



FELIX: the New Detector Readout System for the ATLAS Experiment

Gokhan UNEL
UC Irvine

On behalf of the ATLAS TDAQ Collaboration

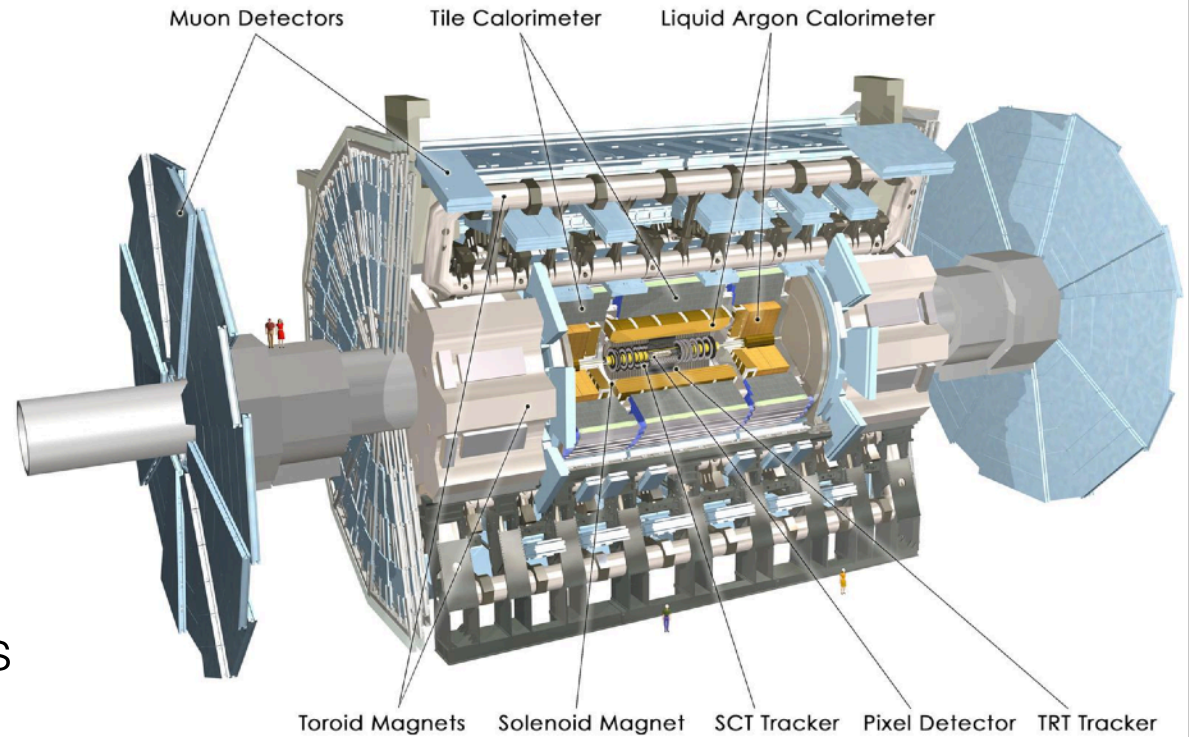


TWEPP 2018 Topical Workshop on Electronics for Particle Physics

Outline

The ATLAS Detector

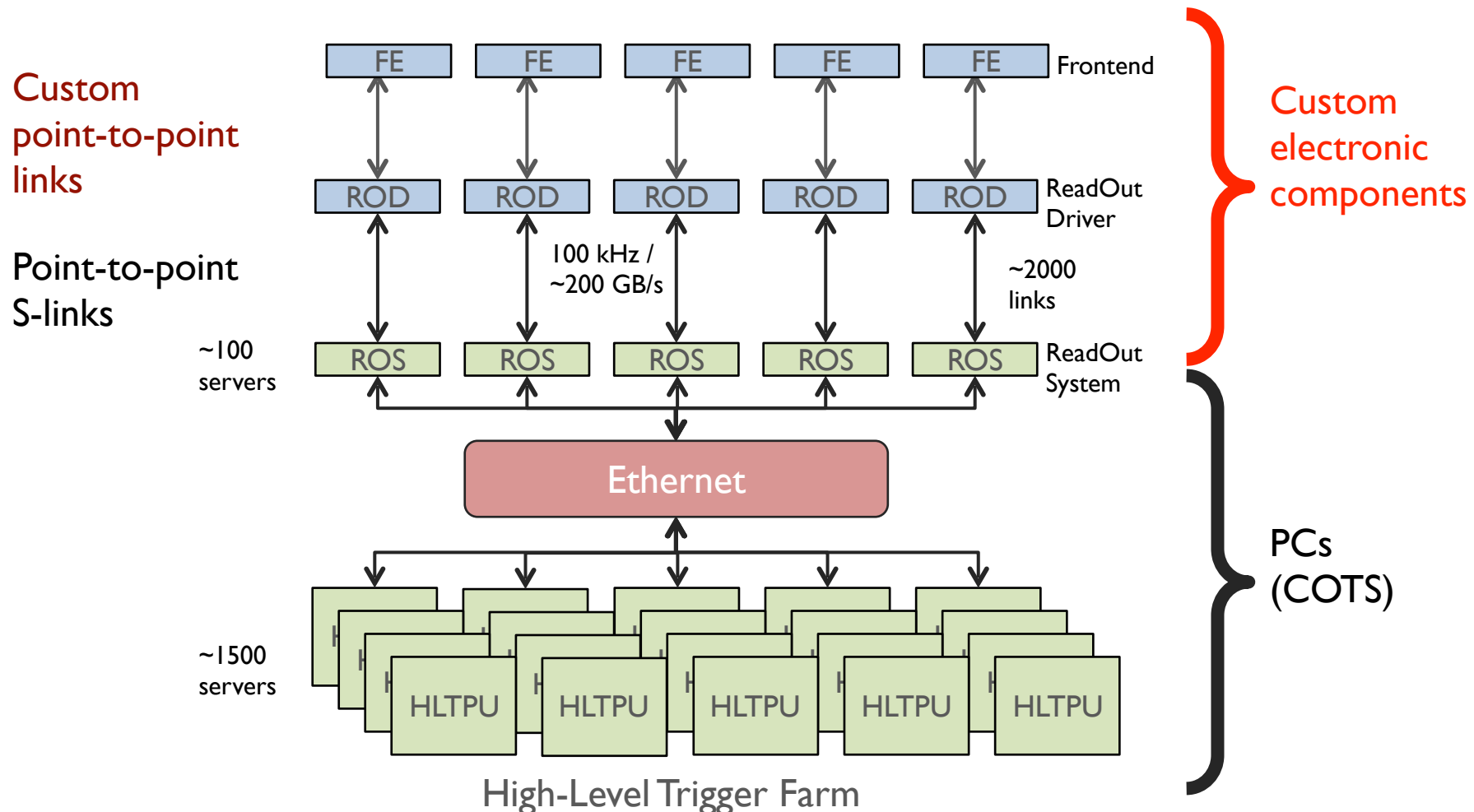
- ATLAS DAQ
 - Today
 - Upgrades
- FELIX
 - Hardware
 - Firmware
 - Software
 - Integration and tests
- Utilization
 - ATLAS Experiment
 - Others



challenge: readout ~100 million electronic channels and sift through data as fast as possible

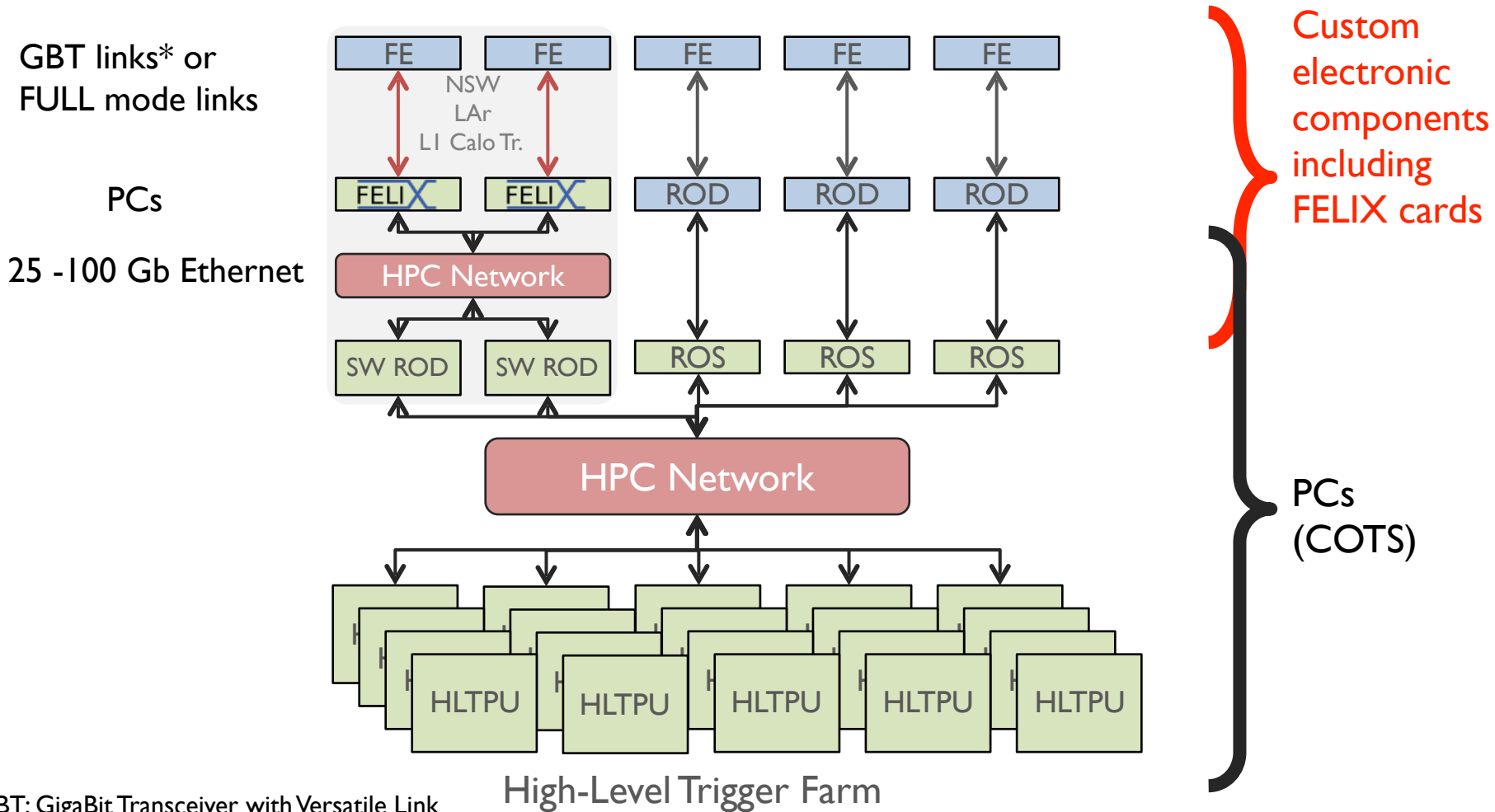
ATLAS DAQ

TODAY: 2MB events, ~50GB/s network bandwidth needed, 1.5GB/s recording data taking finishes by the end of 2018.



Phase-I Upgrade

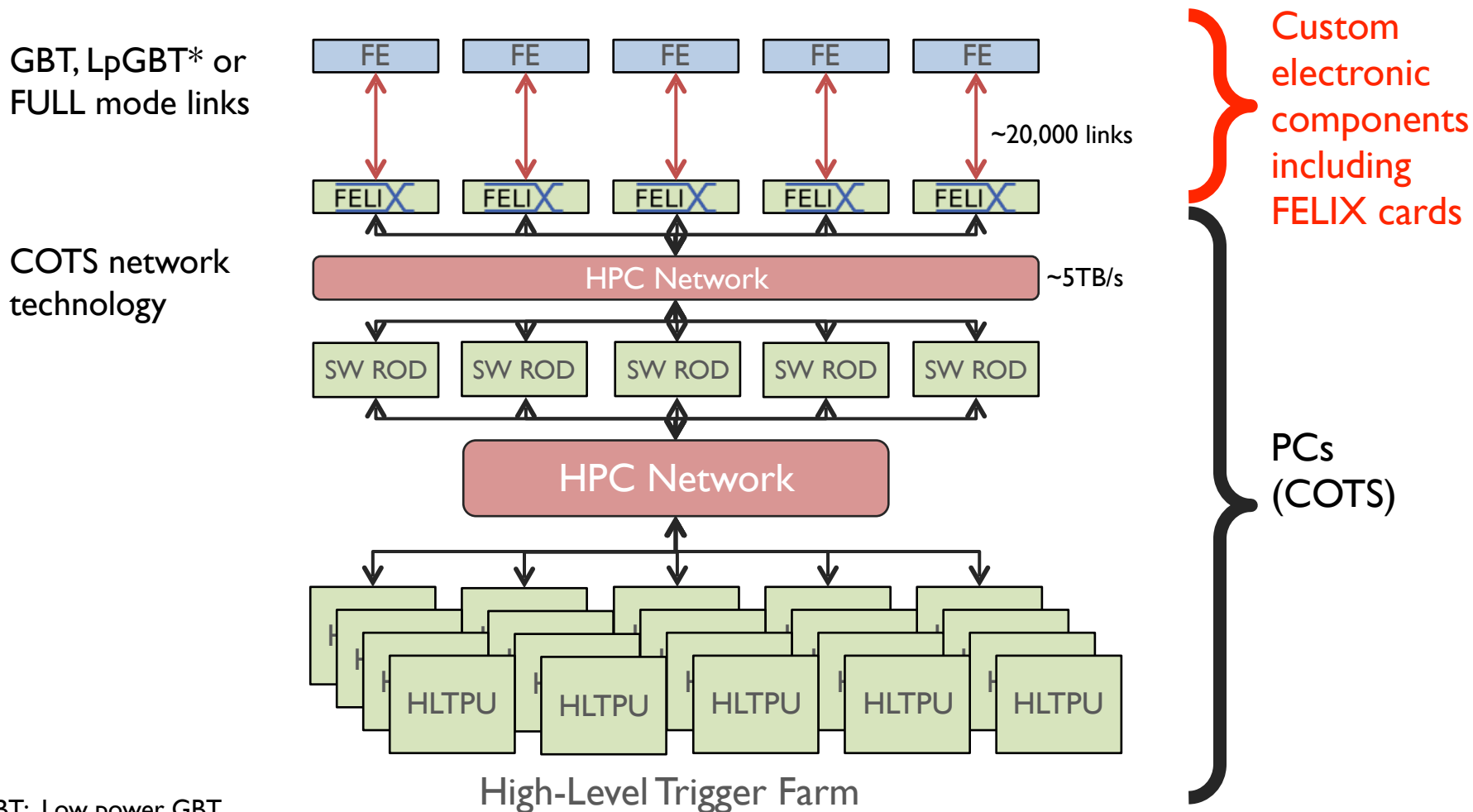
data taking 2021-2023: same requirements as before but will reduce custom components
2MB events, ~50GB/s network bandwidth needed, 1.5GB/s recording



* GBT: GigaBit Transceiver with Versatile Link

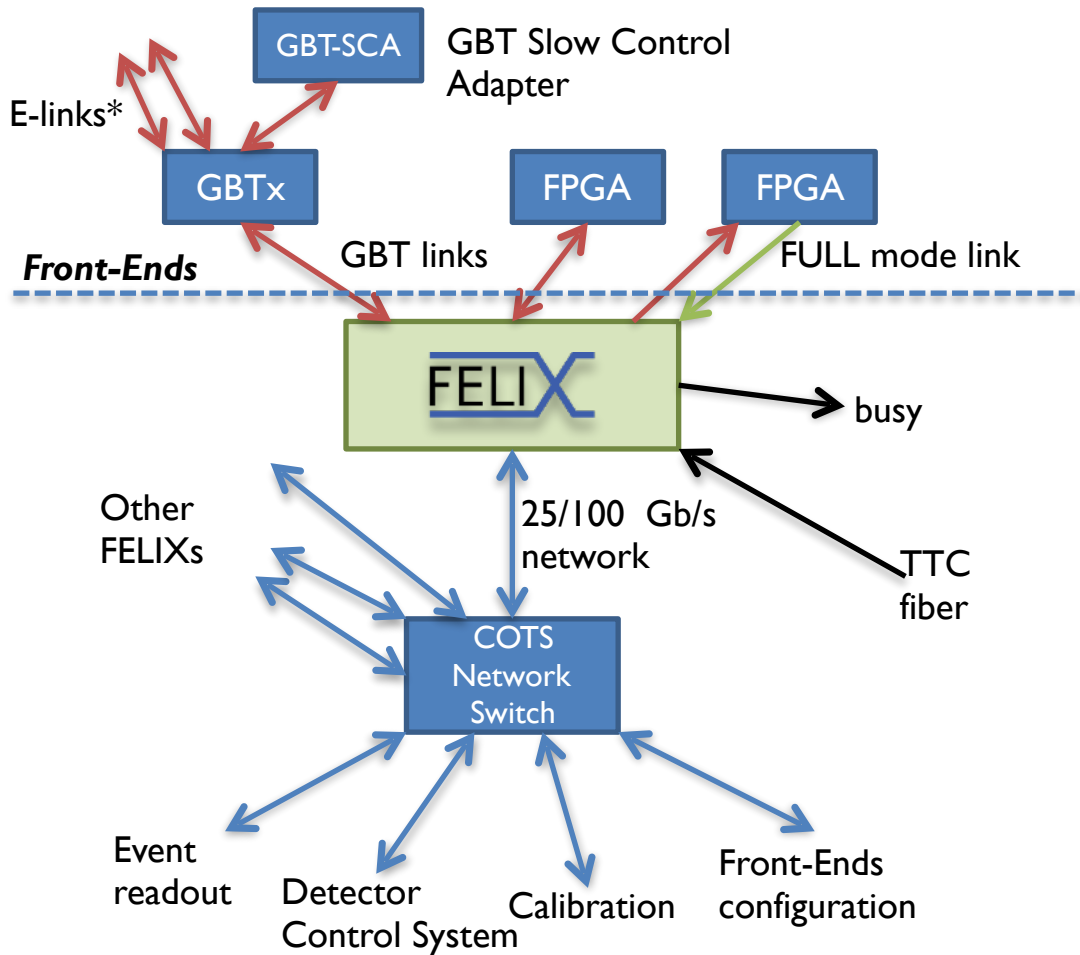
Phase-II Upgrade

data taking 2026-2030: severe requirements, further reduce custom components
 5MB events, ~5TB/s network bandwidth needed, ~50GB/s recording



*LpGBT: Low power GBT

FELIX functionality

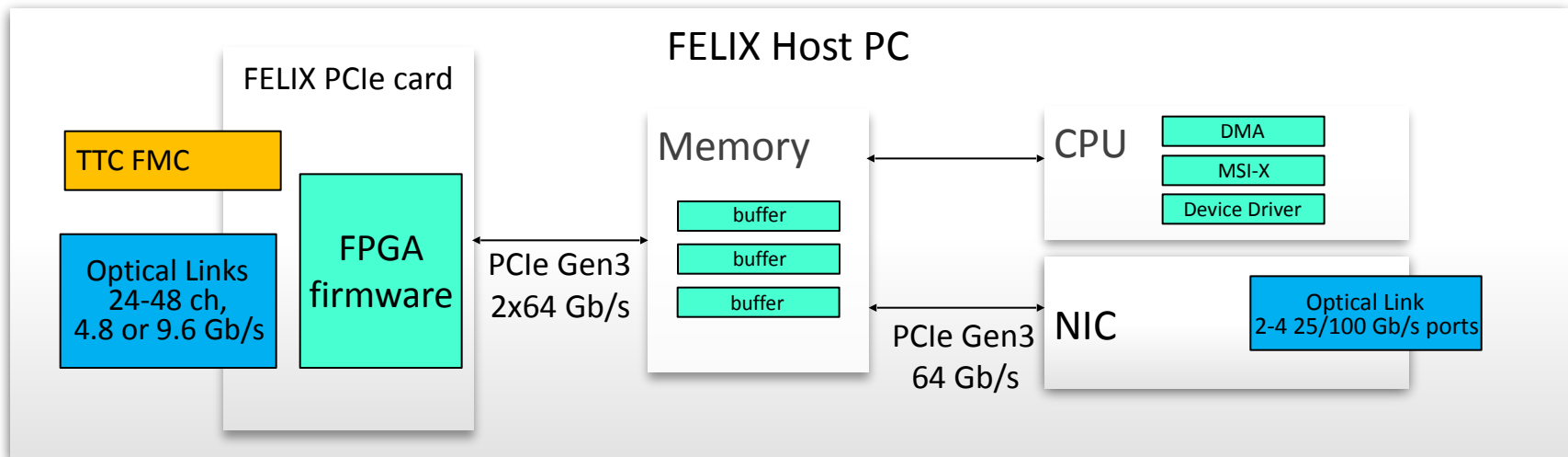


- Scalable architecture
- Routing of event data, detector control, configuration, calibration, monitoring
- Connect the ATLAS detector Front-Ends to the DAQ system, for both the to and from FE directions
- Configurable E-links in GBT Mode
- Detector independent
- TTC (Timing, Trigger and Control) distribution integrated

* **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 hardware design

PCIe card with FPGA chip + Host PC + NIC



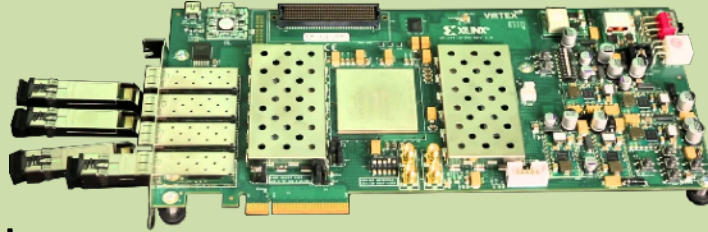
TTC (Time, Trigger and Control) : LHC protocol used to distribute global clock (40.08MHz) and Level 1 trigger information

FELIX hardware components

mini-felix for small systems & prototyping

○ VC-709 from Xilinx

- Virtex7 X690T FPGA
- **FLX-709** or miniFelix
- 4 optical links (SFP+)
- PCIe Gen3 x8



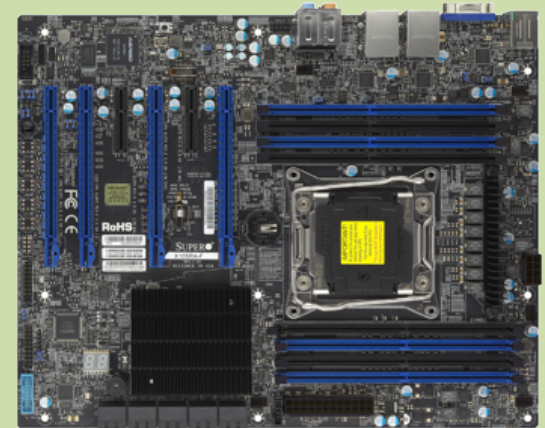
○ TTCfx (v3) mezzanine card

- TTC input
- ADN2814 for TTC clock-data recovery
- Si5345 jitter cleaner



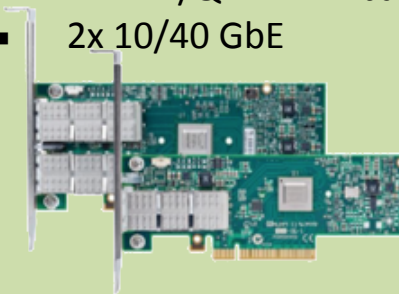
○ SuperMicro X10SRA-F used for prototyping

- Broadwell CPU, e.g. E5-1650V4, 3.6GHz
- PCIe Gen3 slots



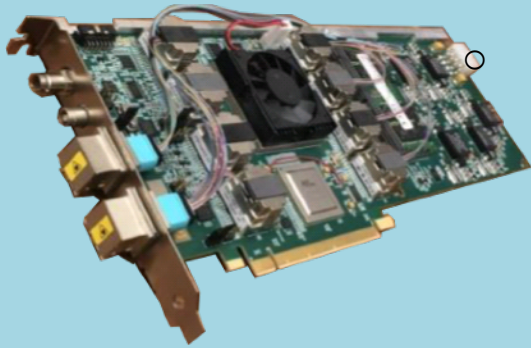
○ Mellanox ConnectX-3

- 2x FDR/QDR Infiniband
- 2x 10/40 GbE



FELIX hardware components

Final Felix system



BNL-712 (or *FLX-712*) from BNL

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

○ MB not yet finalized

- Intel Xeon E5-1660 V4
- \geq two x16 PCIe 3.0
- \geq one x8 PCIe 3.0
- \geq 250GB SSD



TTC



TTC-PON



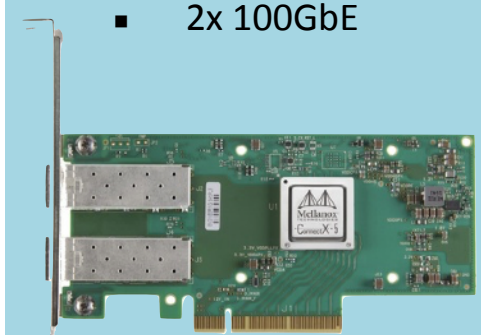
White Rabbit

Timing mezzanine card

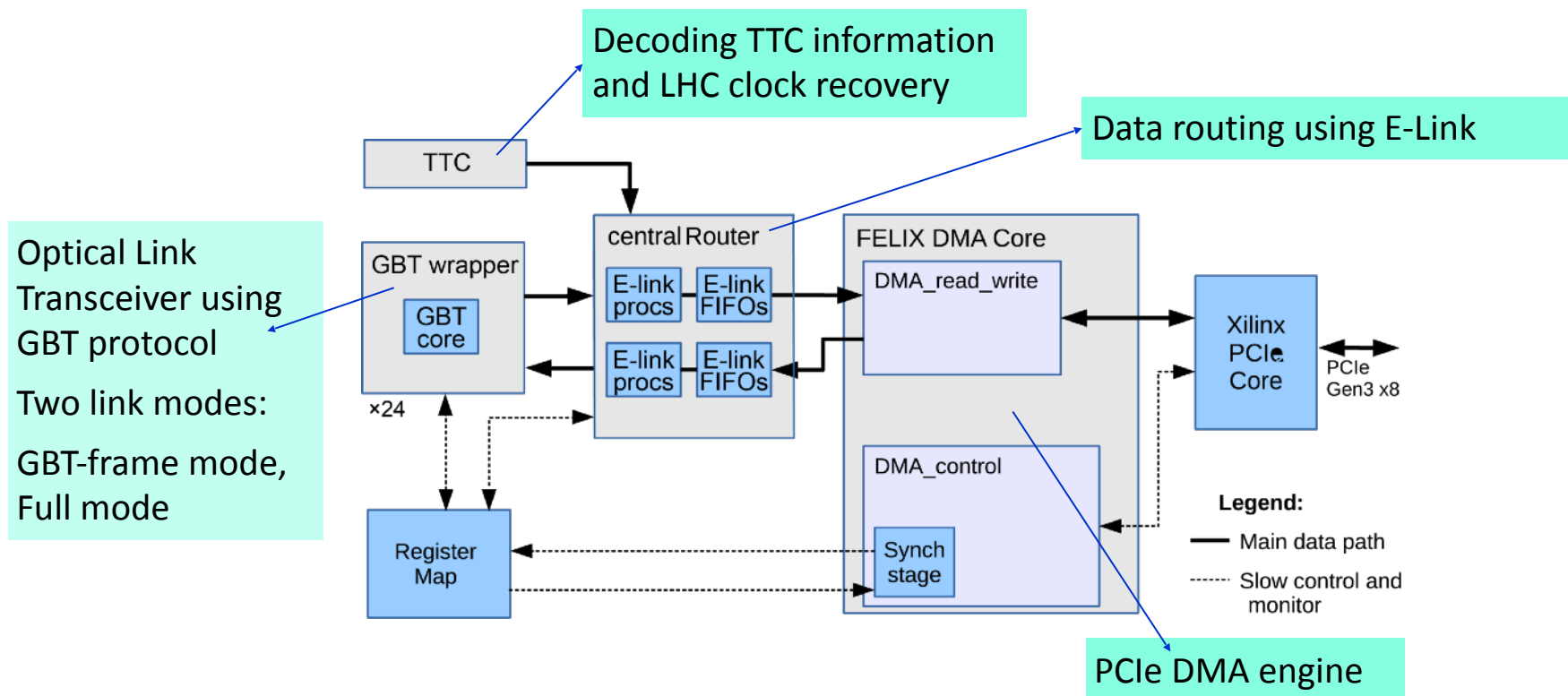
- Supports TTC, TTC-PON, White Rabbit
- TTC configuration is for phase-I

○ Mellanox ConnectX-5

- 2x 100GbE



FELIX firmware design



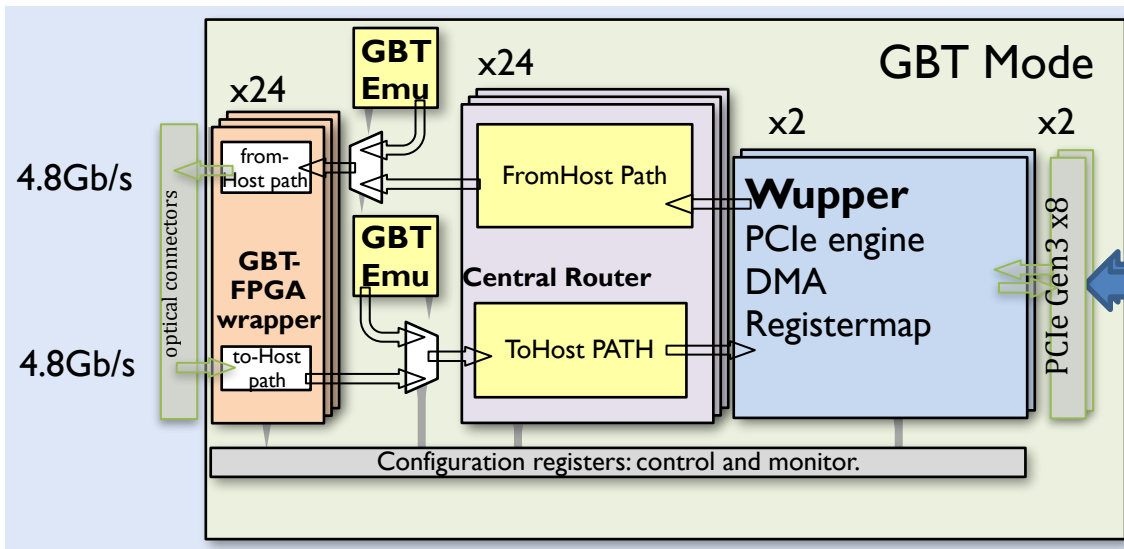
E-Link : data multiplexing protocol designed for ATLAS

Available up to 24 GBT or FULL mode links

Fixed latency transmission to FE using LHC clock

Maximum PCIe throughput to the host PC (2x64 Gb/s)

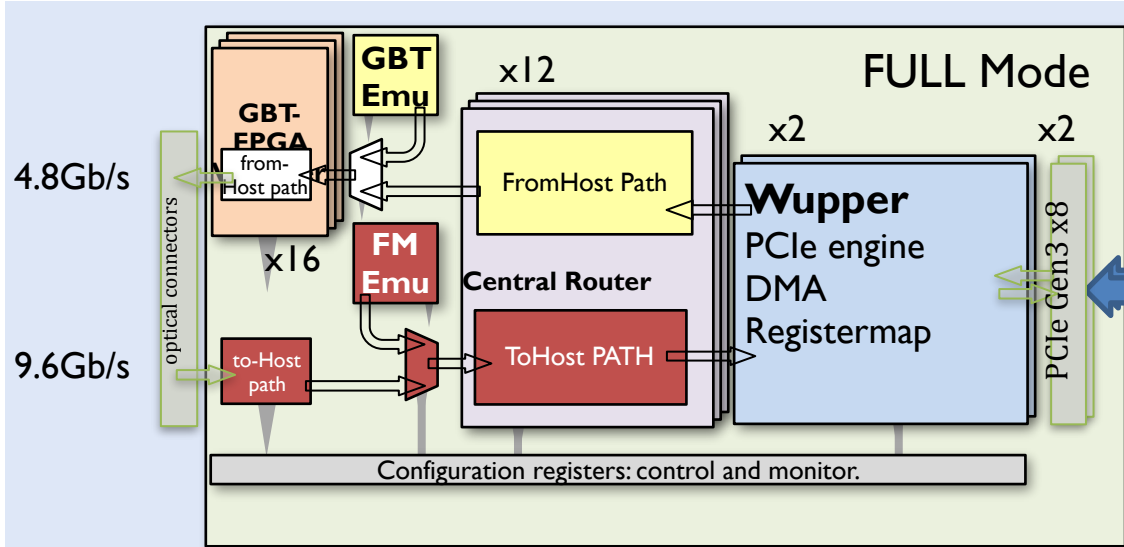
FELIX modes of operation



GBT mode



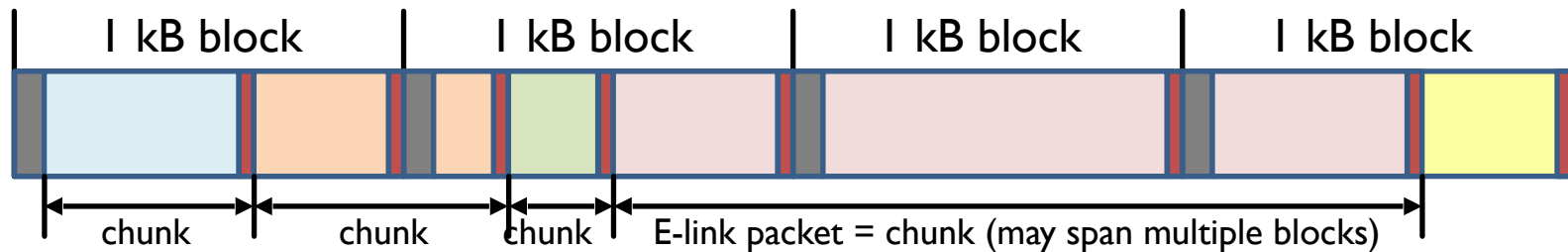
- Line rate: 4.8 Gb/s
- Up to 24 bidirectional optical links
- 3.2 Gb/s payload with FEC or 4.48 Gb/s payload
- Routes TTC information
- Optical link divided in E-Links
- Communicate with GBTx & GBT-SCA



FULL mode

- Line rate: 9.6 Gb/s
- Up to 12 bidirectional optical links OR up to 24 if not fully used
- Routes TTC information
- 7.68 Gb/s payload: 8B/10B encoding
- CRC
- BUSY-ON and OFF
- 4.8 Gb/s GBT links to FE

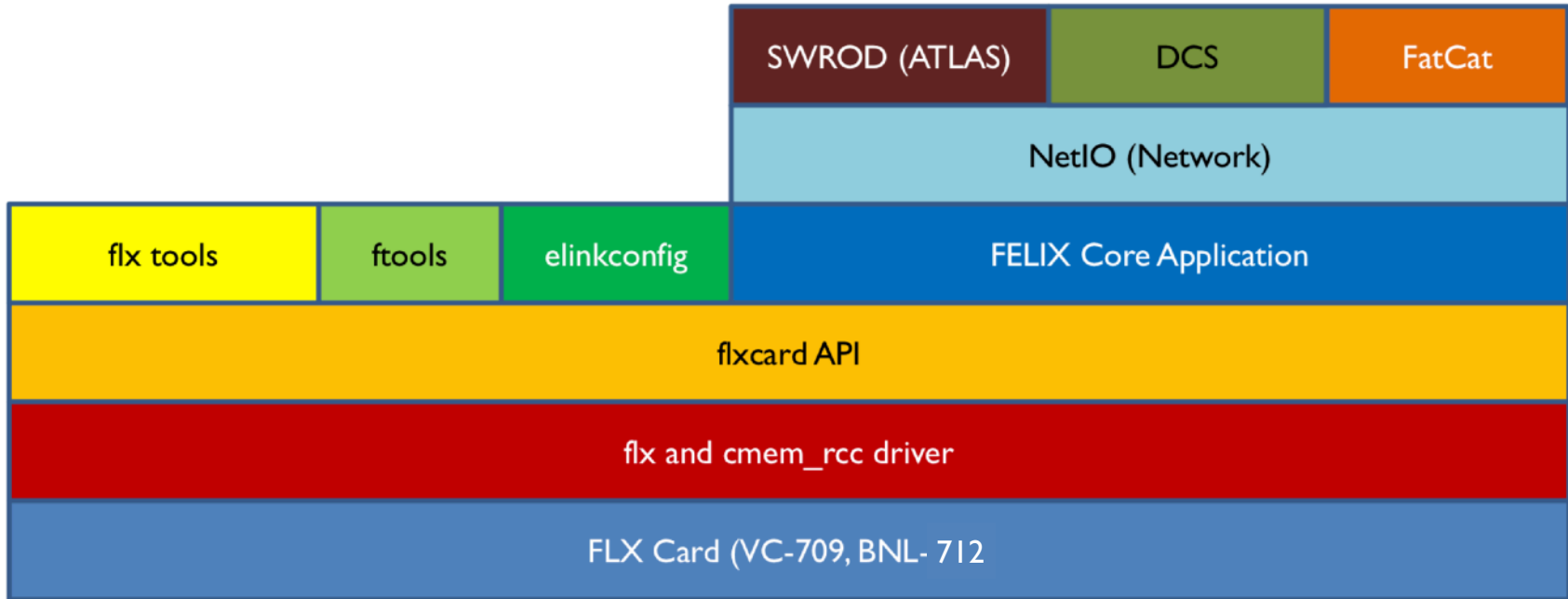
Data format



- Data buffered in the FPGA per E-link or per FULL mode link and transferred under DMA control
- Fixed block size of 1 kB by design choice
- 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-712).
- 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

FELIX software



- SLC6 and CC7 are the two supported built platforms.
- Evaluating control software candidates for FELIX Apps: manage logs, keep the core alive, ...
<http://supervisord.org>
- Classification
 - Drivers & Low level software: ftools, flx-tools etc..
 - FLX Core allows monitoring via web interface
 - NetIO allow shipping of data to the SWROD ~ Read Out System

configuration & monitoring

FELIX E-link Configurator

From FLX: 1 (710) | Read... | File: Open... | Save...

GBT Link 0 | Wide | Normal | Replicate... | Generate/Upload...

Max Chunk 2-bit: 512 | 4-bit: 512 | 8-bit: 512 | 16-bit: 512

Enable: 0x6628 Mode: 0x11011101 #BBS: 4+0+2+2+4+0+2+2 | Enable: 0x0000 Mode: 0x00000000 #BBS: 0+0+0+0+0+0+0+0

From GBT | To GBT

FELIX v1.3.3 10-Sep-2015

FELIX configuration
Graphical User Interface

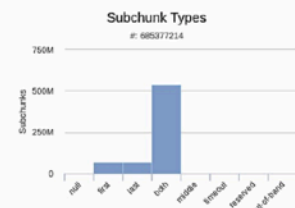
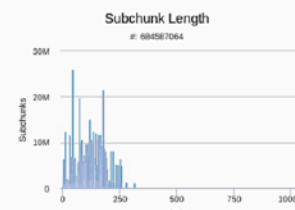
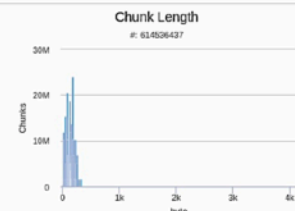
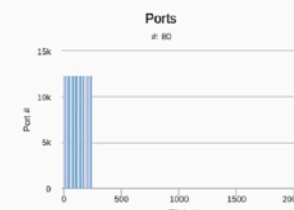
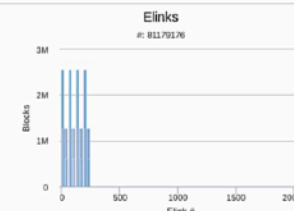
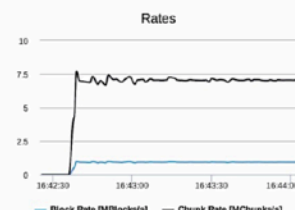
FELIX core application
monitoring web server

Info Configuration Statistics Bus

General	
timestamp	1479619942
blocks	81 MBlocks
chunks	69 MChunks
shortchunks	0.54 GChunks
block_rate	0.94 MBlocks / s
chunk_rate	7.1 MChunks / s
throughput	0 Mbyte / s
max_queue_size	74 Blocks
avg_queue_size	15 Blocks

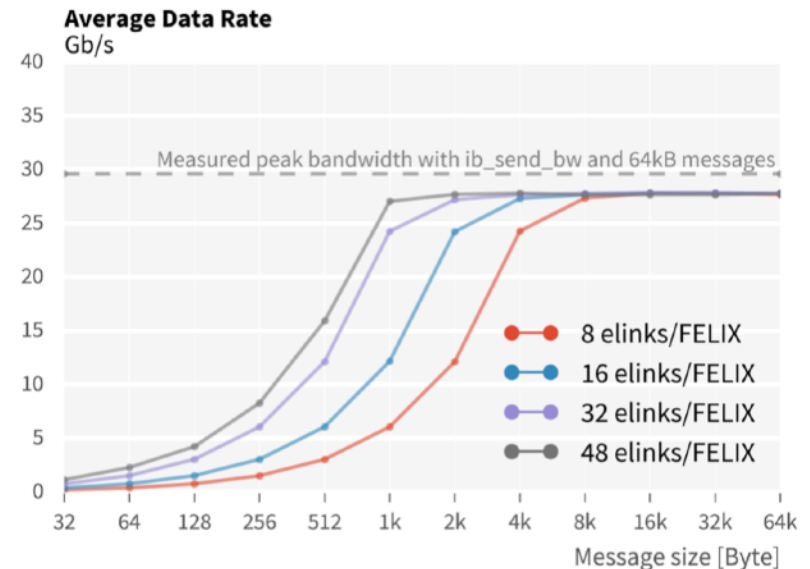
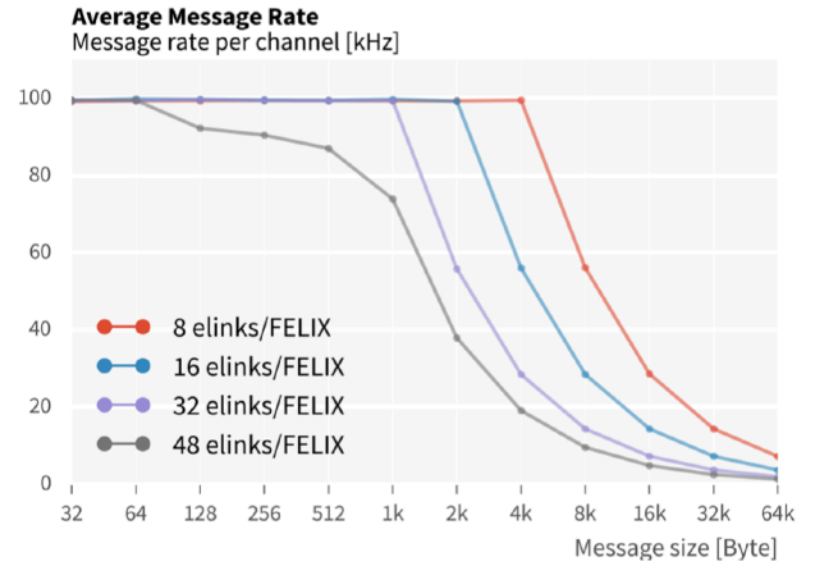
Card	
blocks_free	2.1 MBlocks

Errors	
malfomedSubchunks	0 Chunks
malfomedChunks	0 Chunks
malfomedBlocks	0 Blocks
size_error_chunks	0 Chunks
error_chunks	0 Chunks
truncated_chunks	0 Chunks

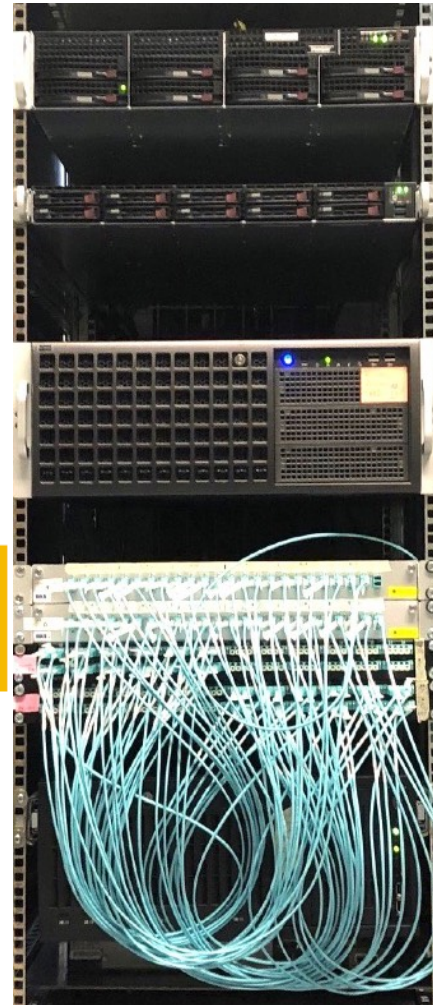
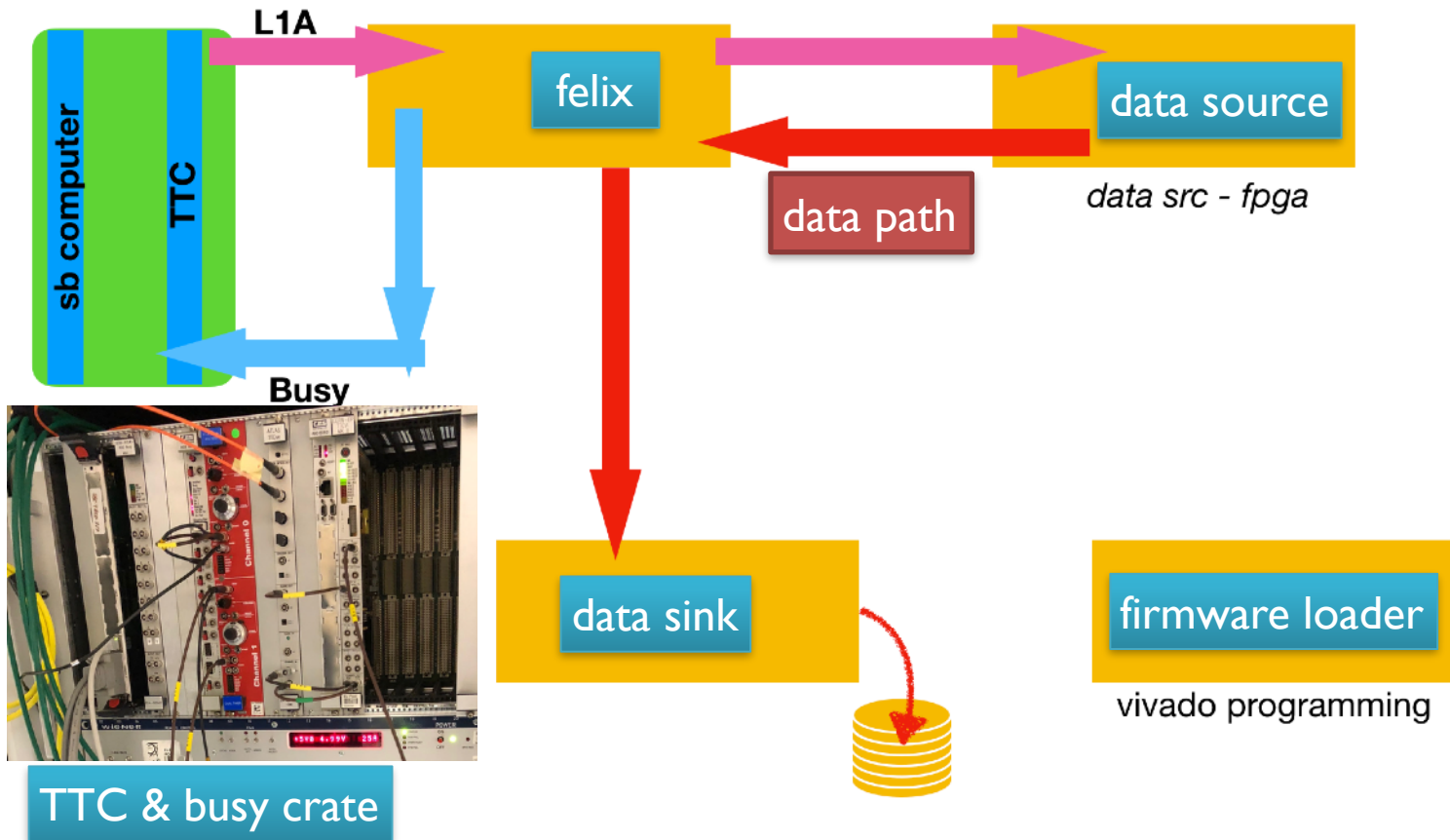


NetIO

- The generic message-based networking library “**NetIO**” has been developed for FELIX to communicate with swROD
- Provides **low-latency** or **high-throughput patterns**
- **Publish/subscribe** messaging system
- Performance of message size greater than 1kB is limited by network bandwidth



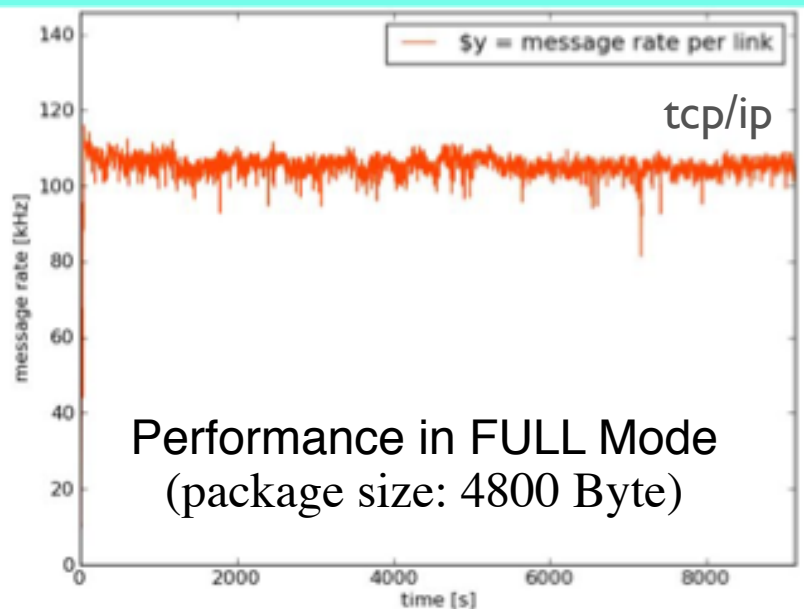
validation test setup



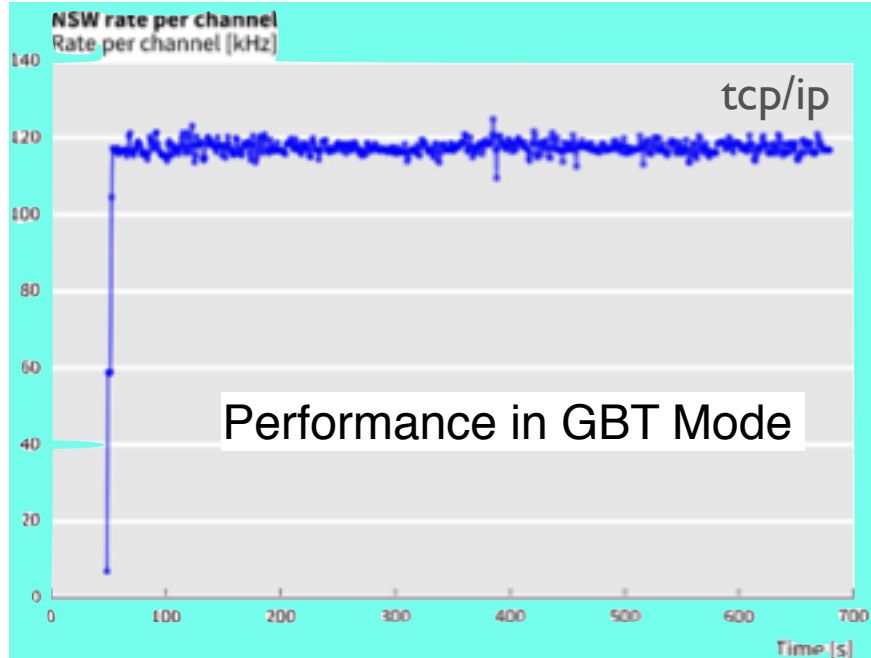
A test setup at CERN allows testing of major felix functionalities & speed
 receiving trigger & data from FE
 pushing data to swrod
 Busy mechanism + (*data to FE electronics soon*)
 Automated tests for different hardware, firmware and software versions

Felix Performance

.....



The L1Calo case is the most demanding workload for FELIX in FULL Mode.
~105-110 kHz per channel is achieved



NSW is most demanding for GBT mode in Run 3

About 120 kHz rate can be achieved in the worst-case NSW GBT-mode configuration

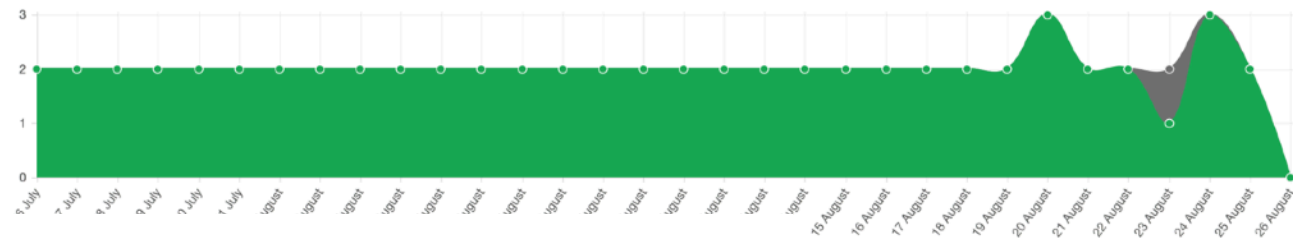
Currently studying RDMA (Remote Direct Memory Access) instead of TCP/IP: it completely eliminates the copy of the data into the system TCP buffer.

we expect better performance in NetIO.

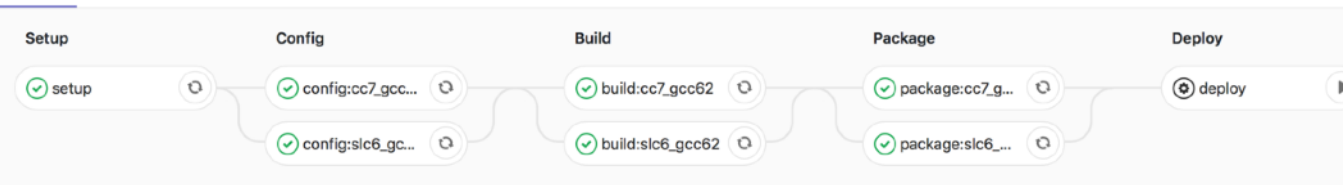
firmware & software tests

- GIT Continuous Integration

Pipelines for last month (27 Jul - 26 Aug)

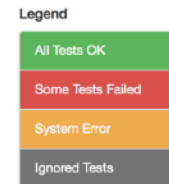


Pipeline Jobs 8



- Firmware tests

Nightly Firmware Test



FW/SW Test Results

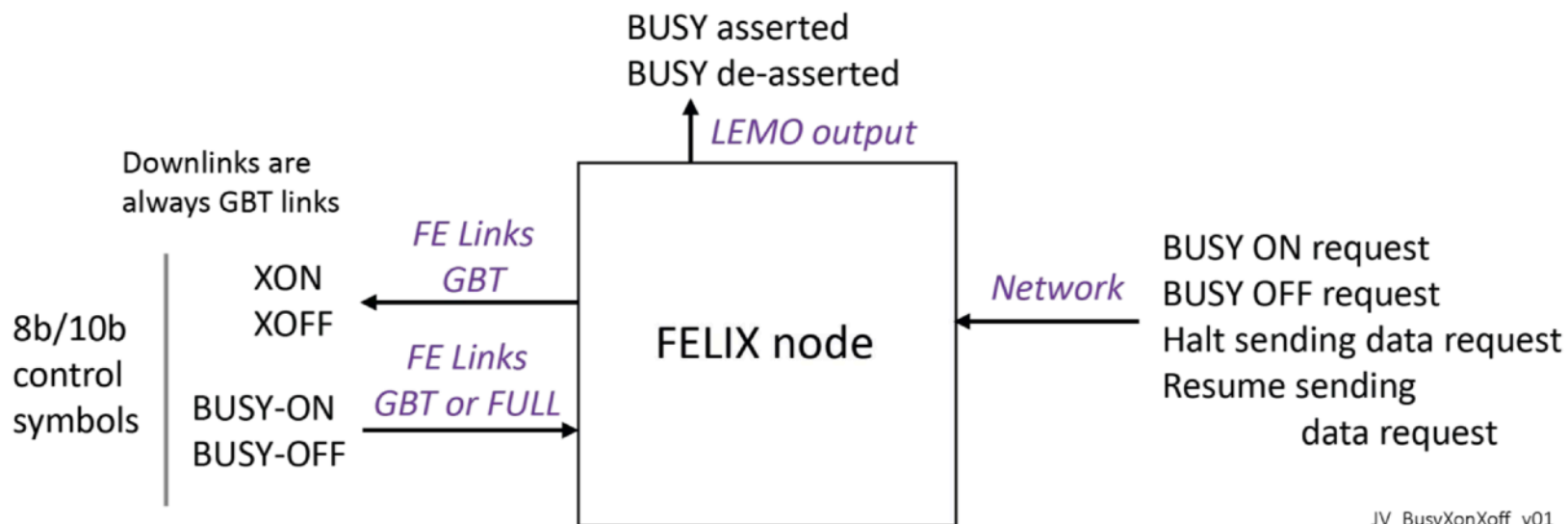
Report Name	Felix nightly tests
Work Area	CERN
Test Date	2018-06-28
Active	true
Status	Failed Tests: 4

	Firmware Tag	Hardware Type	Software Tag	Node	Passed	Failed	Ignored
📁	GBT_RM0402_4CH_CLKSELECT_GIT_RevertFromHost_rm-4	FLX709	felix-03-04-02-48-g39e7667	pc-tbed-felix-01.cern.ch	6	1	1
📁	GBT_RM0402_4CH_CLKSELECT_GIT_RevertFromHost_rm-4	FLX709	felix-03-04-02-48-g39e7667	pc-tbed-felix-01.cern.ch	6	3	1

- Software tests

FELIX BUSY & XON/XOFF

- XON and XOFF signals used to throttle the transmission of data from the front-end to FELIX to prevent buffer overflow.
- BUSY assertion expresses an emergency situation with impending buffer overflow, resulting in all ATLAS data taking being paused, and is not intended (and should not be used) for normal flow control.
- The BUSY signal will be de-asserted once there is no BUSY condition satisfied. Possible sources of BUSY include BUSY-ON or BUSY-OFF commands received from front-ends via input links and the internal state of the firmware and software.
- BUSY assertion has been tested in the validation test setup



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
- LDPB (LAr Digital Processing Blade): integration testing ongoing with MiniFELIX in FULL mode

• Level-1 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 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

- Several experiments outside the ATLAS collaboration expressed interest in the FELIX system to control and readout their detectors
 - A number of them is actively evaluating FELIX as a possible read out solution
- The current most important group is the ProtoDUNE 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

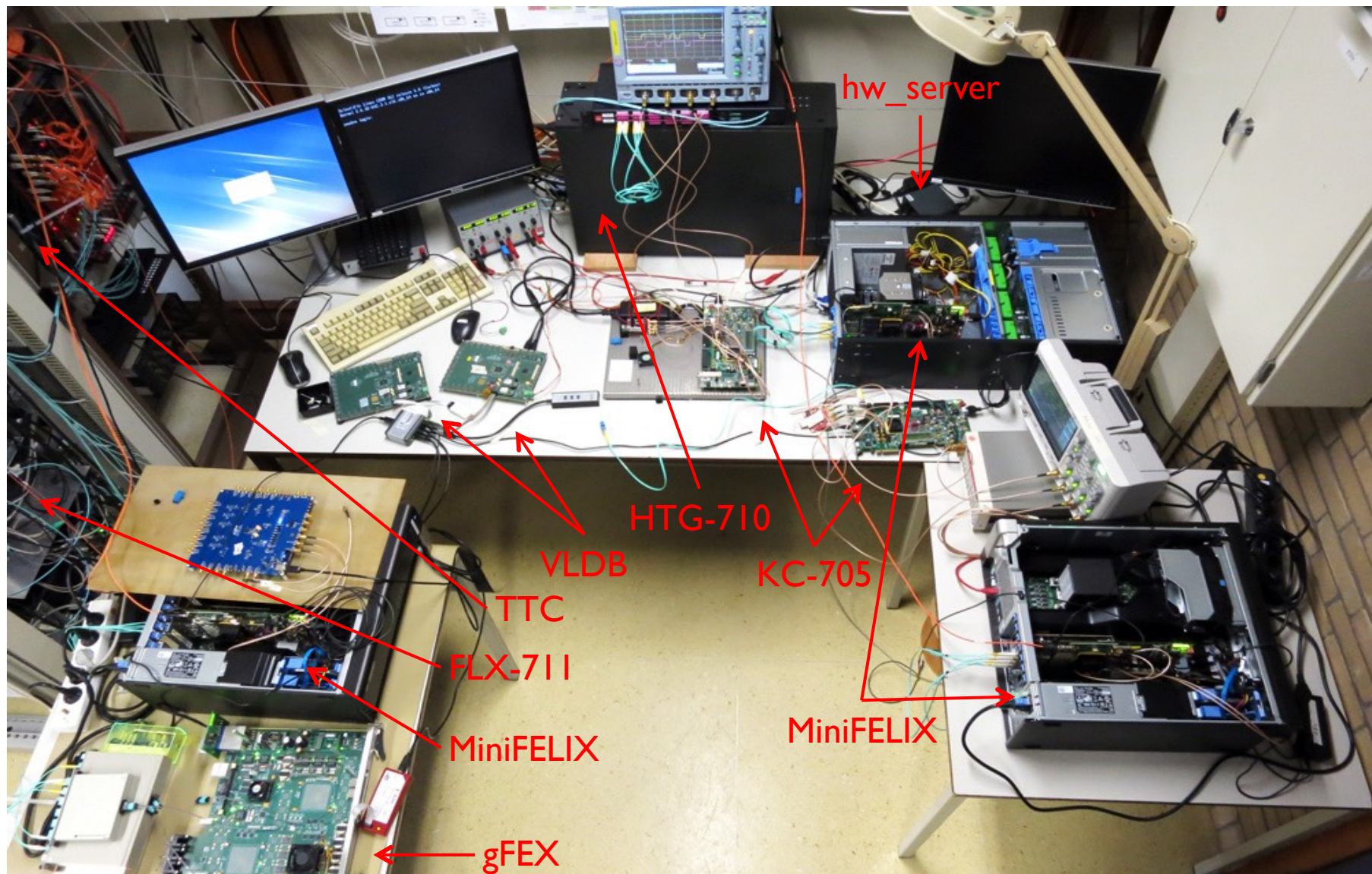
Summary & Outlook

- **FELIX is a router between custom 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.
- **Status:**
 - FELIX GBT Mode: the firmware and the software reached a development status sufficient to be distributed to ATLAS Sub-Detectors Front End developers.
 - FELIX FULL Mode reached a development status sufficient to be distributed to the Front End developers.
 - Supported hardware platforms for both modes: FLX-709 (Xilinx VC-709) and FLX-712
- **Ongoing efforts**
 - Increase overall system reliability.
 - System testing and integration: firmware to control of GBTx ASIC via its IC port, GBT-SCA ASIC via EC port.
 - A C++ API and an OPC server and client are in progress to fully support the GBT-SCA ASIC.
- **Procurement is ongoing, installation planned for 2019**

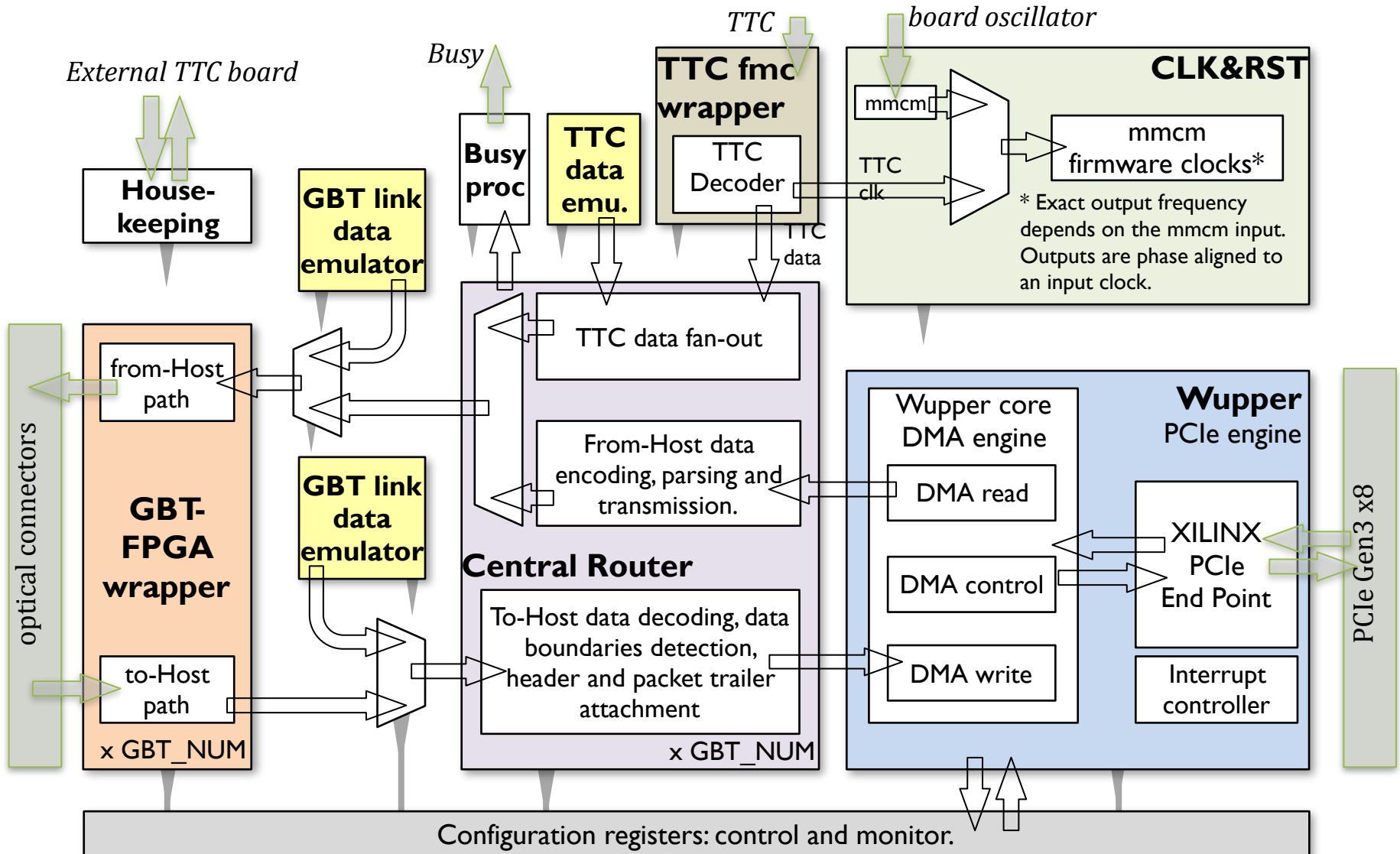
*Thank you for your
attention...*

backups

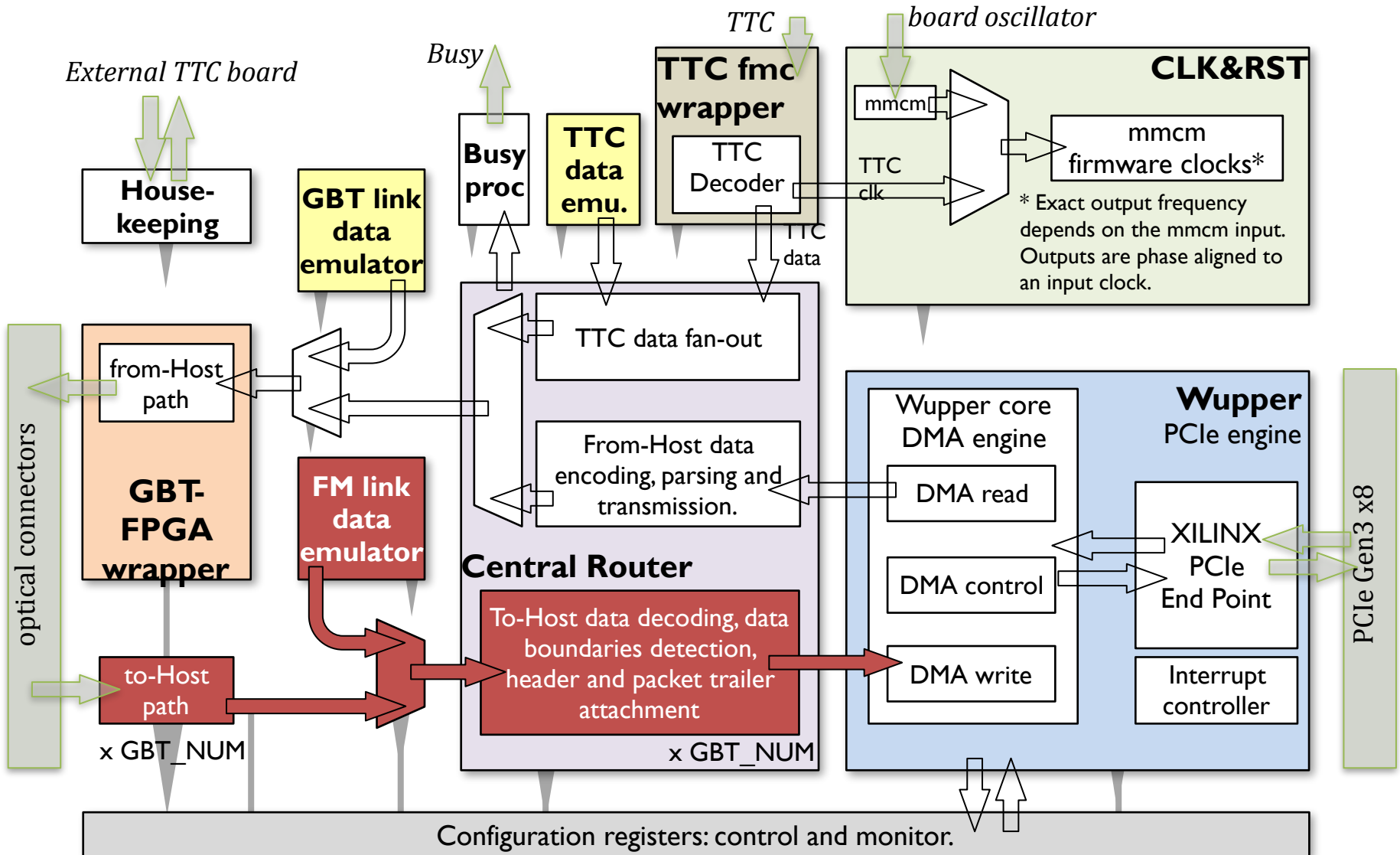
Integration workshop setup



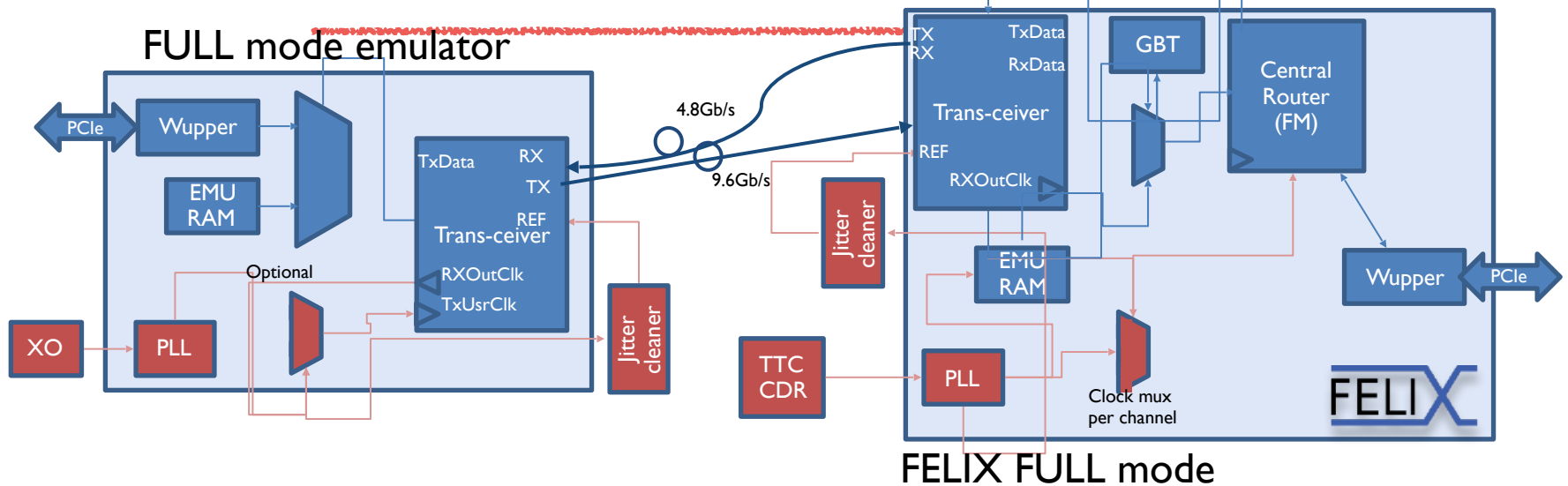
FELIX firmware: GBT mode



FELIX firmware: FULL mode

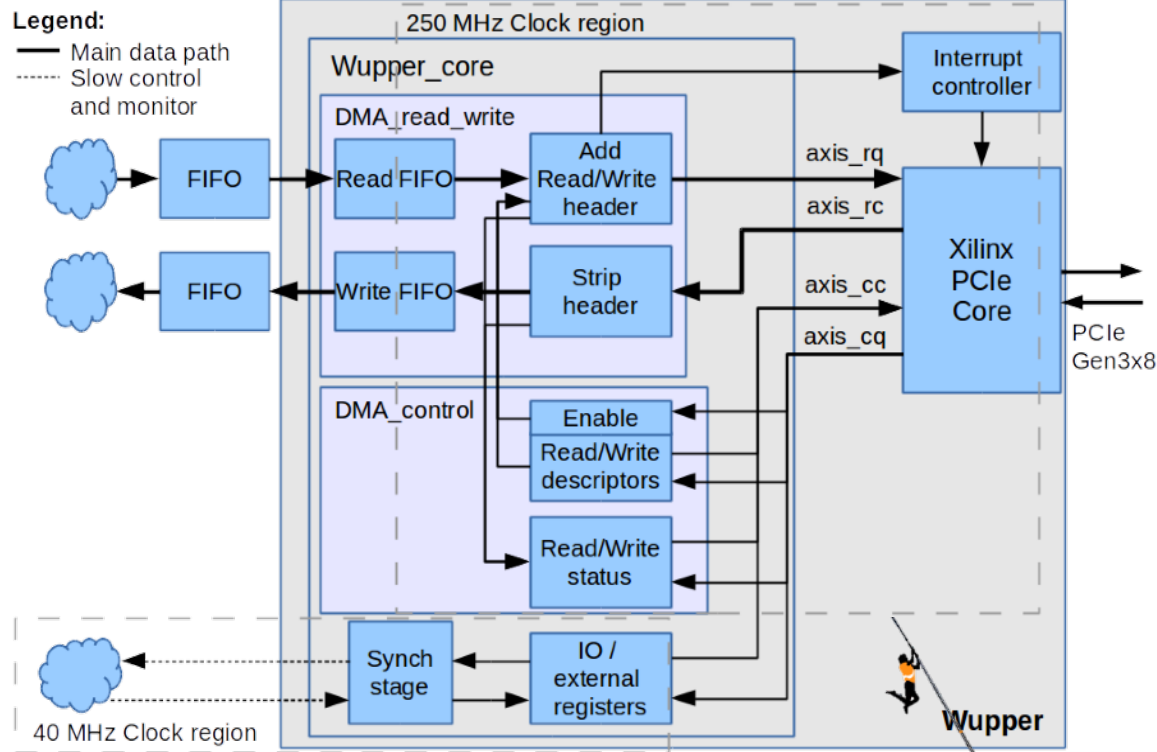


FULL mode chain and clocking



- 9.6Gb/s Link tested with **32-bit PRBS31** generator and checker.
 - No error occurred for ~72 hours run. **BER < 1E-15**.
- Complete design tested with different FPGA based emulated data generators (gFEX, VC709, VC707)
 - 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

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 to the Xilinx Virtex-7 (Kintex Ultrascale) PCIe Gen3 Integrated Block for PCI Express
- Xilinx AXI (ARM AMBA) Stream Interface (UG761)
- MSI-X compatible interrupt controller
- Applications access the engine via simple FIFOs
- Register map for programmed I/O synchronized to a lower clock speed