FELIX

Front End LInk eXchange

ATLAS EXPERIMENT

FELIX: the New Detector Interface for the ATLAS Experiment

Weihao Wu
Brookhaven National Laboratory, New York, USA
weihaowu@bnl.gov

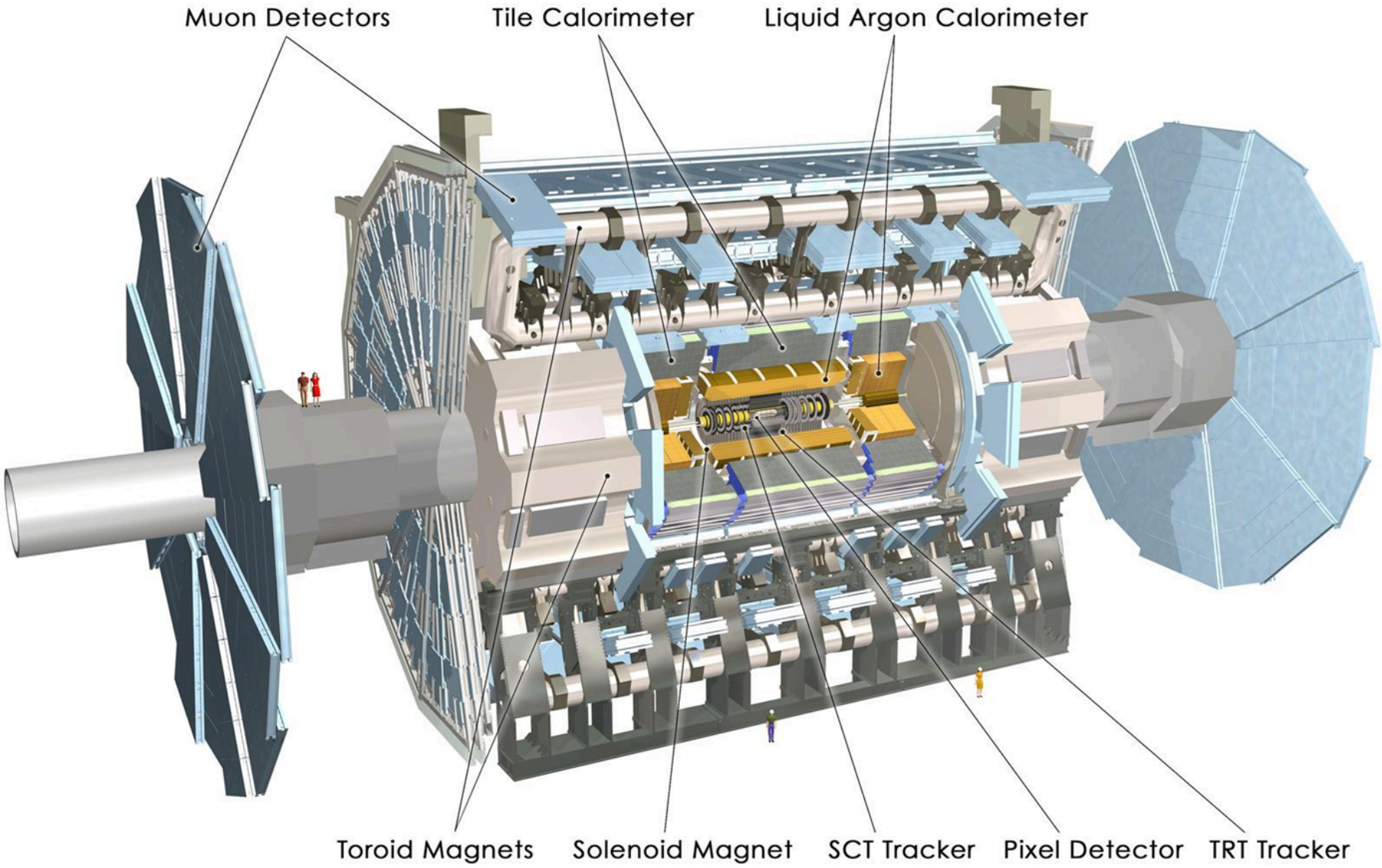
On behalf of the ATLAS TDAQ Collaboration

BROOKHAVEN NATIONAL LABORATORY

Outline

- ATLAS overview
- ATLAS trigger and DAQ system
  - Evolution plans in Run-3 and Run-4
- FELIX
  - Hardware
  - Firmware
  - Software
- Integration with detector front-ends
- Summary

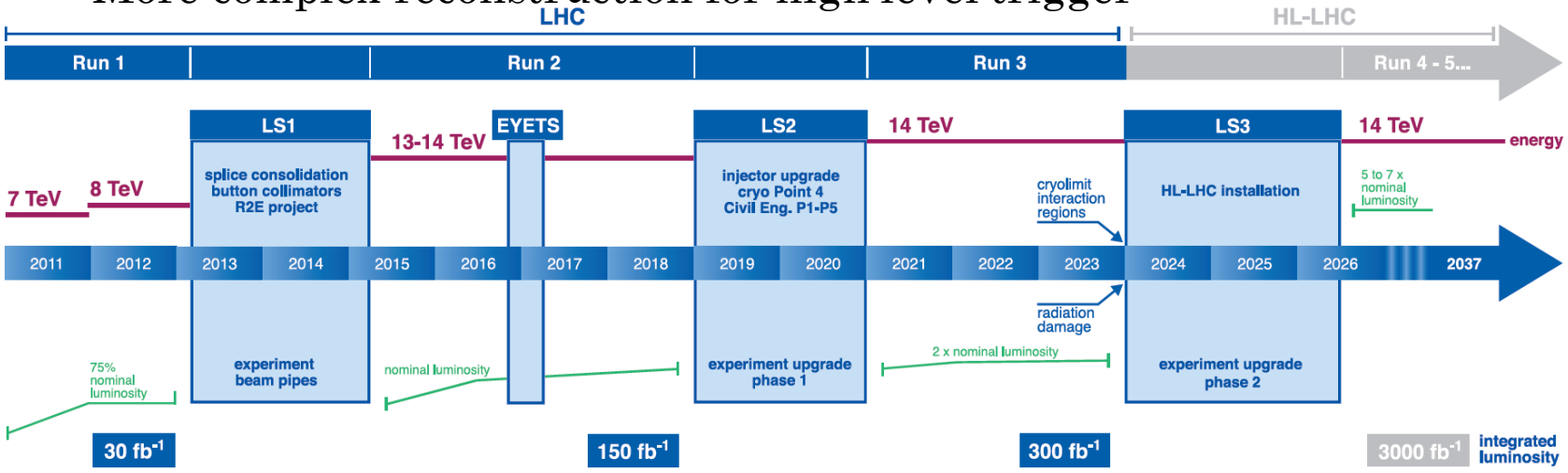
The ATLAS Experiment



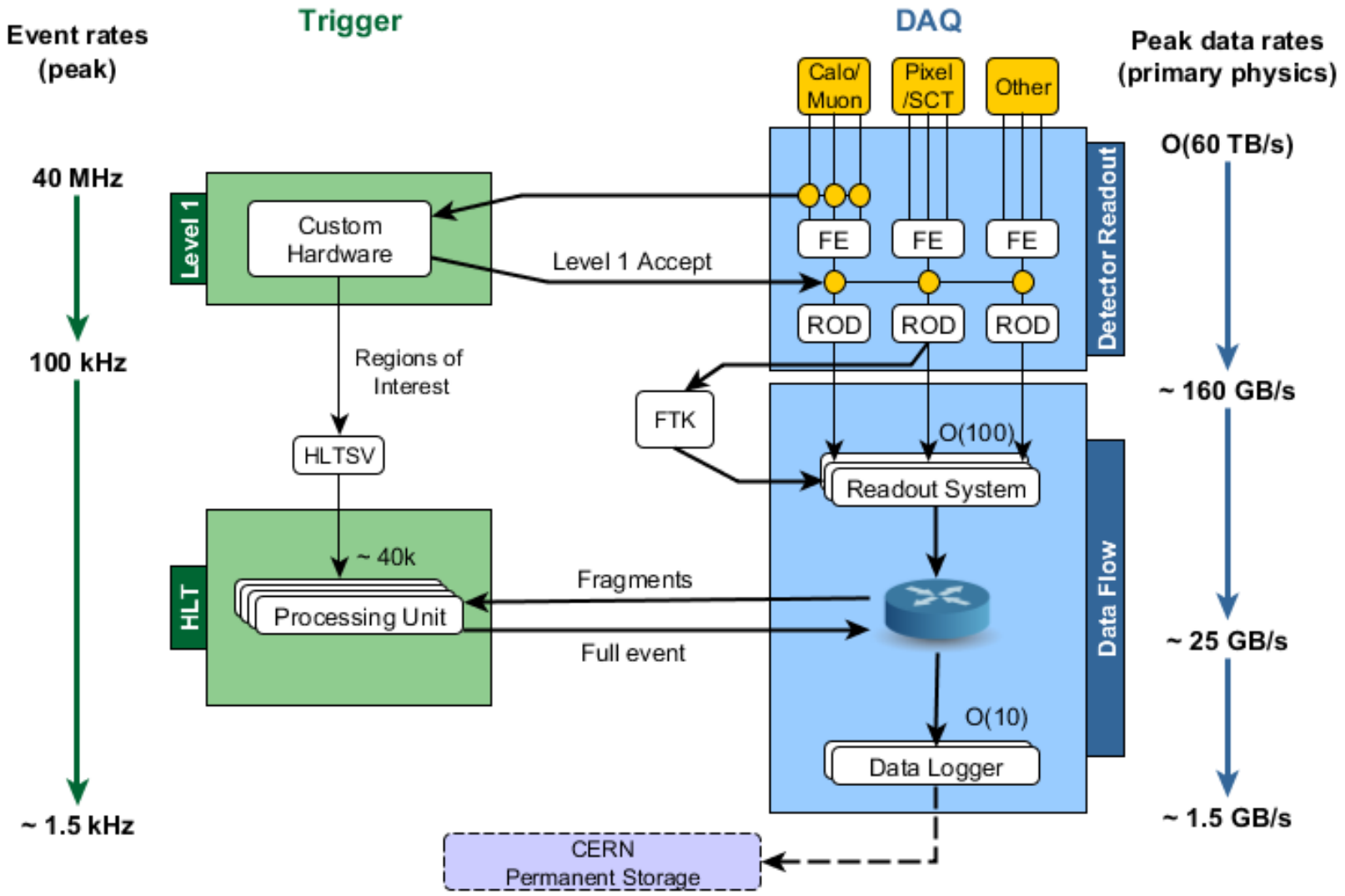
~ 100 million electronic channels

ATLAS Trigger Environment in Run-2

- Level-1 trigger system:
  - Fast analysis of all collision events with reduced resolution
  - Select only interesting events
    - 1.5 MByte × 40 MHz = 60 TByte/s (impossible to record all of LHC data)
- LHC Run-2:
  - Energy increase 8→13 TeV
  - Peak luminosity increase 0.8→21.4 × 10^{33}cm^{-2}s^{-1}
- Strategies for managing trigger rates
  - Better feature identification algorithms for level-1 trigger
  - Increase detector granularity to handle pileup
  - More complex reconstruction for high level trigger



ATLAS Trigger/DAQ in Run-2

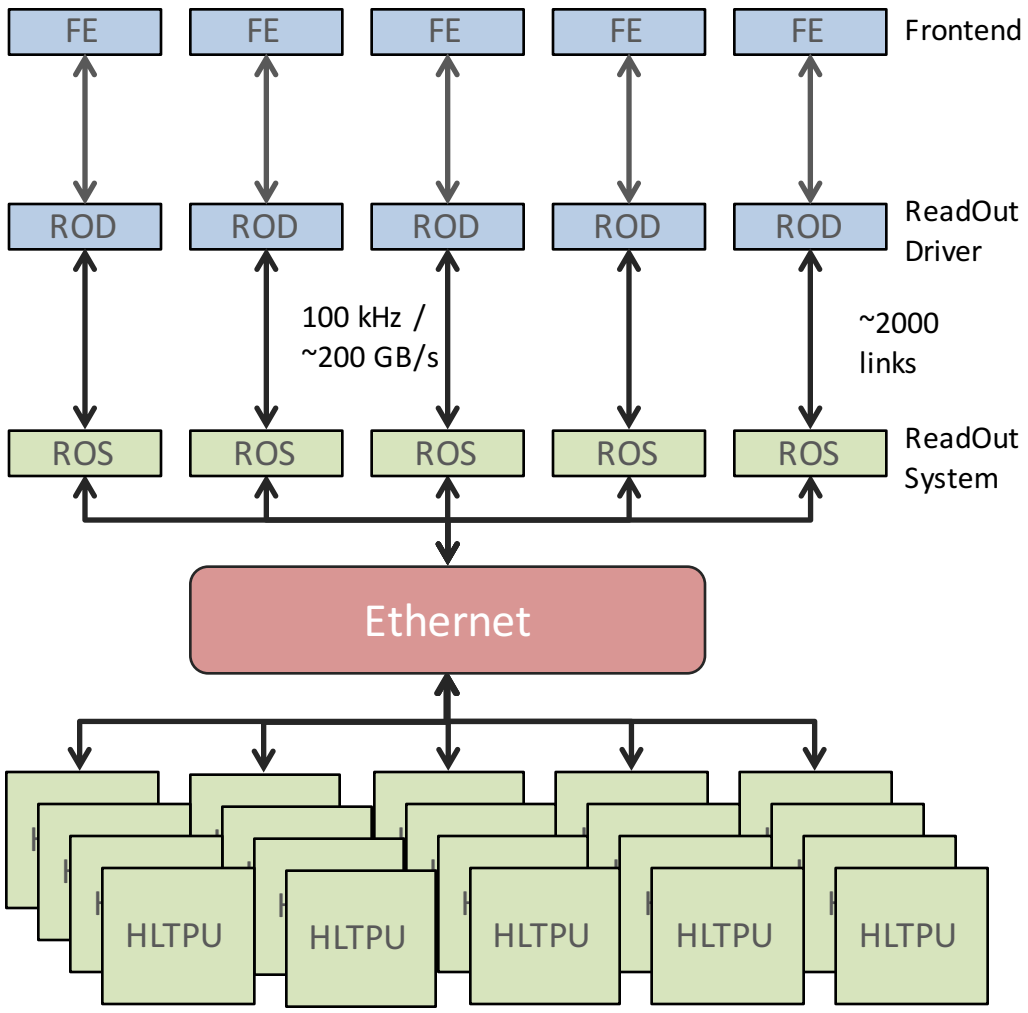


ATLAS DAQ Today

FELIX

Custom point-to-point links

Point-to-point S-LINKs*



Custom electronic components

PCs (COTs)

**S-LINK* is a CERN specification for an easy-to-use FIFO-like data-link which can be used to connect front-end to read-out.

* COTs: Commercial off-the-shelf

Today

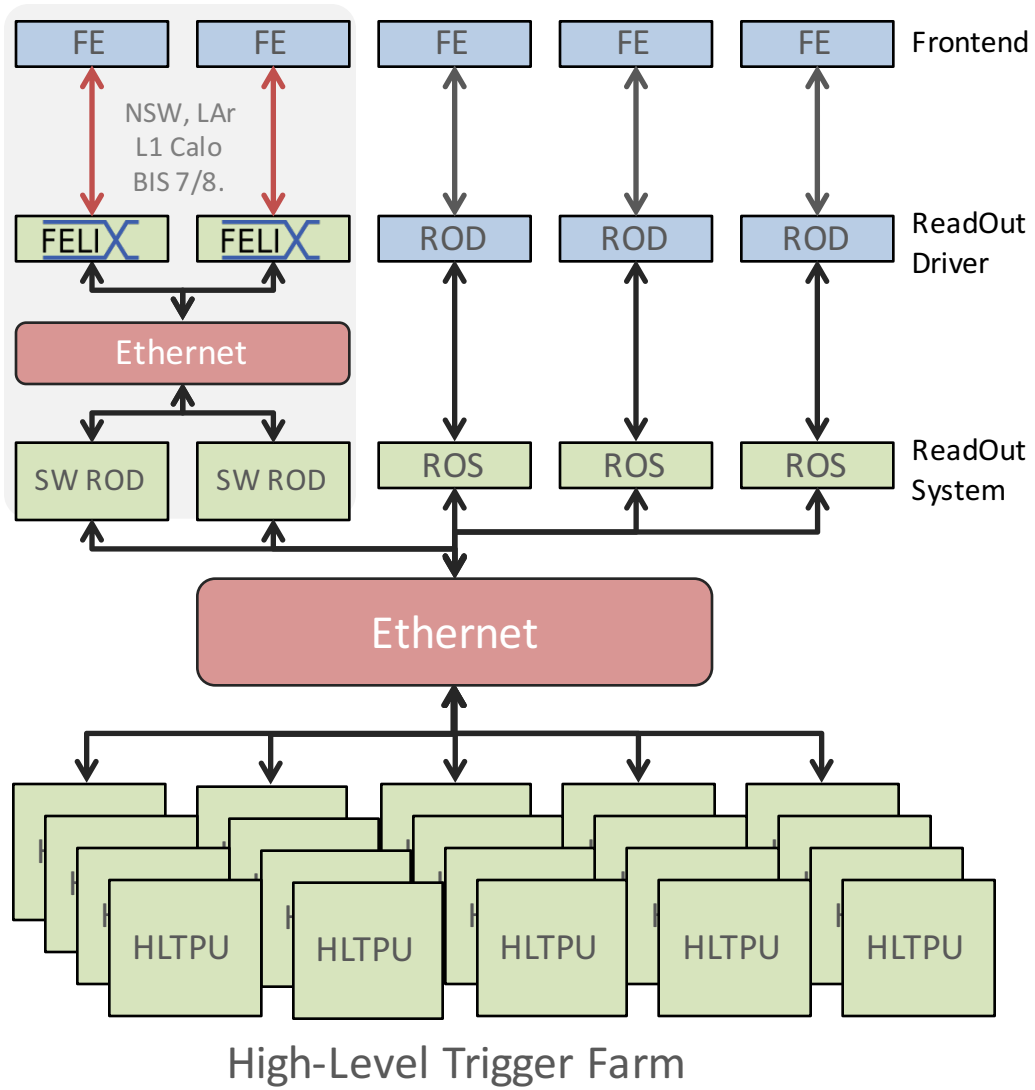
Run 2 2015 2017 | Run 3 2019 2021 2023 | Run 4 2025

Upgrade for Run 3

GBT links* or
FULL mode links

PCs

100 Gb Ethernet



Custom electronic components including FELIX cards

PCs (COTs)

* GBT: GigaBit Transceiver with Versatile Link

* COTs: Commercial off-the-shelf

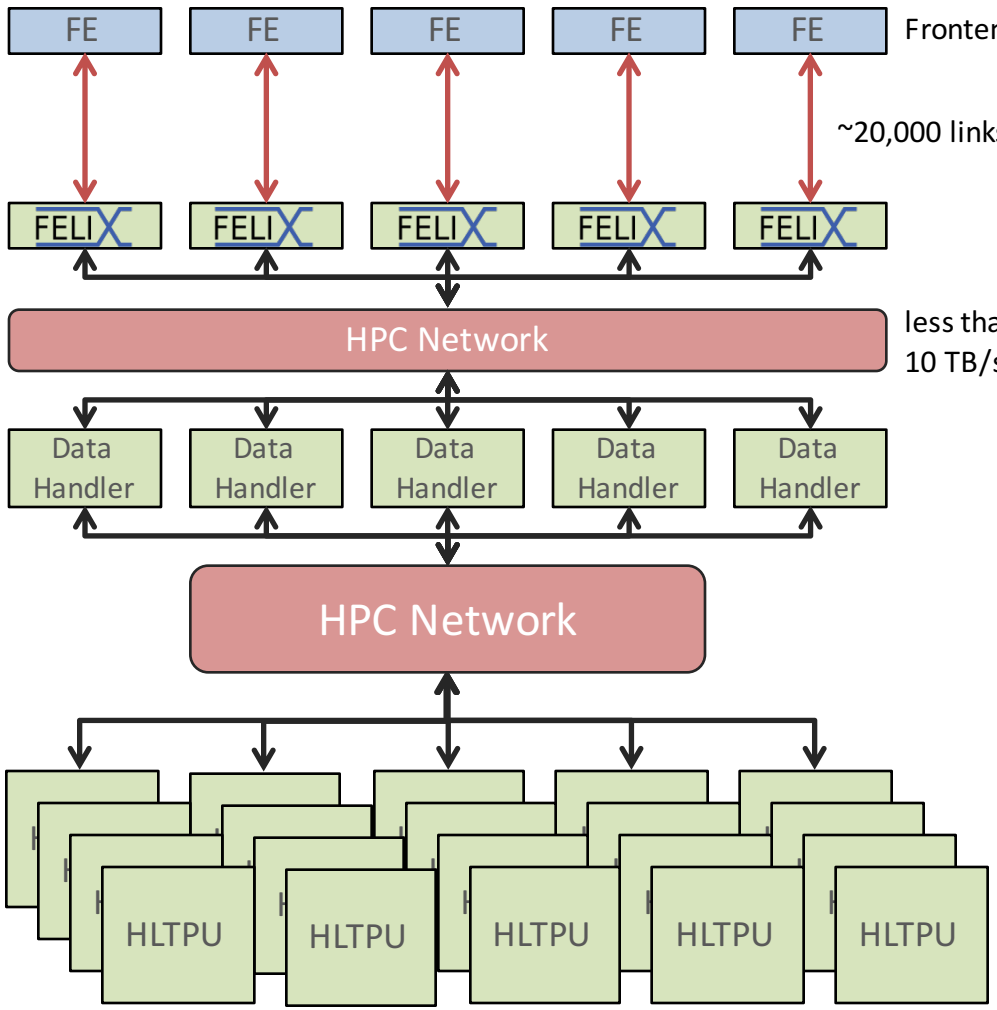
2019

Run 2 2015 2017 2019 Run 3 2021 2023 Run 4 2025

Upgrade for HL-LHC

GBT, LpGBT* or FULL mode links

COTS network technology



Custom electronic components including FELIX cards

PCs (COTs)

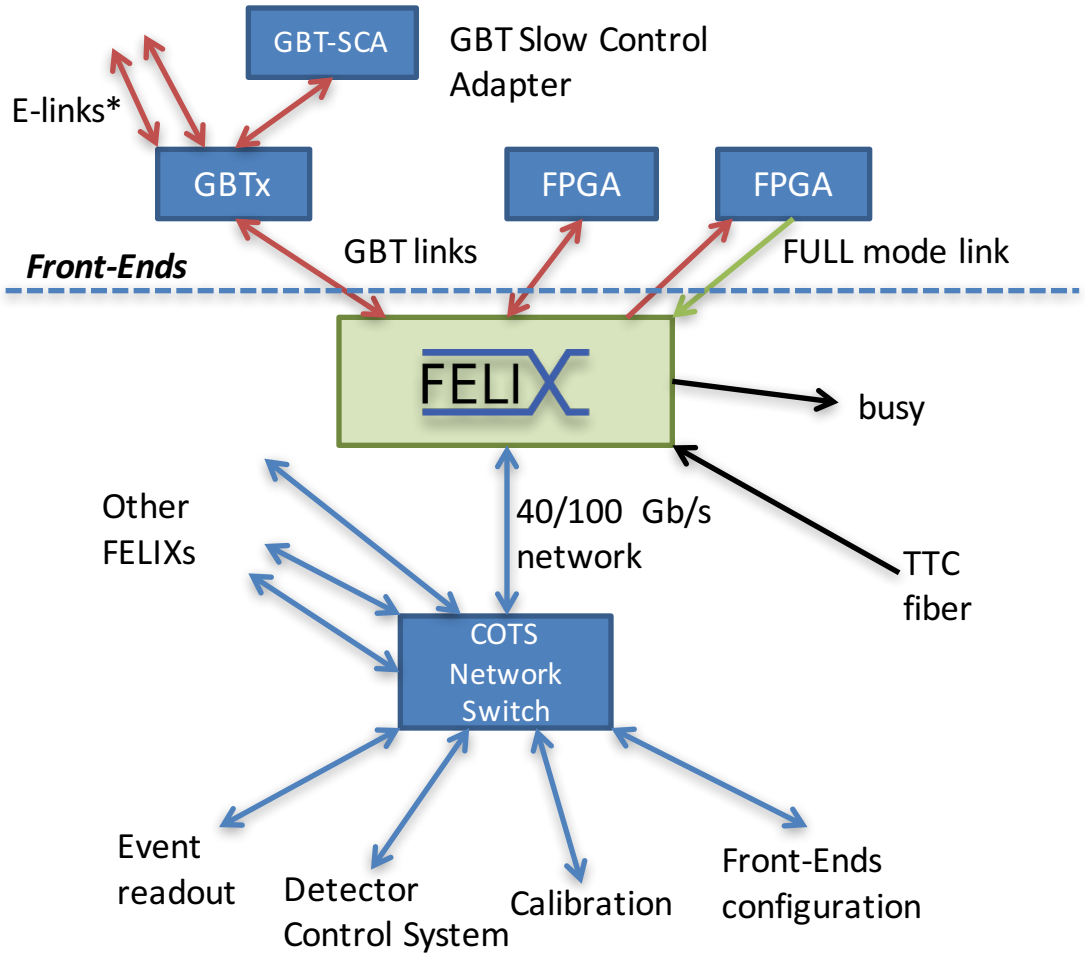
*LpGBT: Low power GBT

* COTs: Commercial off-the-shelf

2024

Run 2 2015 2017 2019 Run 3 2021 2023 Run 4 2025

FELIX Functionality

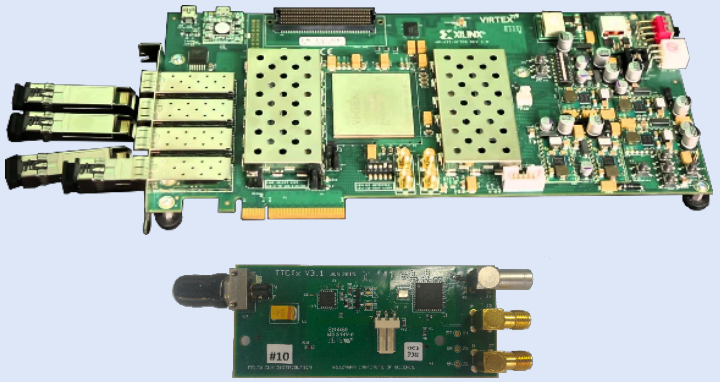


- FELIX is a router between front-end serial links and a commodity network, which separates data transport from data processing.
- Routing of detector control, configuration, calibration, monitoring and detector event data
- TTC (Timing, Trigger and Control) distribution integrated
- Configurable E-links in GBT Mode
- Detector independent

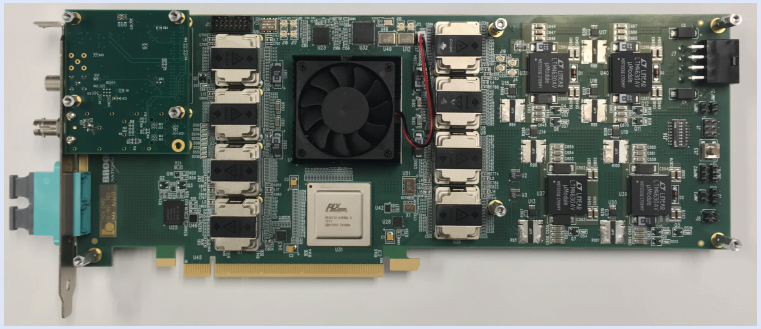
*** E-link:** variable-width logical link on top of the GBT protocol. Can be used to logically separate different streams on a single physical link.

FELIX I/O Cards

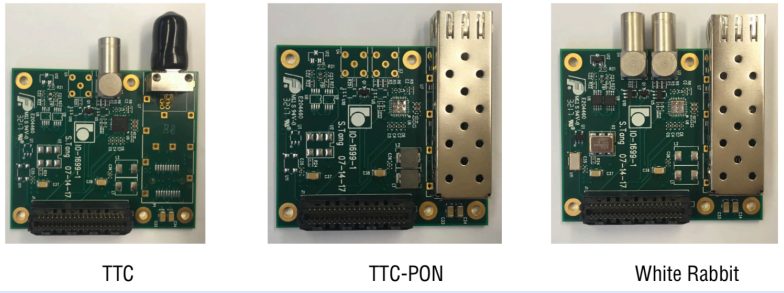
- **VC709 from Xilinx (FLX-709 or MiniFELIX)**
  - Only for development
  - Virtex7 X690T FPGA
  - 4 optical links (SFP+)
  - PCIe Gen3 x8
- **TTCfx (v3) mezzanine card**
  - TTC input
  - ADN2814 for TTC clock-data recovery
  - Si5345 jitter cleaner



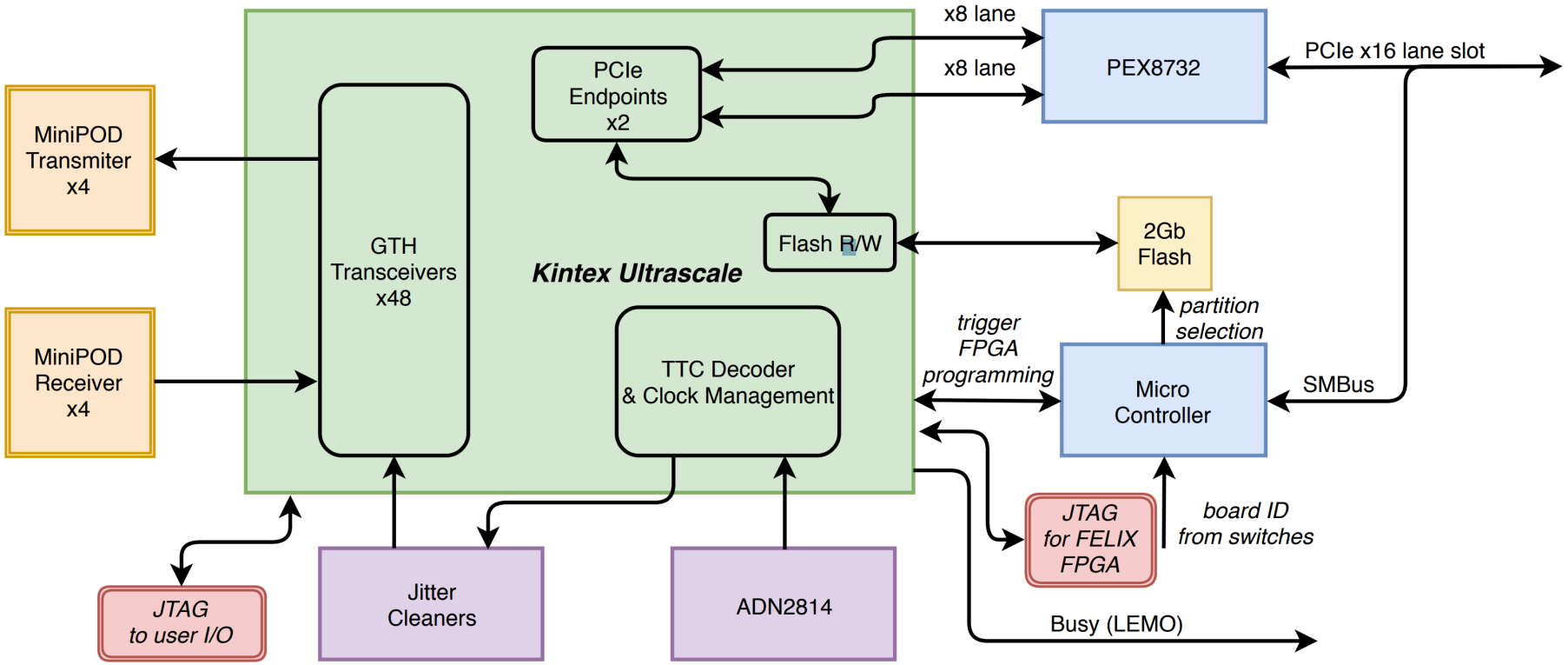
- **FLX-712 (or BNL-712)**
  - Final production board
  - Xilinx Kintex UltraScale XCKU115
  - 48 optical links (MiniPODs)
  - TTC input ADN2814
  - PCIe Gen3 x16 (2x8 with switch)
  - Si5345 jitter cleaner



- **Timing mezzanine card**
  - Supports TTC, TTC-PON, White Rabbit
  - TTC configuration is for Run 3

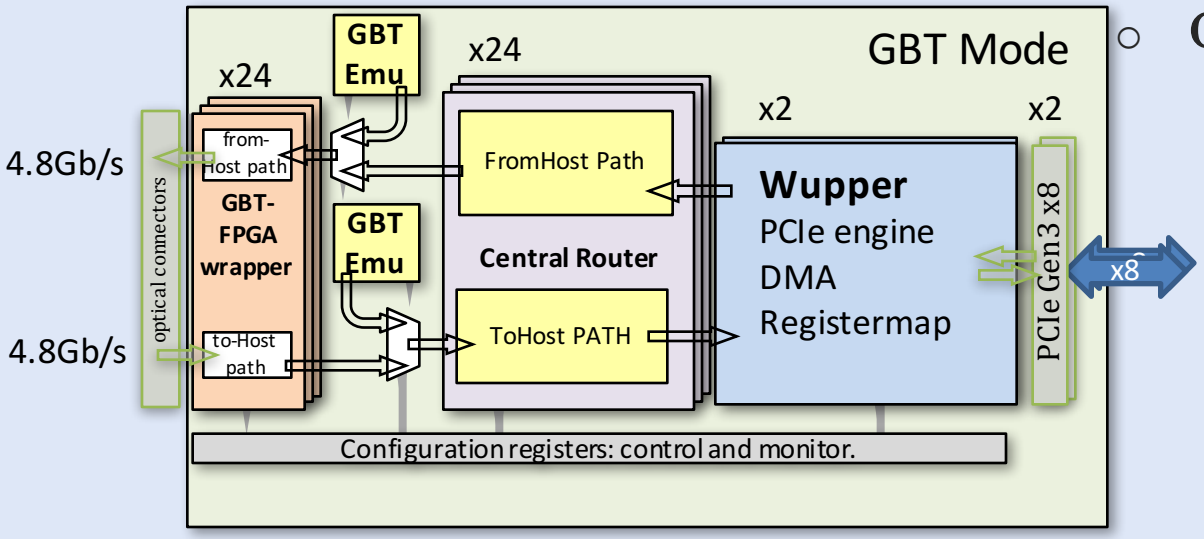


FLX-712 PCIe Card Features

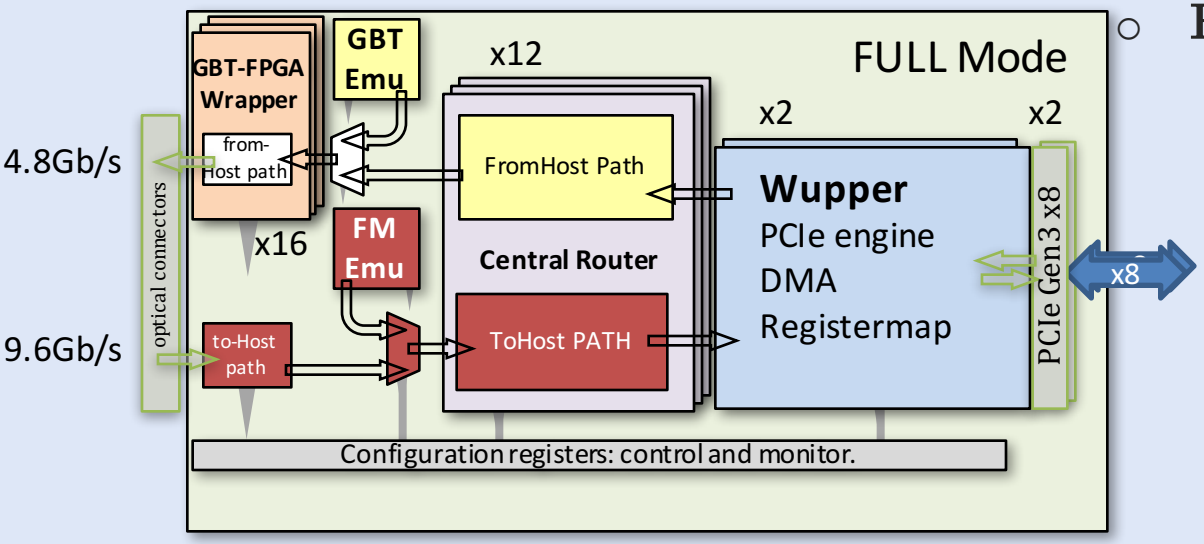


- FPGA: Kintex UltraScale XCKU115
- 8 MiniPODs to support 48 bidirectional optical links
- 16-lane PCIe Gen3 (two 8-lane Endpoints with a switch)
- Flash and Micro-controller to support firmware update
- On-board 0-delay jitter cleaner of Si5345
- Timing mezzanine to interface to the TTC system

FELIX: Modes of Operation

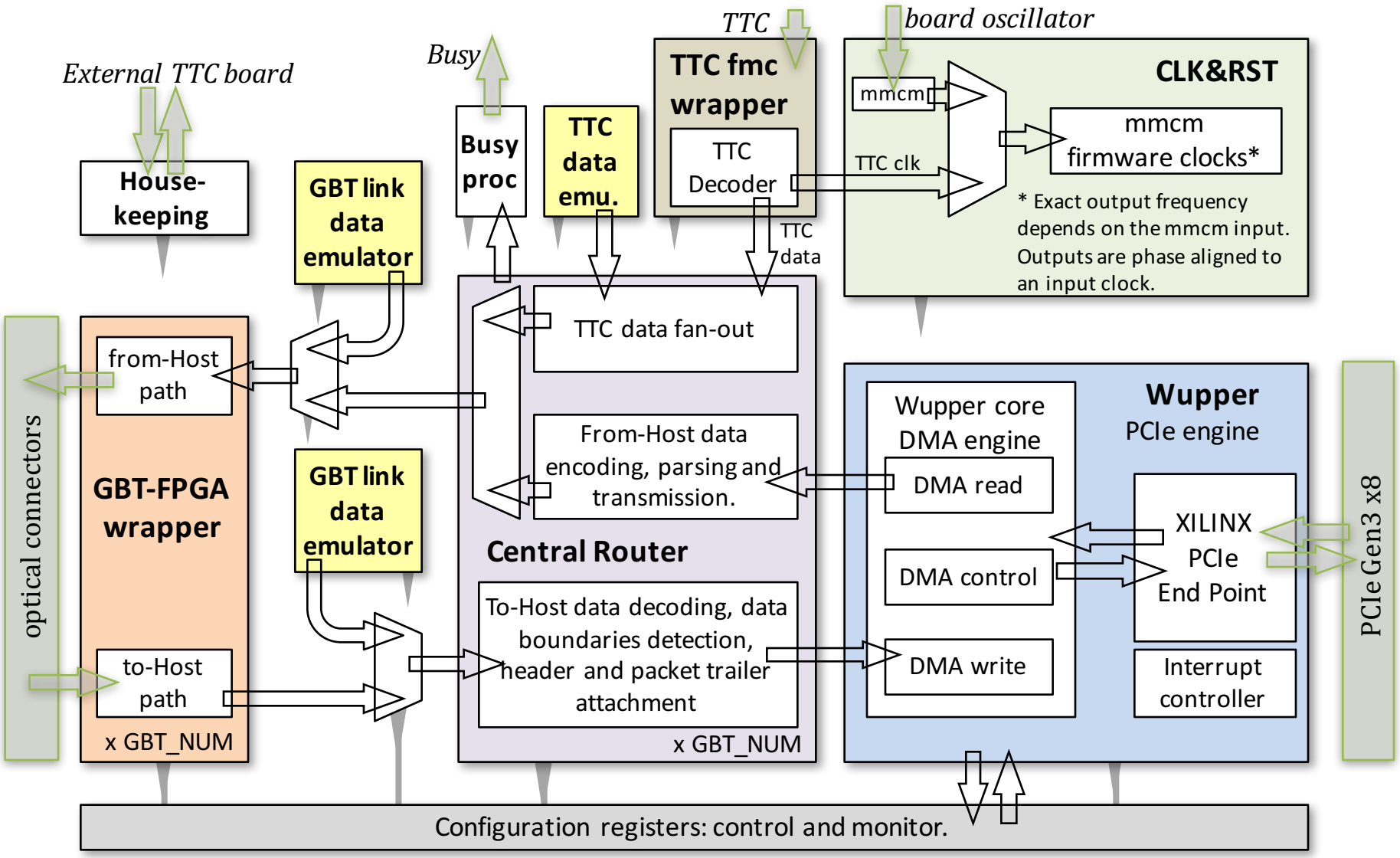


- **GBT mode**
  - Line rate: 4.8 Gb/s
  - Up to 48 bidirectional optical links
  - 3.2 Gb/s payload with FEC or 4.48 Gb/s payload
  - Routes TTC information
  - Optical link divided in E-Links
  - BUSY-ON and OFF
  - Communicate with GBTx, GBT-SCA & GBT-FPGA



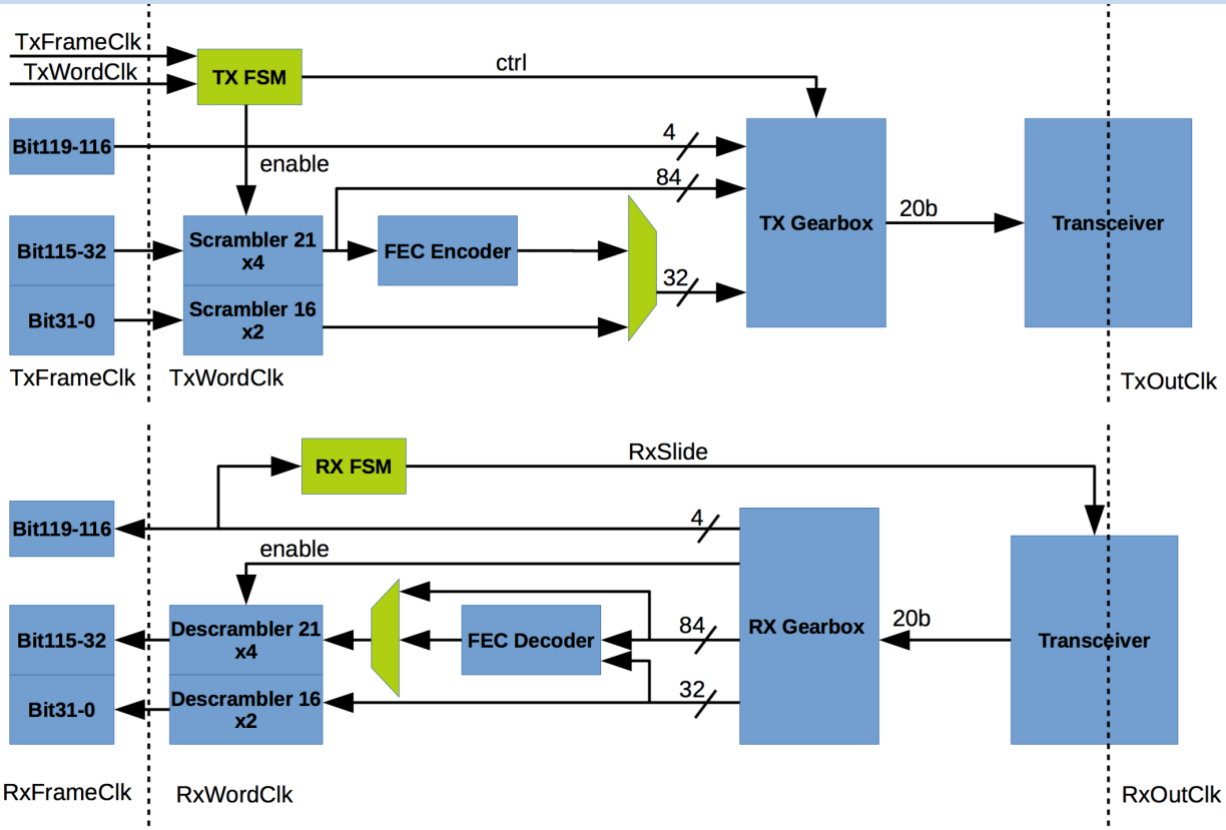
- **FULL mode**
  - Line rate: 9.6 Gb/s
  - 7.68 Gb/s payload (8b10b)
  - Up to 24 bidirectional optical links
  - Routes TTC information
  - CRC
  - BUSY-ON and OFF
  - XON and XOFF
  - 4.8 Gb/s GBT links to FE

FELIX GBT Mode Firmware Block Diagram

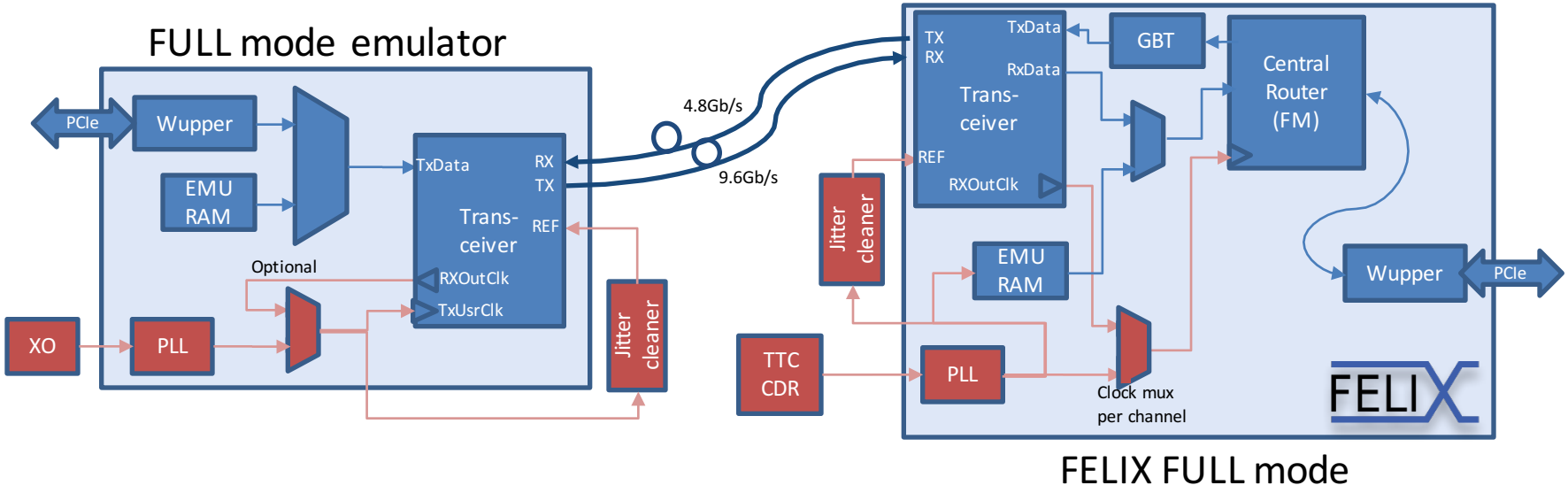


GBT Transceiver Wrapper for FELIX

- FELIX GBT Wrapper is based on CERN GBT-FPGA, with some improvements:
  - Separated GBT firmware from transceiver block.
  - Run-time choice of GBT mode : Normal (FEC) mode or Wide-Bus mode.
  - Lower fixed latency (Tx: 27.8~32 ns; Rx: FEC mode 56.4ns; Rx: Wide mode 43.9 ns).
    - The GBT encoding/decoding are in the 240 MHz domain.
  - https://doi.org/10.1088/1748-0221/12/07/P07011



FELIX FULL Mode Chain and Clocking



- 9.6Gb/s links tested with 32-bit Prbs-31 generator and checker.
  - No error occurred for ~72 hours run. BER < 1E-15.
- Complete design tested with different FPGA based emulated data generators
  - No errors occurred for several TB of data transmitted
- Optional RX clock recovery for TX in emulator
- Clock recovery in FELIX, or local clock for internal emulator

FELIX Software Package

Low-Level Software Tools

Test Software

Production Software

FLX Card Drivers

cmem_rcc, io_rcc, flx

Development and Debug

pepo, fel, …

The flx-tools Suite

FLX Card API, flx-init, flx-config, flx-info, …

The f-tools Suite

fec, fic, fupload, felink, …

E-Link Configuration

elinkconfig

FELIX Core Application

felixcore

FELIX TDAQ Integration

FELIX Monitoring

felix-mon, felix-web

NetIO

netio-cat, …

FELIX Discovery

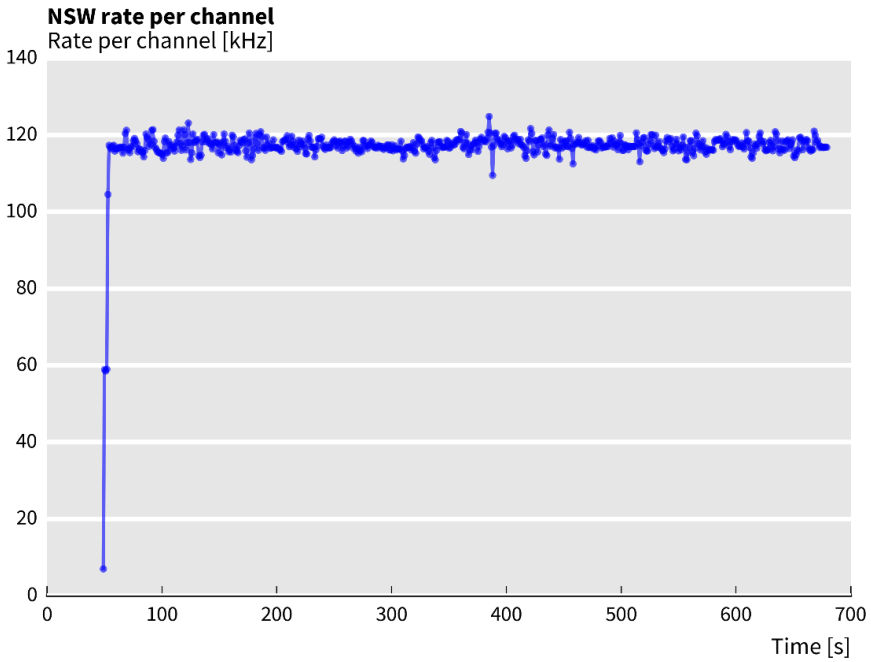
felix-bus

Test clients

felix-client, felix-dcs, fatcat

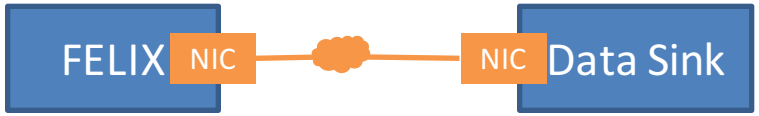
FELIX Core Performance in GBT Mode

- The FELIX Core application is the central process, handling the communication between one or more FELIX Cards on one side and a set of NetIO clients on the other side.
- FELIX Core performance is determined by two critical components
  - 2 threads reading data from the FLX card and copying it ("source threads")
  - Variable number of worker threads processing the data and sending it over the network ("worker threads")
- FELIX core performance with GBT mode was tested with NSW front-ends.
  - NSW is most demanding for GBT mode in Run 3
- FELIX core is capable of full-load at above ATLAS L1 Accept rate of 100 kHz.



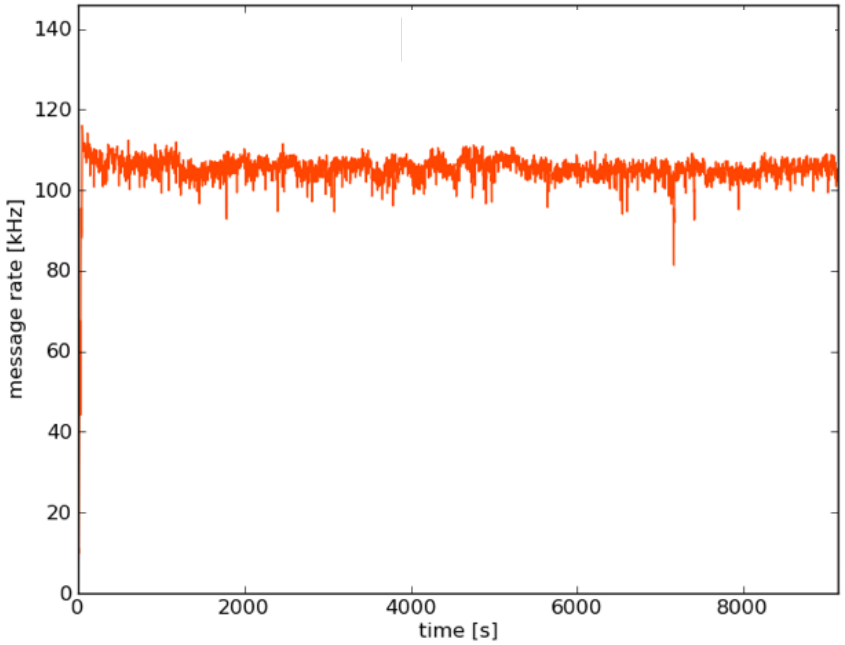
8 E-links per GBT link and 40 Byte chunks

- About 120 kHz rate can be achieved in the worst-case NSW GBT-mode configuration.
- The source threads reading from the cards are using only 10-20% of CPU due to the low data rate.
- The worker threads are the bottleneck of the system.
  - Each packet has to be processed individually!

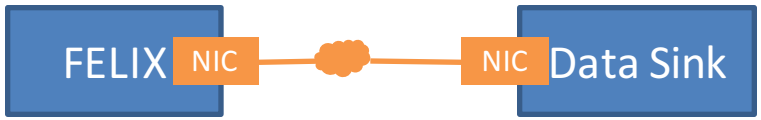


FELIX Core Performance in FULL Mode

- FULL-mode has high bandwidth demands, so the source threads are bottleneck.
  - With 2 threads we can read out around 8 GB/s.
- When using TCP/IP, the worker threads have to copy data into the system TCP buffer. This copy is taking up most of the resources.
- Two ways to improve this:
  - Run more threads (requires more cores than the current 6-core CPU)
  - Use RDMA (Remote Direct Memory Access) instead of TCP/IP, this completely eliminates the copy. RDMA technology is already supported by NetIO and being actively studied by the FELIX team and ATLAS netadmins. Measurements show significantly reduced CPU consumption when using RDMA.



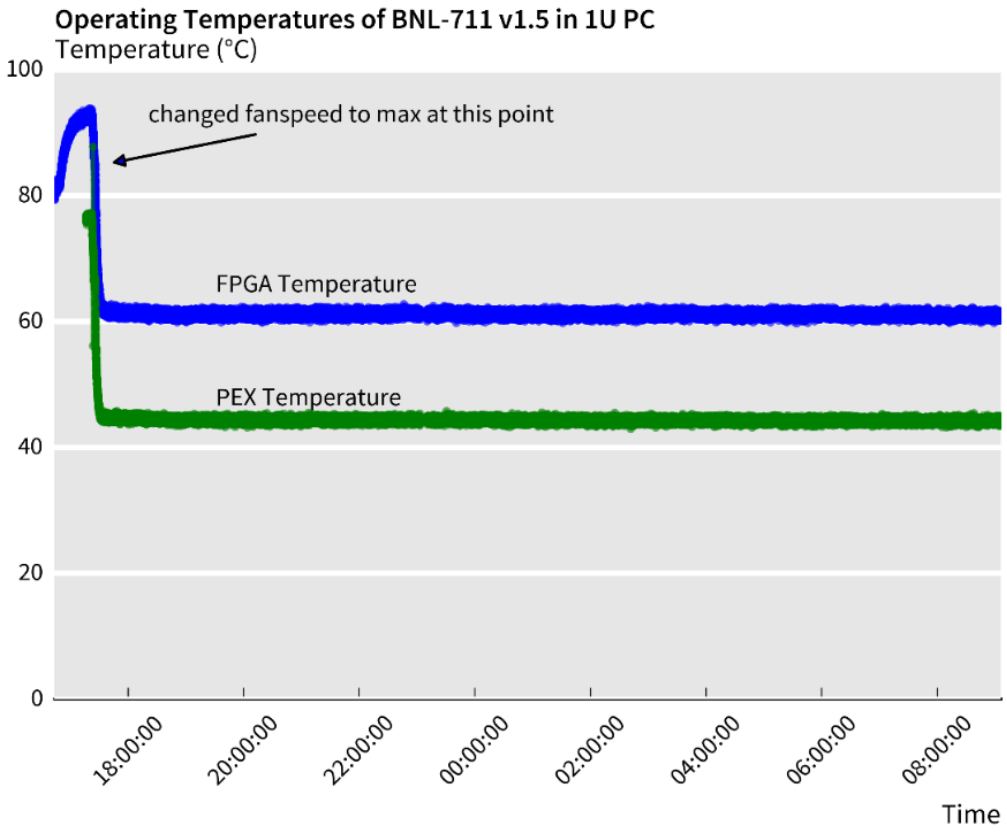
- The L1Calo case is the most demanding workload for FELIX
- Achieved rate is about 105-110 kHz per channel



L1Calo configuration
(package size: 4800 Byte)

Health Monitoring

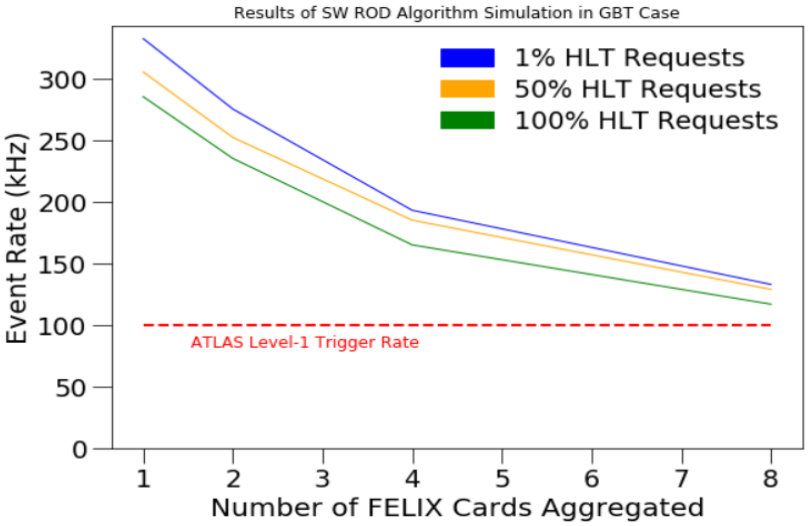
- Power and temperature monitoring are very important.
  - FELIX system will be based on a 1U server and has to host up to two FPGA cards as well as a high-performance network adapter.
  - The ***flx-monitor*** application can monitor both of power consumption and temperature.
  - FPGA, PEX and MiniPODs are all within their thermal tolerance.



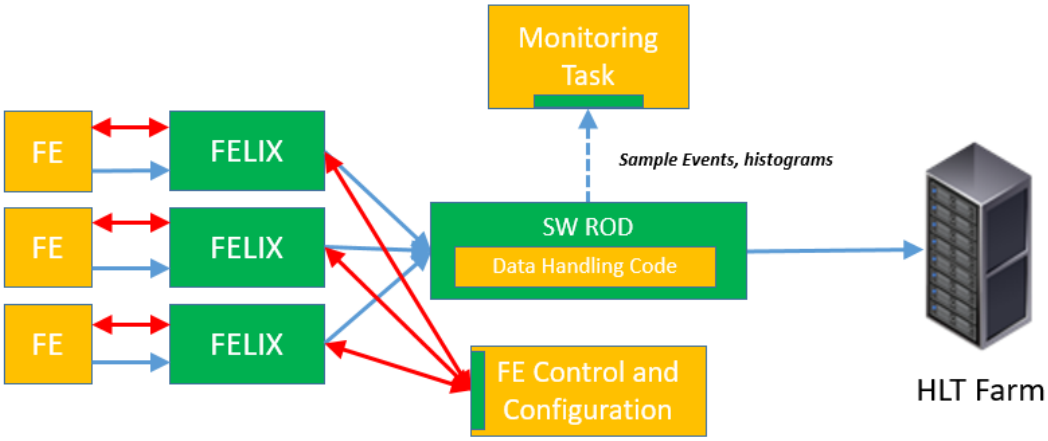
- In the presented measurements all fans of the system are set to full speed.
- The same fan configuration is used in the current ATLAS ROS system. The experience there shows that the fan lifetime is not significantly reduced in this mode.
- Measurements here are done with two FLX cards under full load with the LTDB firmware (this uses all 48 Mini-PODs of a FLX card), as well as with a FULL-mode firmware.

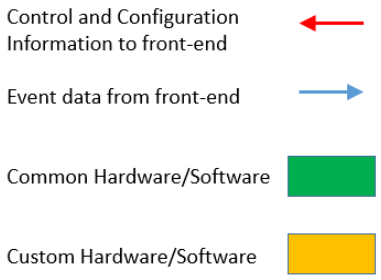
Software ROD Performance

- A Software ROD (ReadOut Driver) is an application running on a commodity server PC which receives data from one or more FELIX systems and performs flexible data aggregation and formatting tasks.
  - Incoming data packets associated with a given ATLAS event are automatically logically aggregated into a larger 'event fragment' for further processing.
  - Data are finally formatted to match common ATLAS specification, as produced by existing readout system, for consumption by High Level Trigger (HLT) on request.



- 'Standalone' runs with simulated input data from multiple FELIX cards, each with 192 E-links and realistic packet sizes
- Data from every E-link aggregated into single fragment at full event rate
- Performing realistic subdetector workloads, reading and modifying the data
- Different scenarios where increasing data fractions are requested by the HLT.



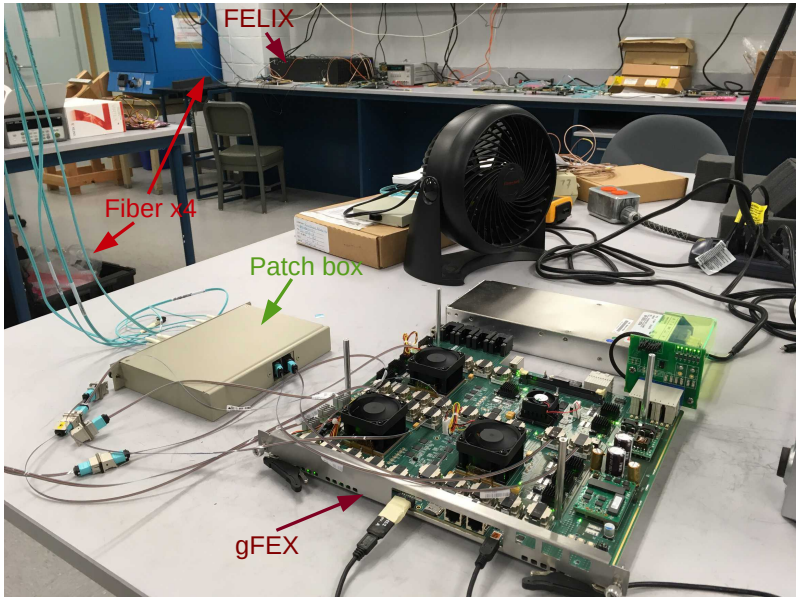


Integration Test with Front-ends

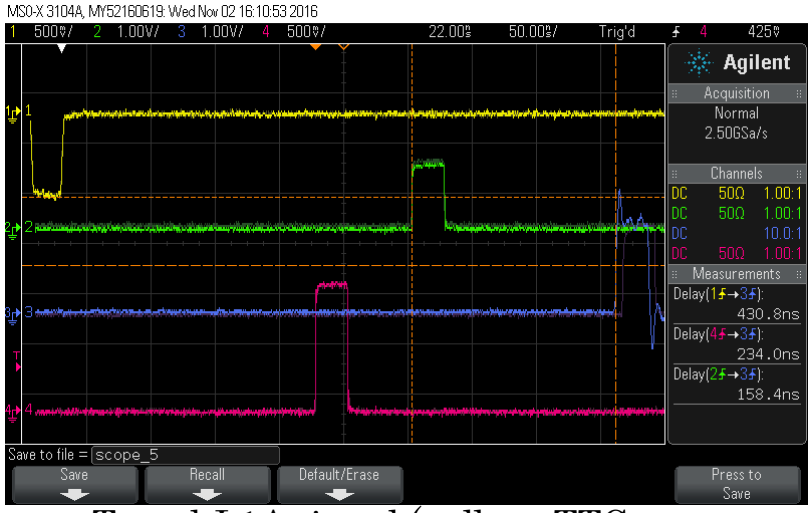
- ATLAS sub-detector test setups
  - Run-3
    - Liquid Argon Calorimeter
      - **LTDB** (LAr Trigger Digitizer Board): 48 channels of TTC distribution and FE control, configuration and monitoring
      - **LDPB** (LAr Digital Processing Blade): FULL mode
    - Level-1 calorimeter trigger
      - **gFEX** (Global Feature Extractor): 12 FULL mode links
      - **ROD**, **Hub** for **eFEX** (Electron Feature Extractor) and **jFEX** (Jet Feature Extractor)
      - **TREX** (Tile Rear Extension) modules
    - Muon spectrometer
      - **New Small Wheels (NSW): sTGC** (Small-strip Thin Gap Chamber) and **MicroMegas** (Micro Mesh Gaseous Structure) detector for muon tracking
      - **BIS78** (Barrel Inner Small MDT (sector 7/8))
  - Run-4
    - Tile Calorimeter
    - ITk Inner Tracker
      - HV-CMOS sensor R&D (CaRIBOu & FE-I4B Telescope)
      - **Pixel demonstrator** readout
      - **Strip demonstrator** readout
- Non-ATLAS detector test setup
  - **ProtoDUNE** collaboration (vertical slice): FULL mode

Integration Test with the gFEX

- gFEX is the global feature extractor in the Level-1 Calorimeter trigger in ATLAS Run 3.
- GBT links from FELIX to gFEX are stable.
  - TTC clock is recovered.
  - GBT link latency is fixed and does not change upon:
    - Transceiver reset
    - Optical link re-connection
    - TTC system power cycling
    - gFEX & FELIX power cycling
- The TTC signals (L1A, BCR, ECR) from TTC system to gFEX through FELIX have been tested.
- FULL mode links from gFEX to FELIX are stable.
  - Prbs-31 data pattern and FULL mode emulator data are tested.
  - BER < 10^{-15}
  - FULL user example from FELIX community is used.

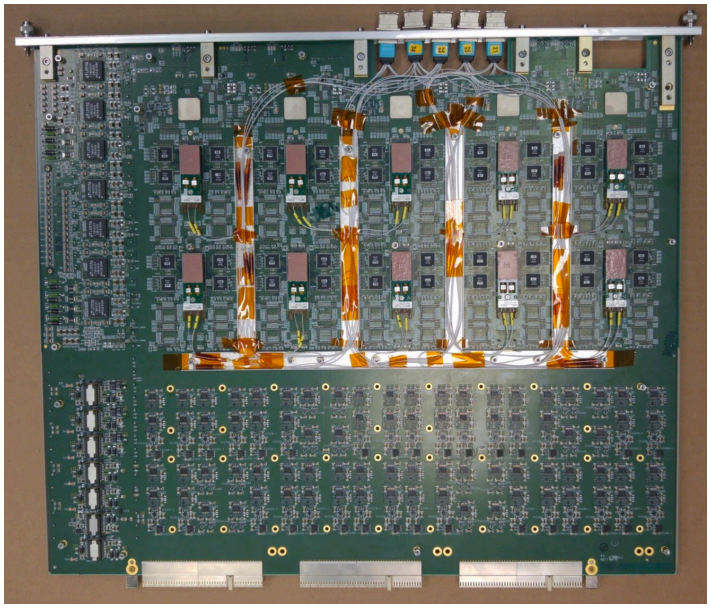


Test setup at BNL

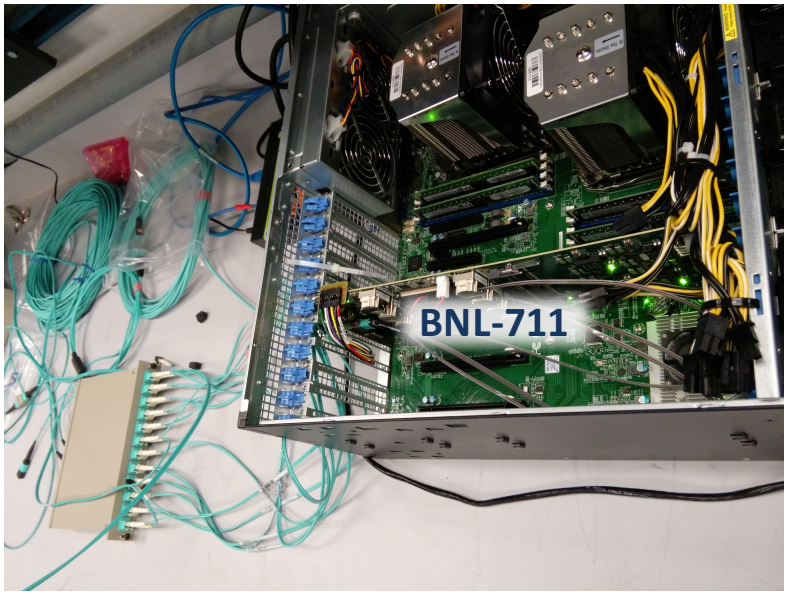


Tested L1A signal (yellow: TTC system, red: TTC decode in FELIX Green: GBT input in FELIX, blue: gFEX)

Integration Test with LTDB Front-ends



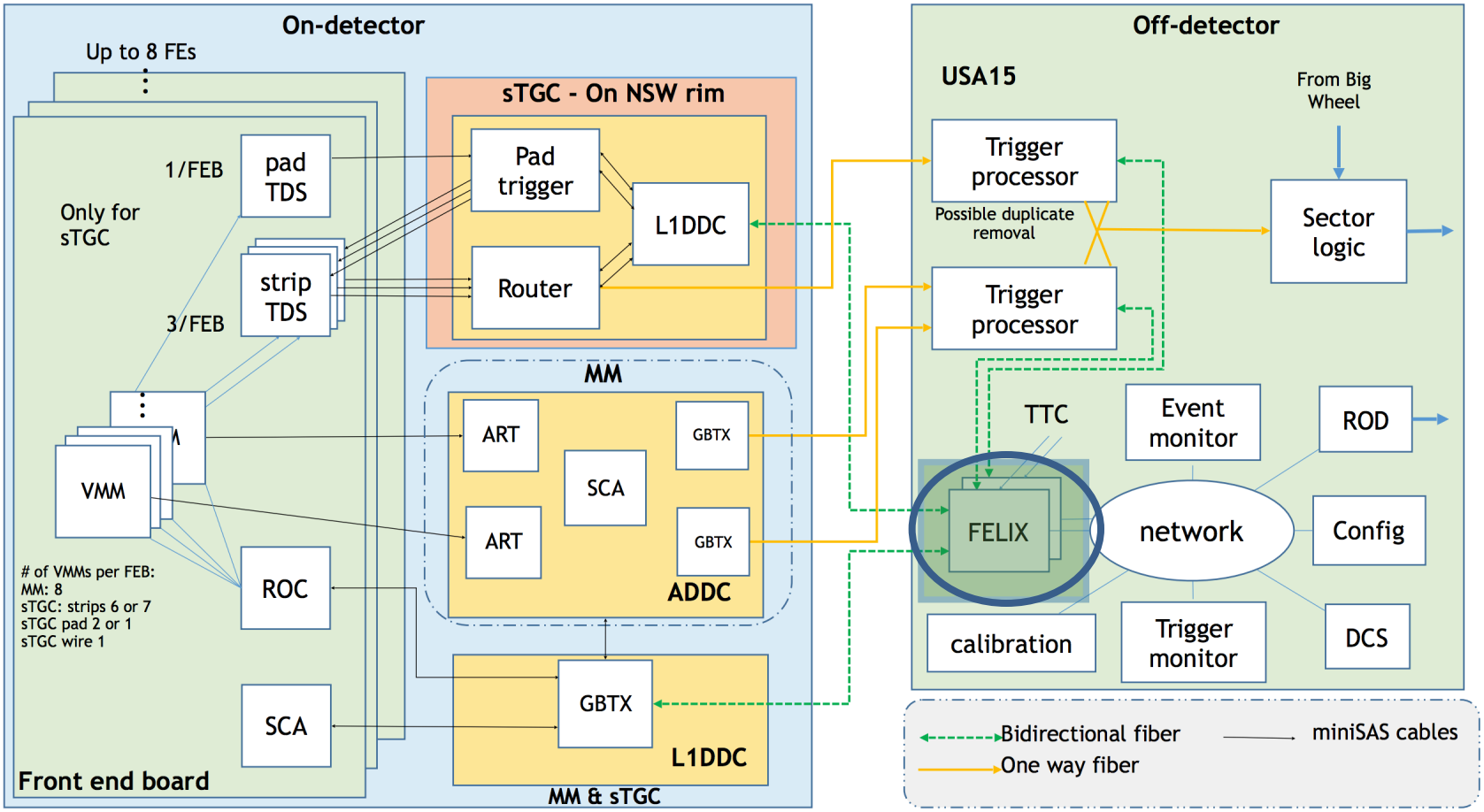
320-ch LTDB pre-production



Back-End for LTDB test

- 320-ch LTDB (Liquid Argon Calorimeter Trigger Digitizer Board) Pre-production
  - With 5 GBTx and 5 GBT-SCA on board
  - FELIX (FLX-711 or FLX-709) system works well.
    - Perform control and monitoring
    - TTC information distribution: clock and BCR (Bunch Counter Reset)
  - The LTDB pre-production was installed in ATLAS pit and had a successful integration test with the FELIX in early 2018.

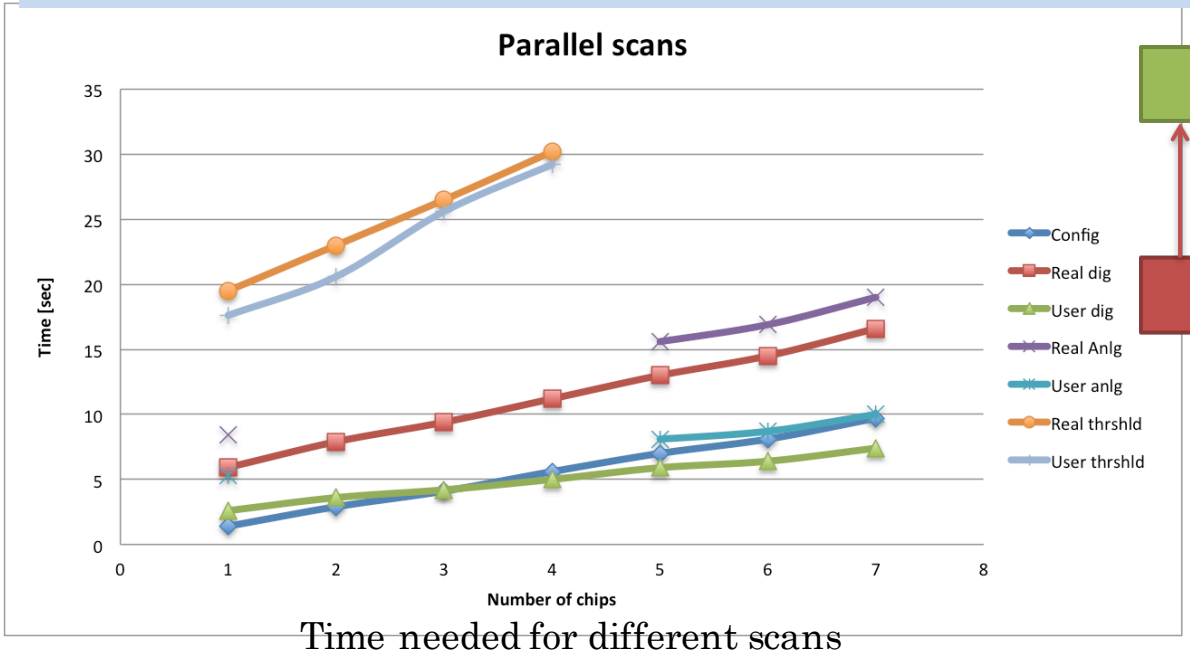
Integration Test with NSW (New Small Wheel)



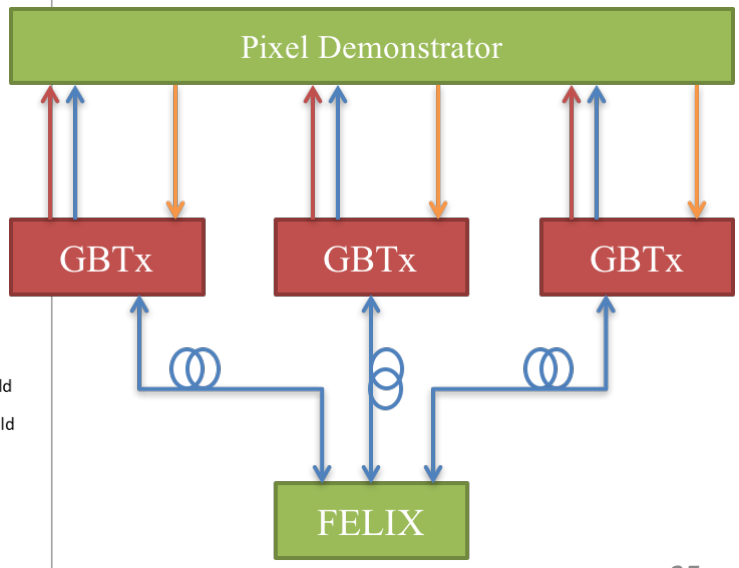
- TTC information can be distributed successfully.
- Data taking over the network was demonstrated, including FELIX full software suite.
- The configuration path is reliable.

Integration Test with Pixel Demonstrator

- Inner Tracker (ITk) will be built to replace the current inner detector in ATLAS HL-LHC upgrade, consisting of Pixel track and Strip track.
- The pixel demonstrator has 7 FE-I4B quad modules (flat section) and 16 double modules (inclined section) on half stave.
- VLDB board is used as the adapter between front-ends and FELIX.
- Quad modules and double modules can be configured and readout correctly.
  - Due to AC-coupling inside quad module, Manchester encoding is implemented in the FELIX firmware.
- FE-I4B calibration has been tested.
  - Multiple FE-I4B chips can be tested in parallel through different Elinks.



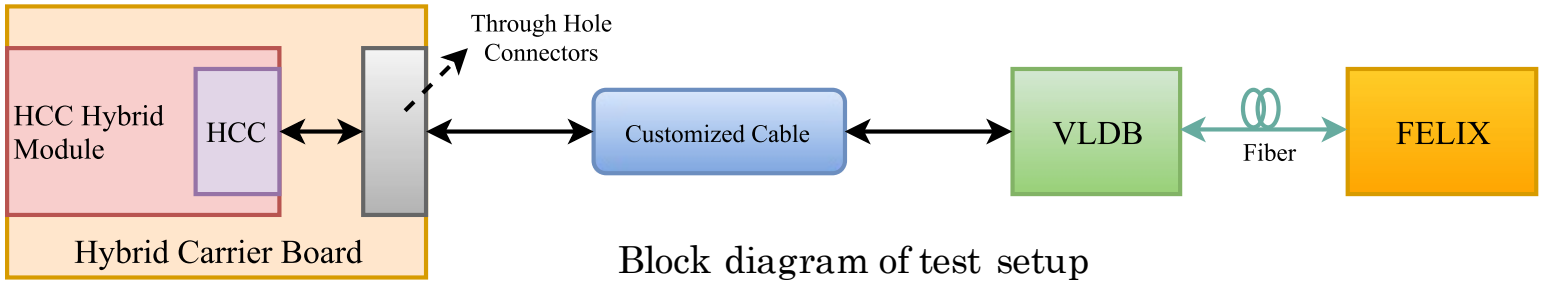
Time needed for different scans



Test setup

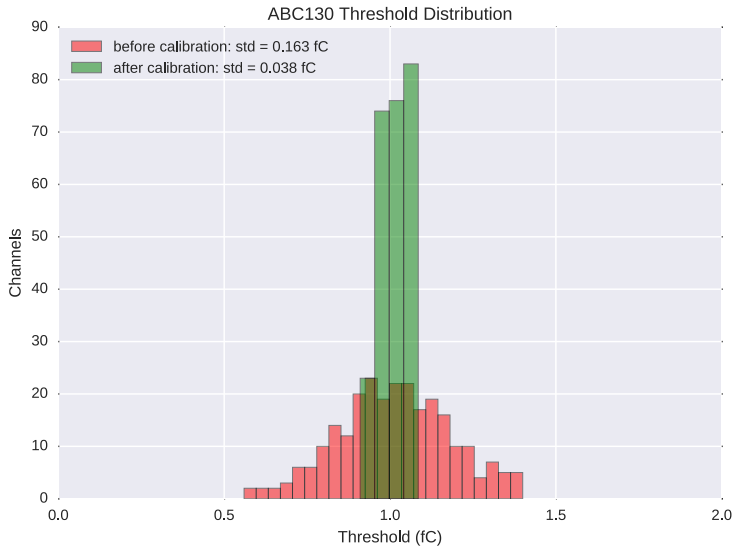
Integration Test with Strip Demonstrator

- The communication between Strip Hybrid modules and the FELIX works well.
  - VLDB board is used as adapter between the hybrid module and the FELIX.
  - Some modifications are made on FELIX firmware to support HCC communication.



Block diagram of test setup

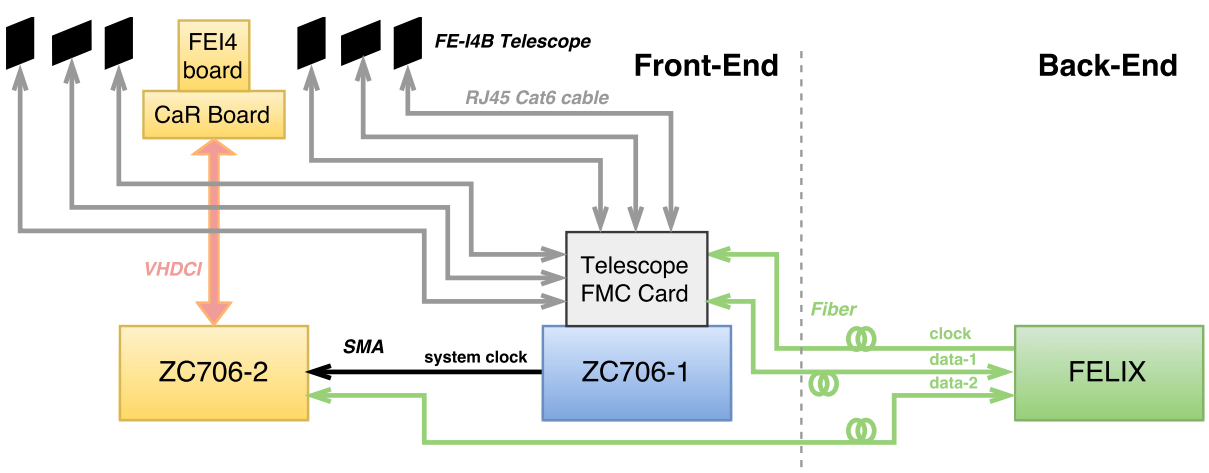
- Some test results:
  - Both of HCC and ABC130 registers can be read and written correctly.
  - 8b10b mode of HCC has been tested.
  - Both of two ABC130 loops have been tested.
  - Both of 160 Mbps and 320 Mbps of HCC output data rate have been tested.
  - Strobe Delay Scan and Three Point Gain
  - Calibration test



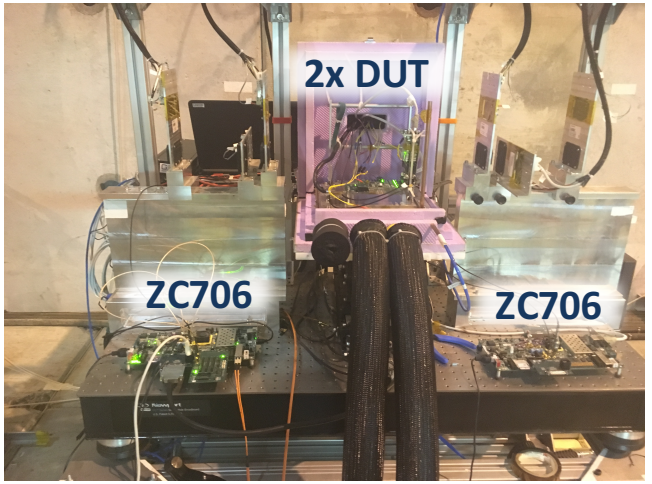
Threshold distribution before and after calibration

FE-I4B Telescope and CaRIBOu Readout

- CaRIBOu is a modular test system for HV-CMOS sensor R&D in the Itk for ATLAS Run 4.
- The FE-I4B Telescope has been built for HV-CMOS sensor characterization in the testbeam , consisting of 6 FE-I4B DC modules.
- The communication between the FELIX and front-ends works well.
  - FELIX software tools are used to send FE-I4B commands and save FE-I4B output data.
  - FE-I4B tuning test
- **The testbeam for the system integration was carried out successfully at CERN in August 2017.**



Block diagram of test setup



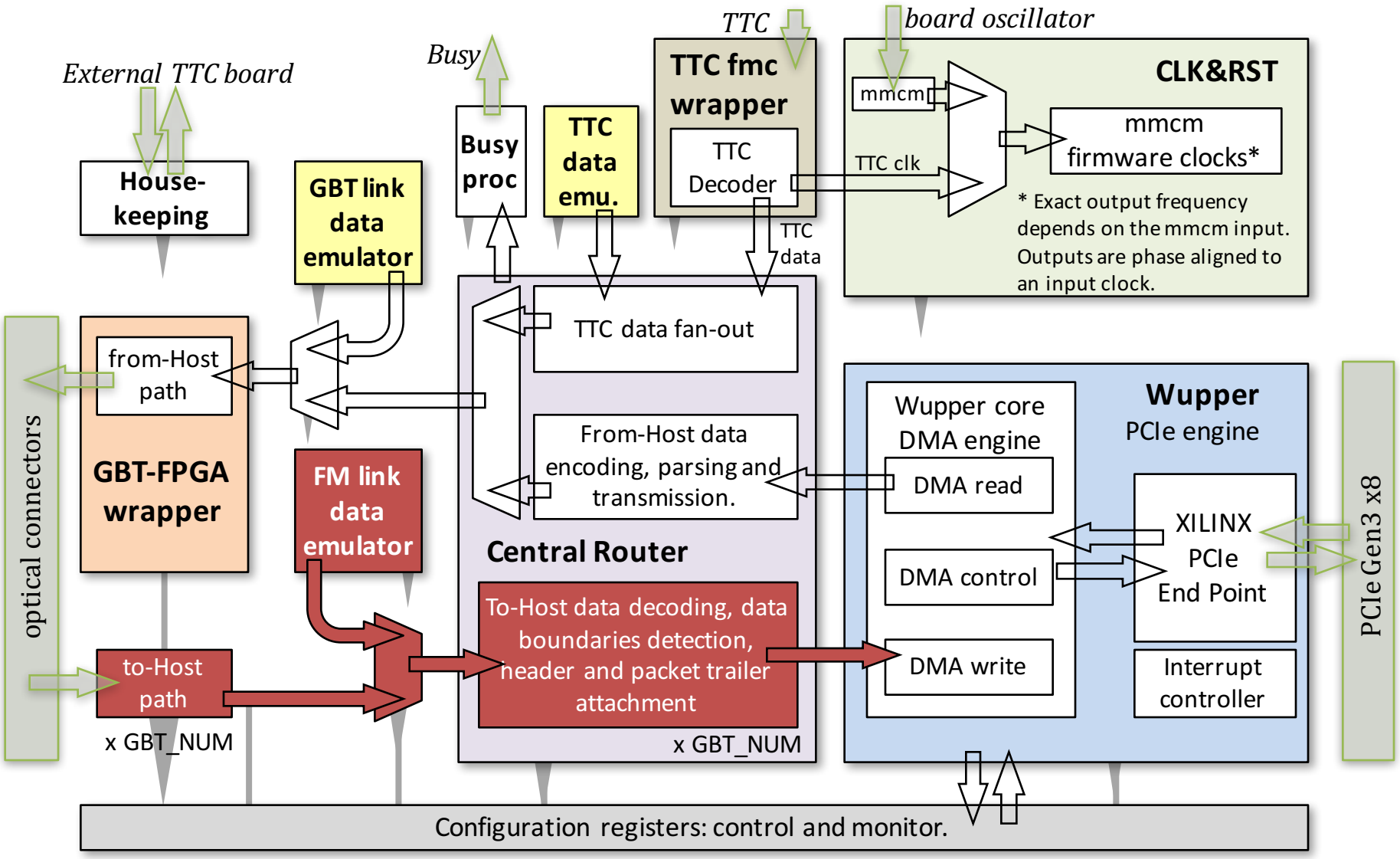
Testbeam at CERN

Summary

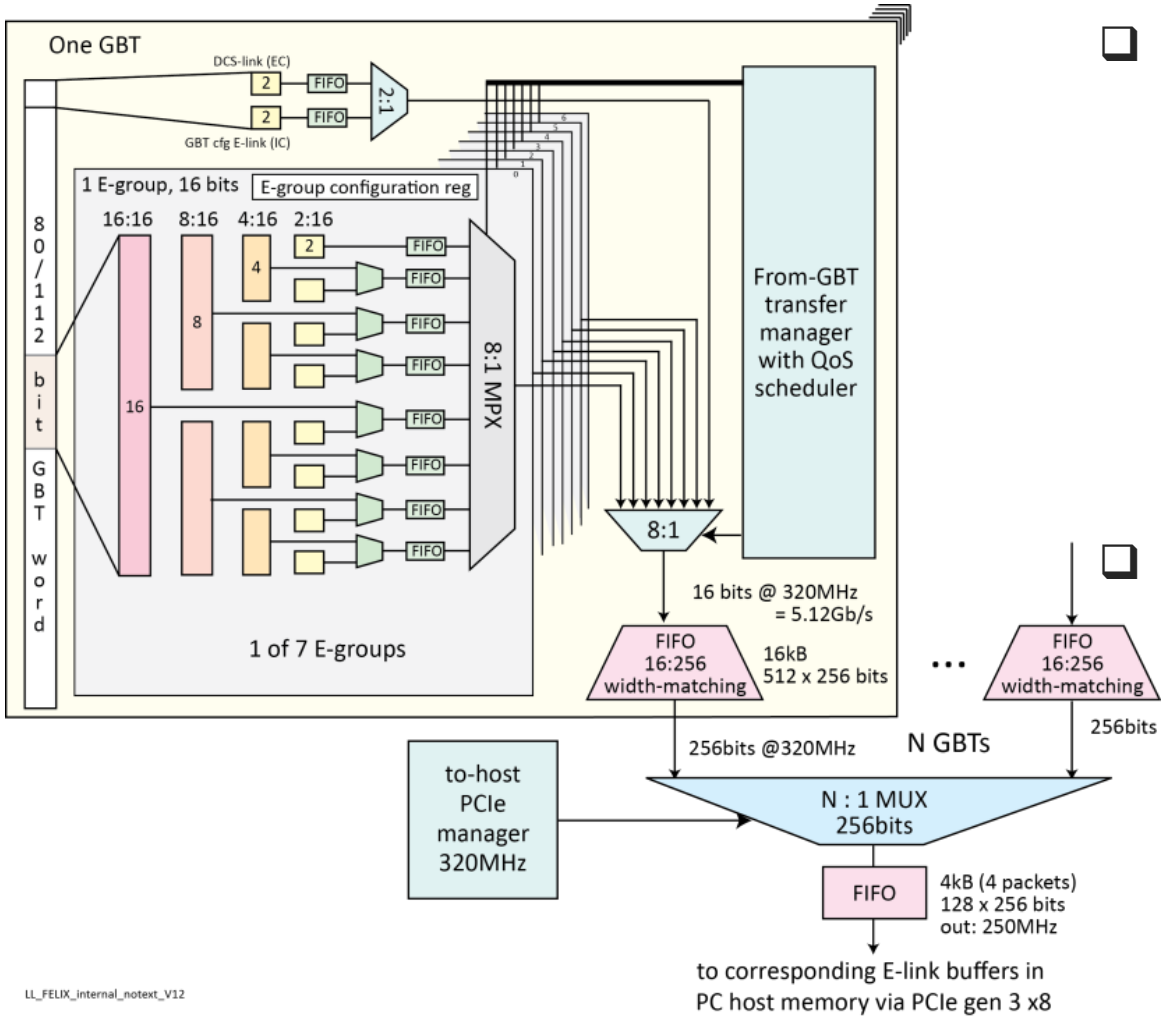
- FELIX is a router between Front-End serial links and a commodity network, which separates data transport from data processing.
- In LHC Run-3 (2021-2023) FELIX will be used by some detectors and trigger systems to interface the data acquisition, detector TTC systems.
- In LHC Run-4 this is planned for all ATLAS detectors.
- FELIX supports GBT mode and FULL mode.
- Status:
  - Reached a development status sufficient to be distributed to the front end developers.
  - Supported hardware platforms: FLX-709 (Xilinx VC-709) and FLX-712 (BNL 16-lane PCIe Gen3 card)
- FELIX has been demonstrated in several integration tests with different sub-detector front-ends.
- Ongoing efforts
  - Increase overall system reliability
  - Final performance benchmarking
- Procurement of Run-3 FELIX in 2018, installation in 2019.

Thank you for your attention

FELIX FULL Mode Firmware Block Diagram



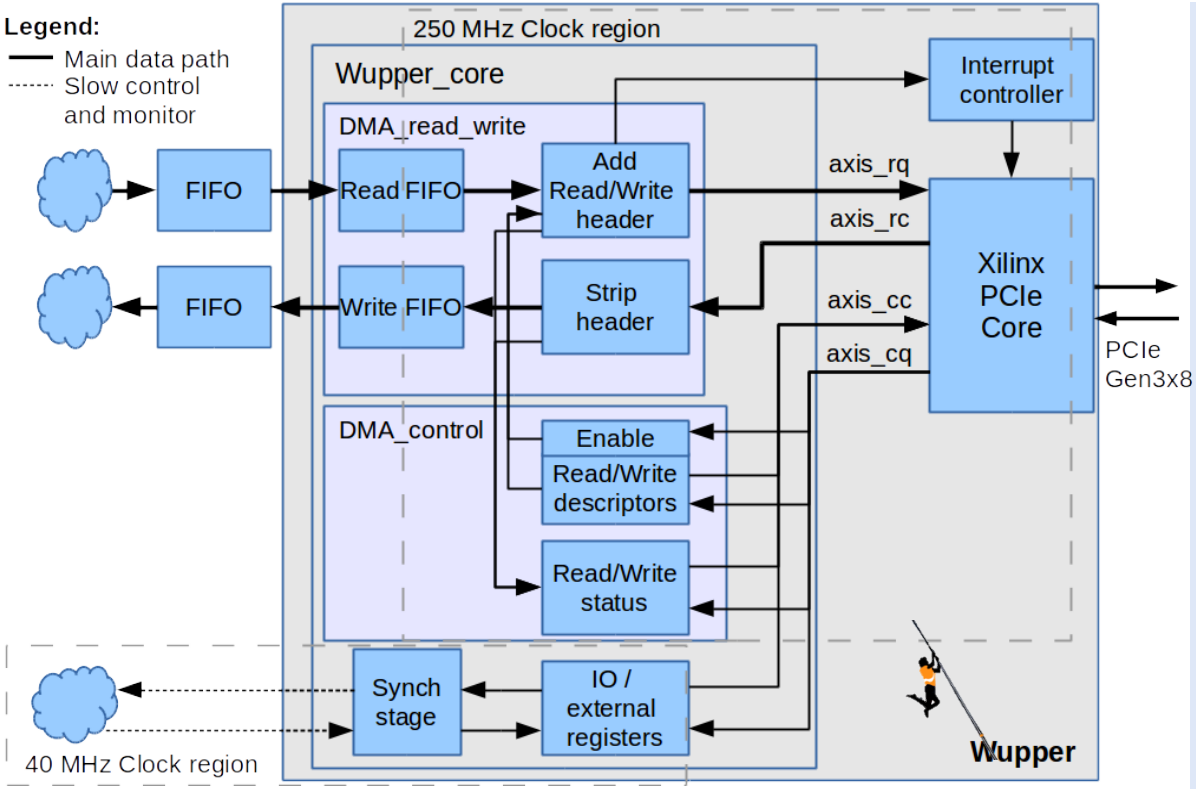
Central Router: Internal Data Multiplexing



- **GBT mode**
  - Handles data streams
  - Routes TTC information
  - Dedicated manager for E-Links in GBT channels
  - Communicate with GBTx & GBT-SCA
  - Main manager toward PCIe Engine
- **FULL mode**
  - Line rate: 9.6 Gb/s
  - Maximum user payload: 7.68 Gb/s: 8B/10B encoding
  - Packets unit: 32-bit
  - Option to include a stream ID for transmitting different logical data streams on same physical link
  - Support for BUSY-ON and BUSY-OFF

Shown: 1 of 7 E-groups of a GBT; 1 of "N" GBTs

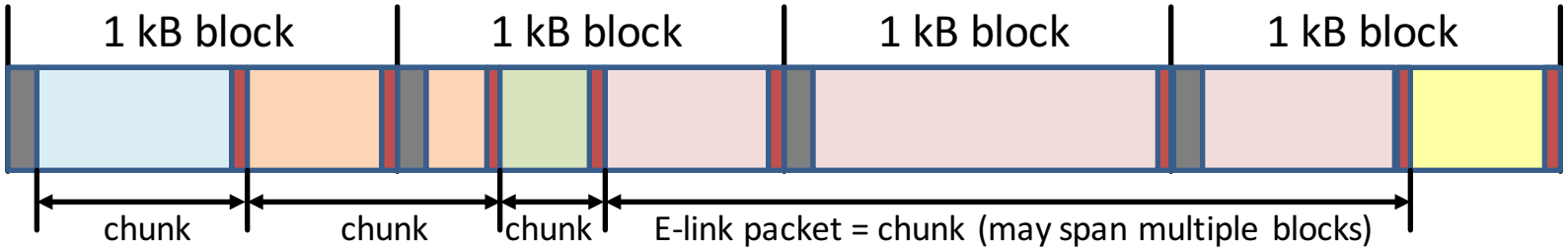
Wupper: PCIe Engine for FELIX



- Developed for use in FELIX
- Published as Open Source (LGPL) on OpenCores http://opencores.org/project,virtex7_pcie_dma
- Core matured to maintenance only phase
- Positive feedback from the community

- PCIe Engine with DMA interface PCIe Gen3 Integrated Block for PCI Express
- For BNL-712: two independent Gen3 x8 lanes EndPoints
- Xilinx AXI (ARM AMBA) Stream Interface (UG761)
- MSI-X compatible interrupt controller
- Applications access the engine via simpe FIFOs
- Register map for programmed I/O synchronized to a lower clock speed

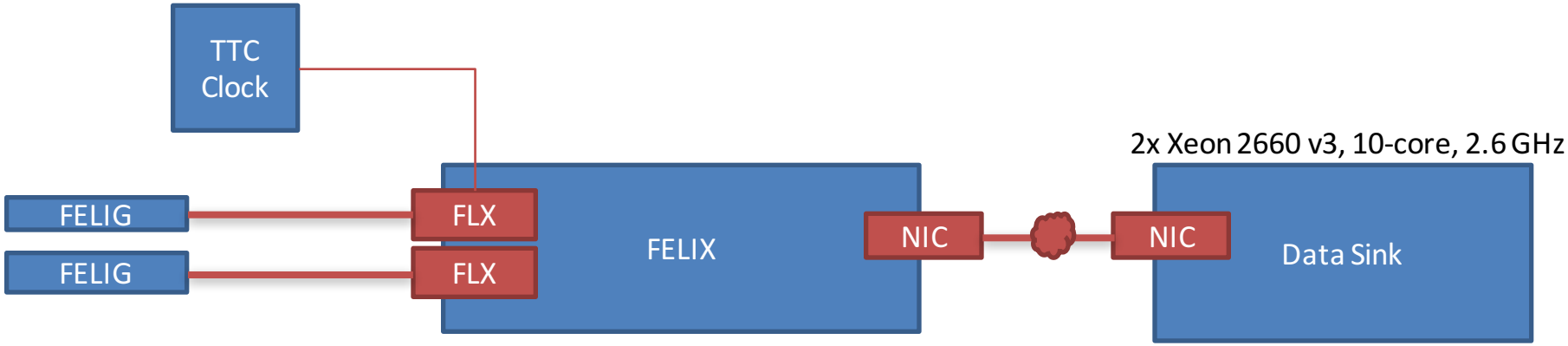
FELIX Data format



- Data buffered in the FPGA per E-link or per FULL mode link and transferred under DMA control
- Fixed block size of 1 kB
- The blocks are transferred into a contiguous area, functioning as a circular buffer, in the main memory of the PC.
- The DMA runs continuously, thereby eliminating DMA setup overheads and achieving high throughput (about 12 GB/s for the 16-lane interface of the FLX-711).
- Event fragments or other types of data arriving via the FE links are referred to as “chunks” and can have an arbitrary size.
- 1 kB blocks of E-links or FULL mode links are multiplexed into a single stream.

- Block header: (32 bits)
  - E-link ID
  - Block sequence
  - Start of block symbol
- Fragment trailer (16 bits)
  - Fragment type
    - First, last, both, middle, null
  - Flags
    - Error, truncation, timeout, CRC error
  - Fragment length
    - 10 bits

Benchmark Setup



All presented benchmarks are based on the FELIX prototype system with an Intel Xeon 1650 v4 (“Broadwell”), 6 physical cores, 3.7 GHz, 32 GB RAM, and Mellanox network adapters. More on PC platform later in this talk.

FELIX Performance Test in GBT mode

- For Phase-I the worst-case scenario is the New Small Wheel.
  - Two types of Front-end systems have to be supported with different requirements:

	NSW sTGC	NSW MM
E-Links per GBT	3	8
Avg. chunk size	38 Byte	22 Byte
Required bandwidth (48 links @100 kHz)	729.6 MB/s	844 MB/s
Chunk rate	19.2 MHz	38.4 MHz

- Note that – different from FULL mode - the required bandwidth is low, but the packet rate is very high.
- The measurement results are for 8 E-links per GBT and 40 Byte chunks (the "hardest" case).
- 48 GBT links in total

FELIX Performance Test in FULL Mode

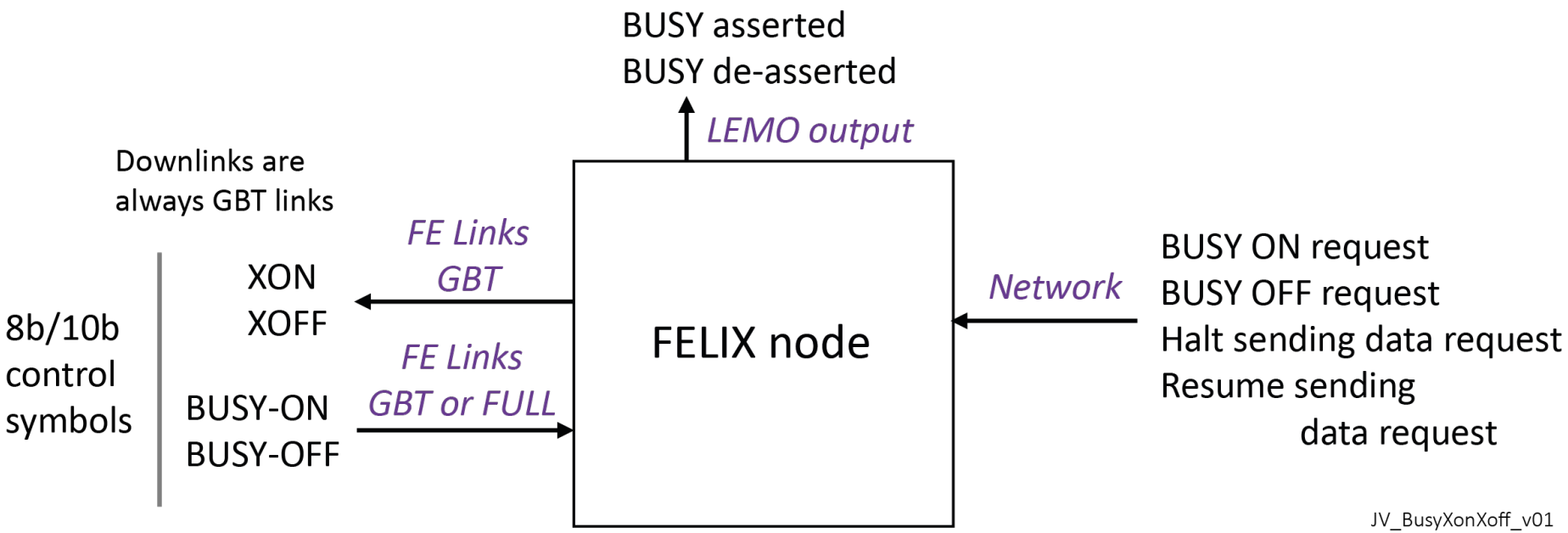
- For Phase-I there are 3 ATLAS sub-systems that use FELIX with FULL-mode

	LAr LDPB	Tile	L1Calo
Max. packet size	3900 Byte	2500 Byte	4800 Byte
Bandwidth (12 links @ 100 kHz)	4.7 GB/s	3 GB/s	5.8 GB/s

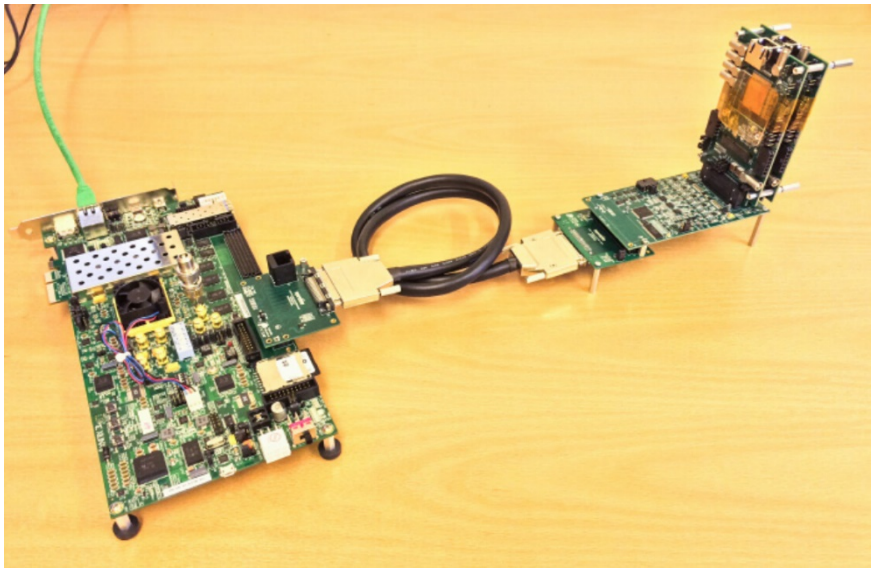
- The FULL-mode FELIX features:
  - 1x FLX-712 with 12 optical links
  - 1x 100 GbE network adapter (dual-port for redundancy)

- The FULL-mode results were obtained using internal emulators of the FLX card. An improved FULL-mode emulator is being worked on. The FULL-mode FELIX is already used in production systems like the ProtoDUNE experiment and LAr

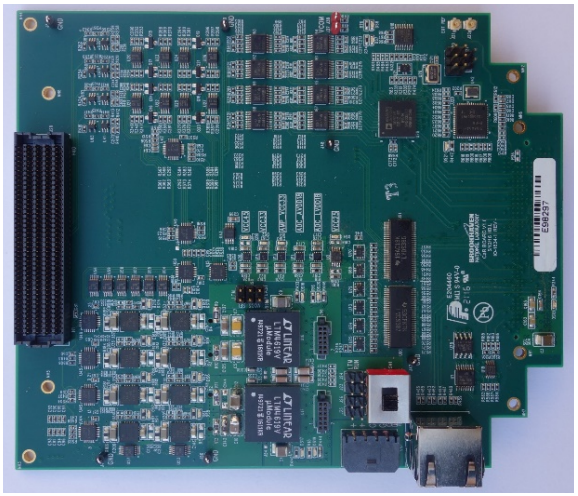
BUSY and XON/XOFF Architecture



Control and Readout Itk Board (CaRIBOu)



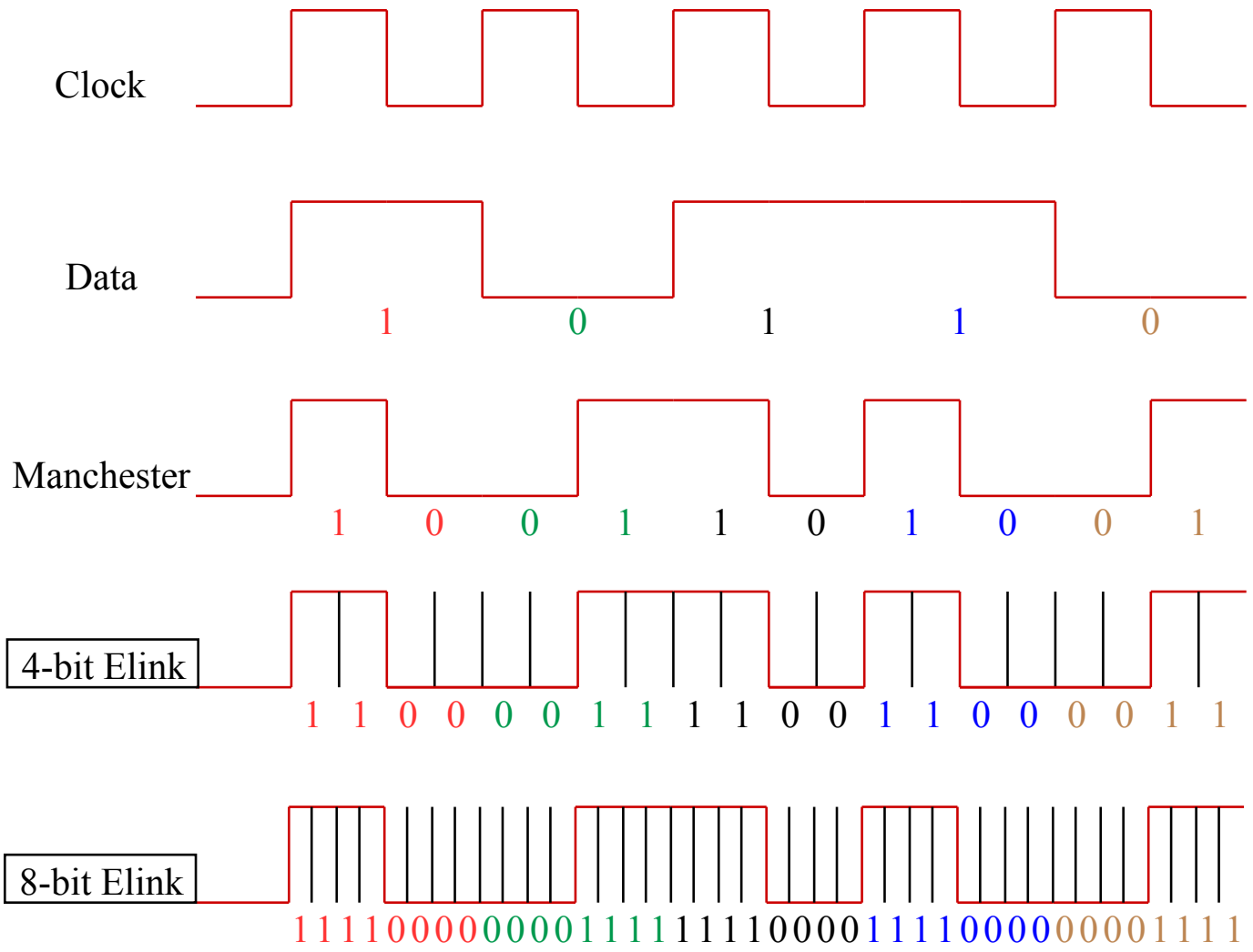
CaRIBOu v0



CaR board V1.0

- CaRIBOu (Control and Readout Itk Board) is a module readout system for HV-CMOS sensor R&D.
  - CaR board:
    - Adjustable power supplies with monitoring for chip board
    - Bias generator for sensor under test
    - Pulse generators for charge injection
    - High-Speed ADC for analog signal sampling
  - DUT board:
    - HV-CMOS sensor and read out chip are mounted on this board
    - Samtec SEARAY right-angle connector is used to connect with the CaR board
    - CCPD, H35DEMO and ATLASPix
  - ZC706 SFP is connected to FELIX via GBT link

Manchester Encoding



FELIX Core Architecture