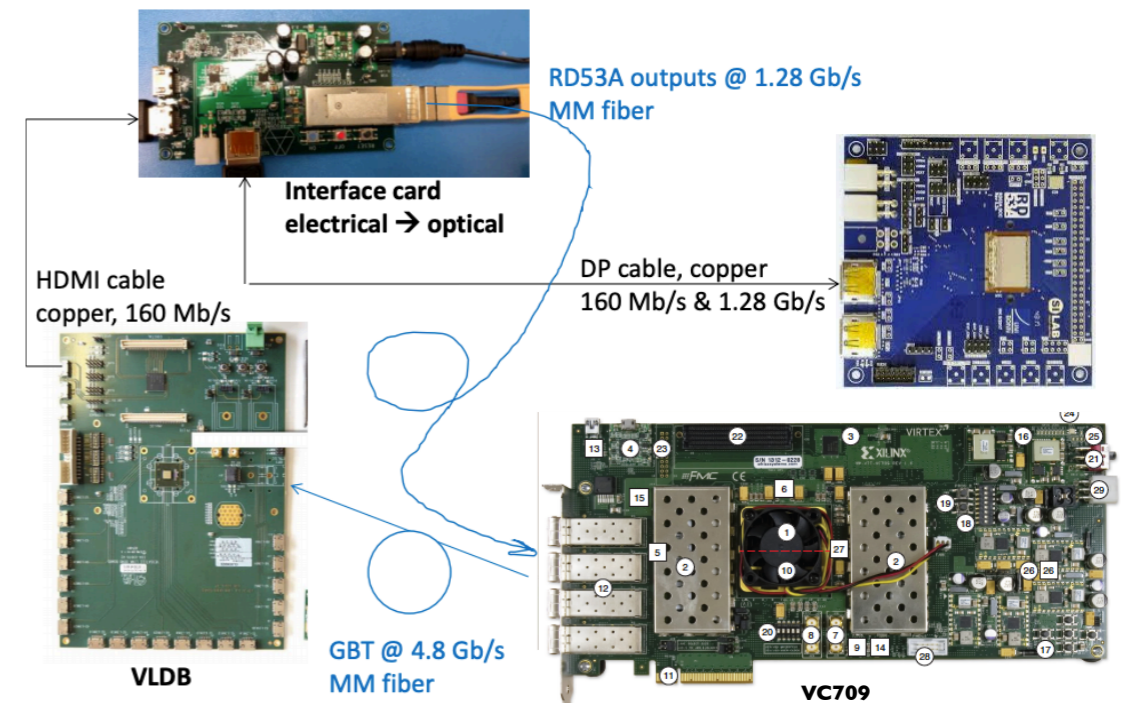
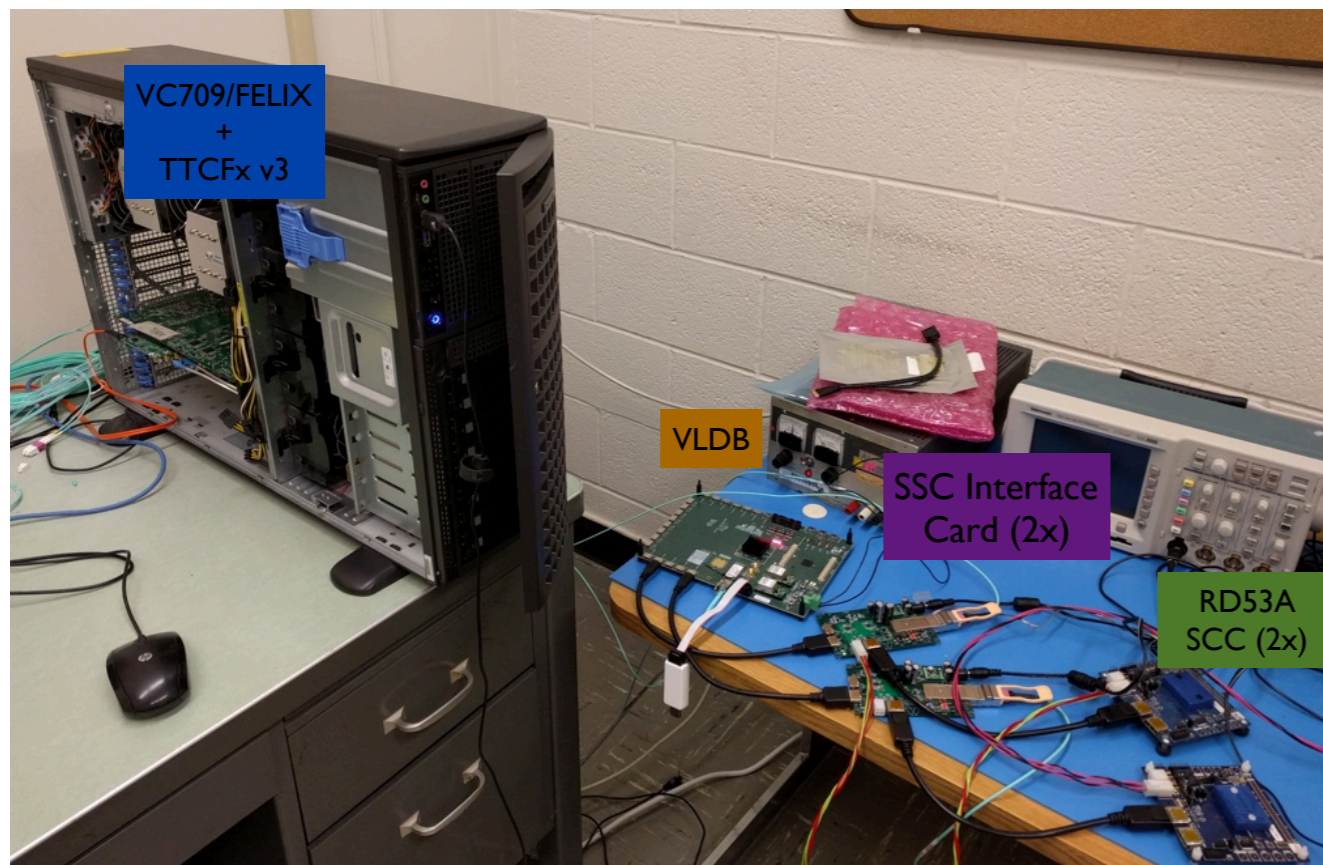


- The converter, initially based on Xilinx Aurora IP, is now custom
  - ▶ original IP is sub-optimal in terms of latency and resource usage
  - ▶ our custom implementation meets the RD53A and IEEE specs.
  - ▶ latency: 125 ns (was ~1 us) if lanes are not skewed
- Output format meets FELIX full mode spec
- The converter receives data from four transceivers (32b gearbox) to allow timing de-skew
  - ▶ de-skew can be turned off via one of the felix registers



# Current Setup @ ANL

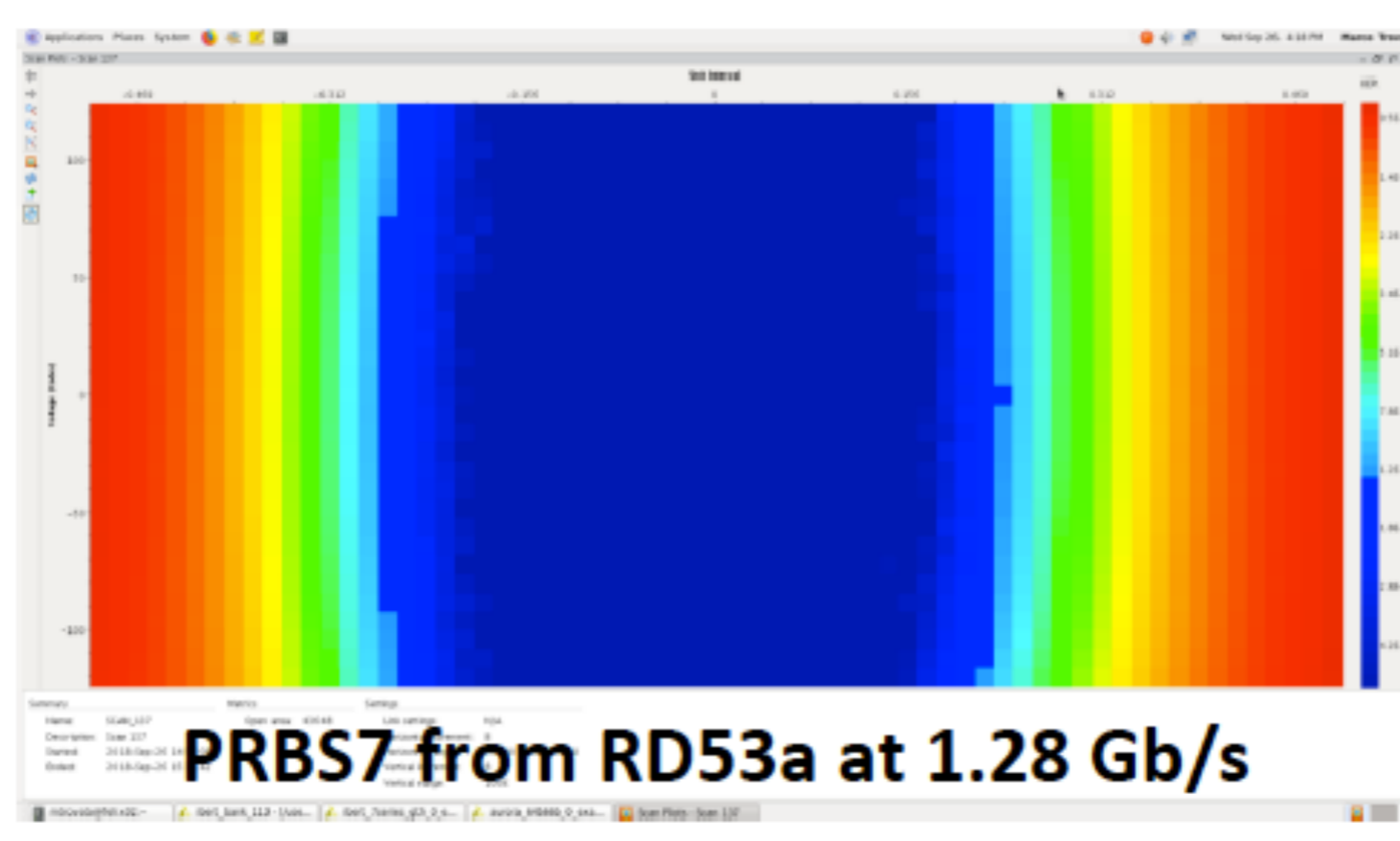
- RD53A Single Chip Card (SCC)
- SCC Interface Card (SCC IC)
- Versatile Link Demo Board (VLDB) w/ GBTx
- VC-709 FELIX board
- TTCfx V3 mezzanine card
- TTC system (not shown)





# Hardware Status

- Great link integrity at full (RD53A speed): hardware is healthy and transceiver settings well configured



# Testing Procedure

- RD53A scans and tunings are performed by using the Yet Another Rapid Readout (YARR) software

- ▶ YARR software, initially used for FE-I4 chips testing, was adapted for RD53As
- ▶ YARR relies on NETIO libraries
- ▶ scans and tunings undergo a common procedure: mask staging, double column loop, and trigger injection
  - ▶ mask staging activates only 1/16-1/32 of the pixel matrix, since not all the pixels can be read out at the same time
  - ▶ injection performed in a smaller pixel subset



- **Main routines:**

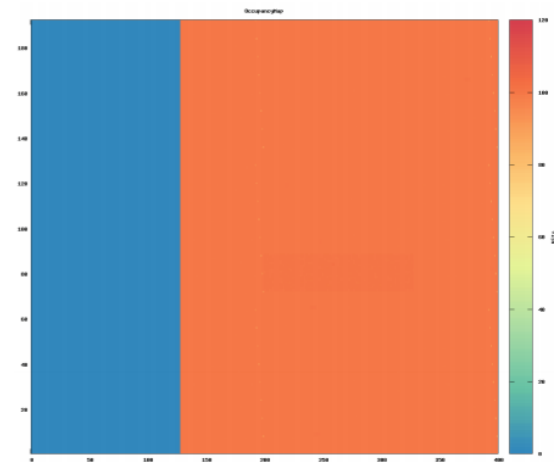
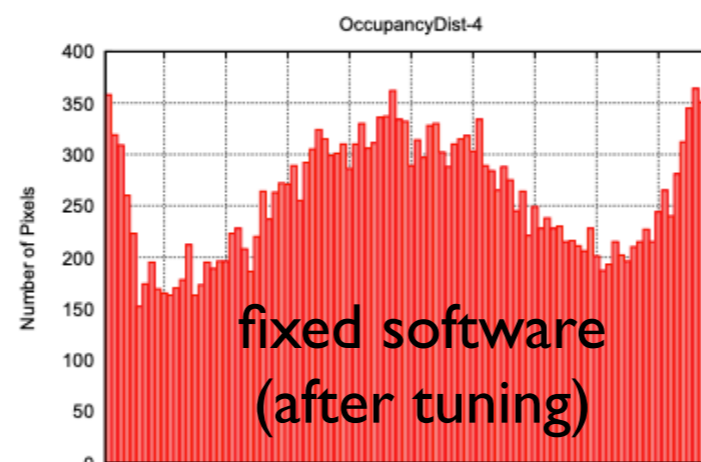
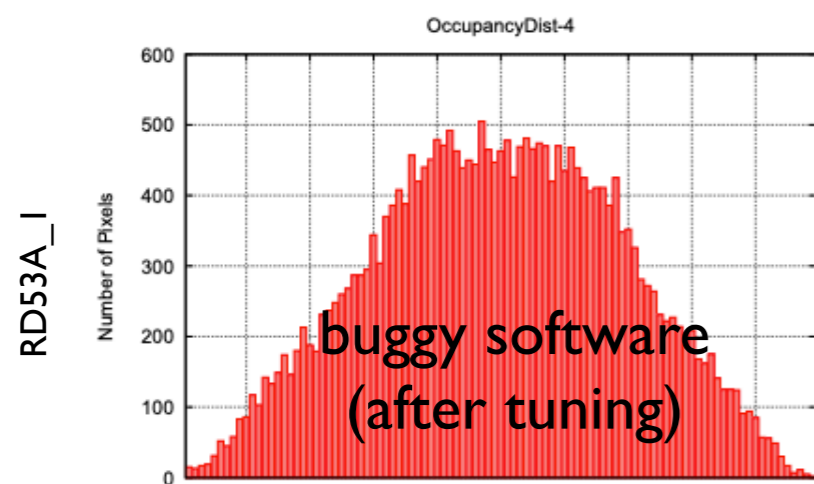
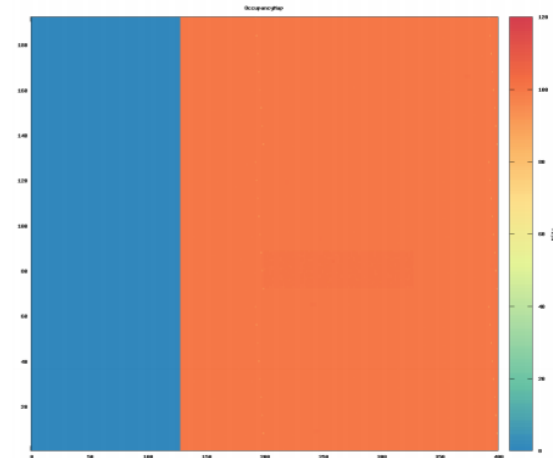
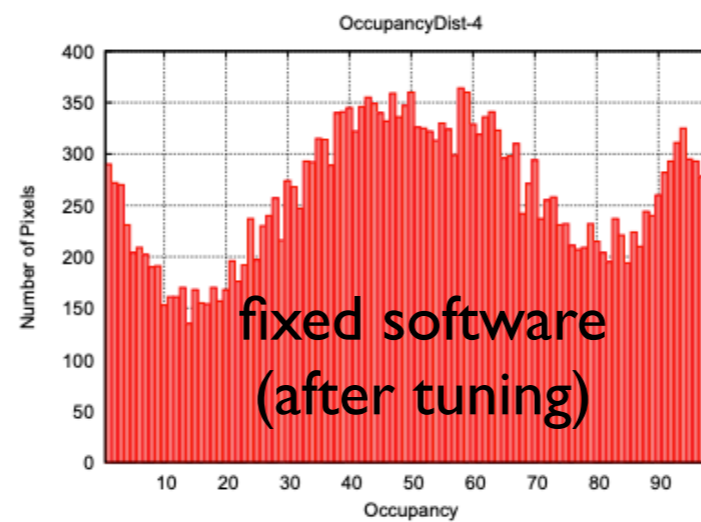
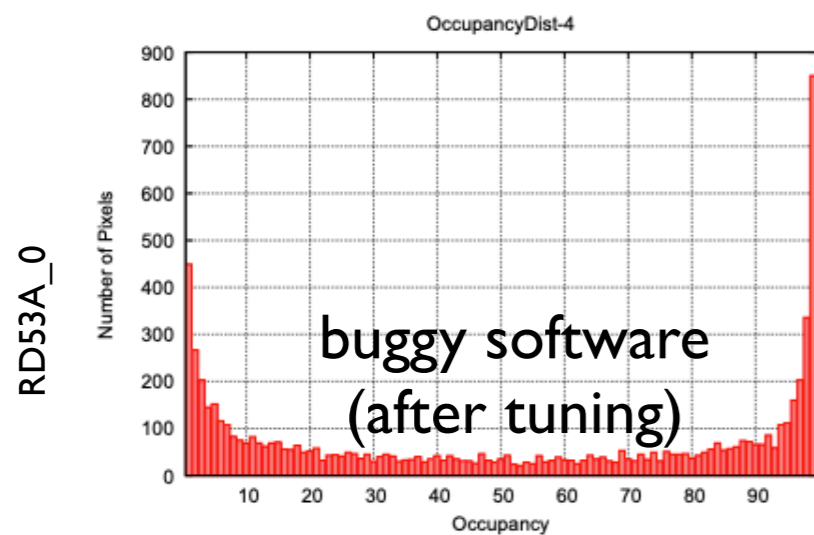
- ▶ analog scan: analog injection of very high charge is performed. Pixel with occupancy <99%, >101% discarded
- ▶ digital scan: same as analog but the injection is digital (ie: injection at discriminator level)
- ▶ global/pixel-by-pixel threshold scan: series of analog scans when the injection charge is varied.
- ▶ occupancy scan: same as analog scan but with charge injection close to the threshold. Global threshold chosen when the mean occupancy over the pixel matrix is 50%.

# Results

- Communication between FELIX and 4 RD53A established at LBL during the TDAQ work in May 2019
- Results shown only for the 2 RD53A results

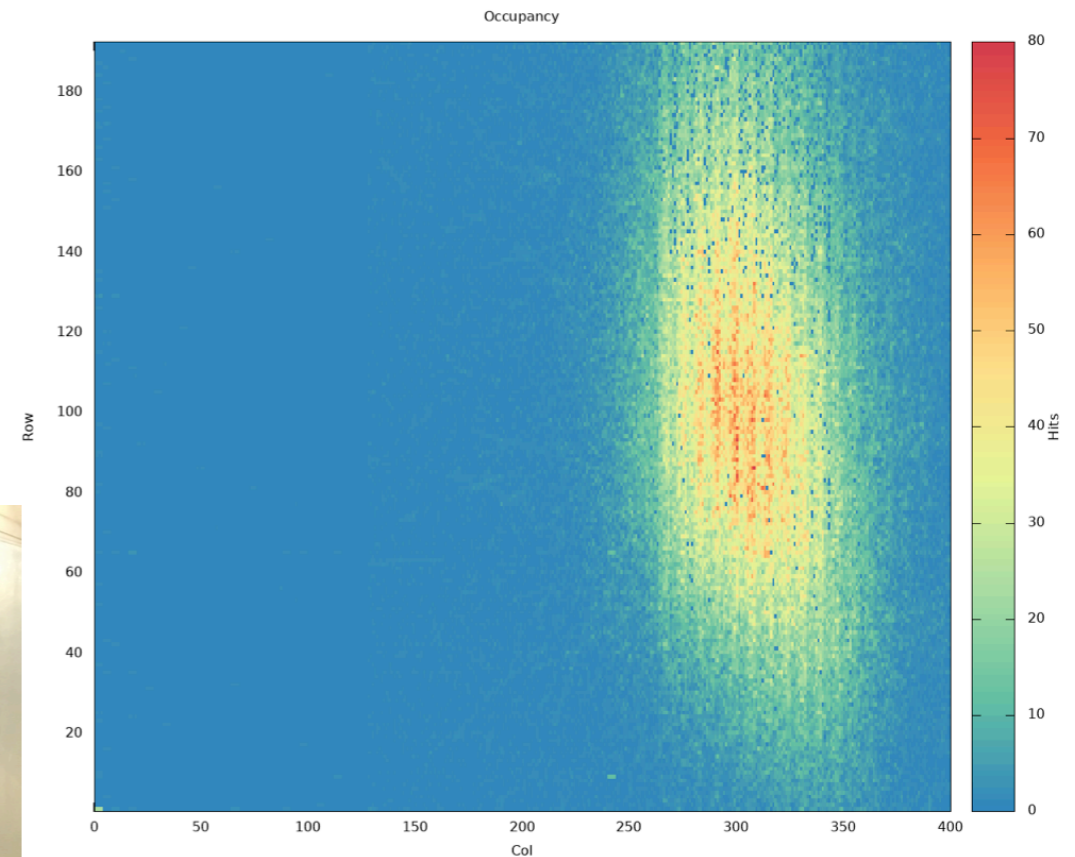
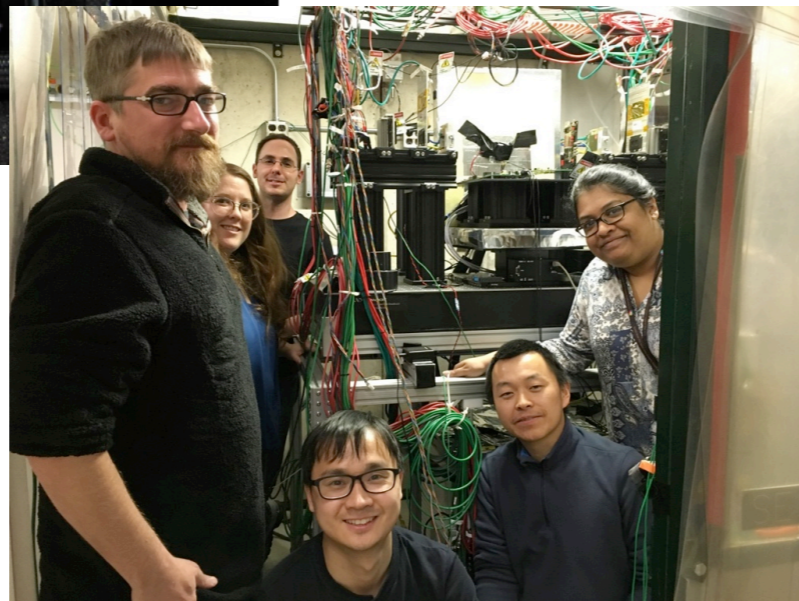
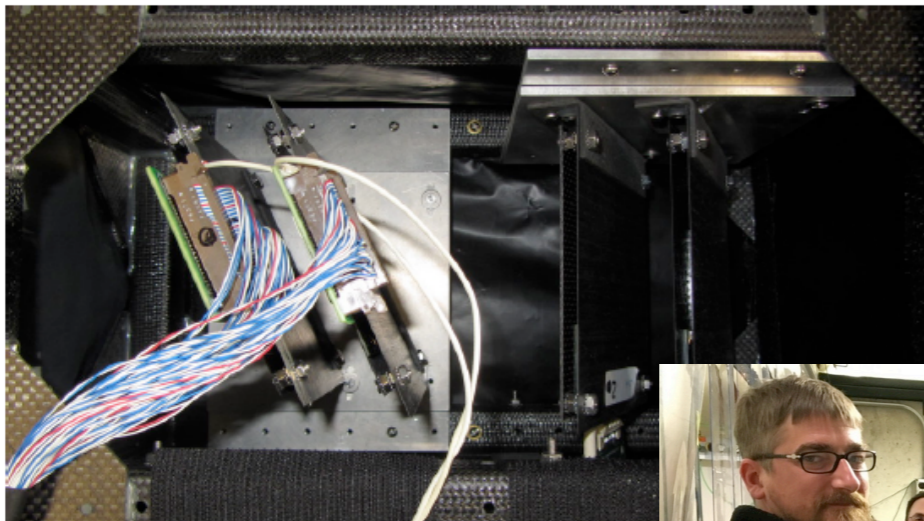


▶ awaiting for more RD53A at ANL



# Results

- Similar setup used at Fermilab during test-beam
- Real modules (sensors + FE chips) deployed in the telescope
- Firmware slightly changed to receive triggers from upstream strip test setup
- Beam spot clearly visible, but spread out
  - ▶ Not enough time to tune the chip
  - ▶ Will test it again in October



# Conclusions and Outlook

- FELIX is a router between FE serial links and commodity networks: it separates data transport from data processing
  - ▶ takes advantage of the latest technology to simplify the ATLAS readout
- In Run-3 (2021-2023) FELIX will be used by some detectors and trigger systems to interface w/ the data acquisition, TTC systems
- In Run-4 (2025-2035) this is planned for all ATLAS detectors
  - ▶ the FELIX PCIe card will be revised to benefit from the latest technology advancements
- FELIX supports GBT and FULL modes; Xilinx VC-709 and FLX-712 (16-lane PCIe Gen3 card) HW platforms
- FELIX has passed several integration tests with different sub-detector FEs
- Ongoing efforts:
  - ▶ Increase overall system reliability
  - ▶ Final performance benchmarking
- Procurement of Run-3 was FELIX in 2018, installation is now
  - ▶ passed the final readiness review...fabricating many FELIX boards

# Conclusions and Outlook

- **ANL has a fundamental role in FELIX. The main focus is on:**
  - ▶ data emulation to test the FELIX system (Phase 1,2)
  - ▶ pixel readout (Phase 2)
  - ▶ TTC interface (Phase 1,2)
  - ▶ increase overall reliability by helping in the debug process (Phase 1,2)
- **Pixel readout effort has been successful so far**
  - ▶ CERN is duplicating the ANL system for the system demo in July 2020
  - ▶ more institutions (Wuppertal, Bologna, Liverpool, LBL, ...) interested in using ANL readout system
- **A lot of experience gained, especially on board design**
- **Ongoing efforts:**
  - ▶ optimize 64b/66b decoder firmware
  - ▶ port firmware to FLX-712 for better scalability
  - ▶ continue module testing both at the lab and at the test-beam
  - ▶ finalize interface card
  - ▶ final performance benchmarking



THANK

A large, stylized blue 'X' graphic that spans across the width of the slide. The 'X' is composed of four thick, blue lines that intersect in the center. The top and bottom lines are horizontal, while the left and right lines are diagonal, meeting at the center. The 'X' is positioned behind the word 'THANK'.

for your attention