FELIX
Front End LInk eXchange

ATLAS EXPERIMENT

FELIX: the New Detector Readout System for the ATLAS Experiment

Frans Schreuder
Nikhef, The Netherlands
f.schreuder@nikhef.nl

On behalf of the ATLAS TDAQ Collaboration

Nikhef

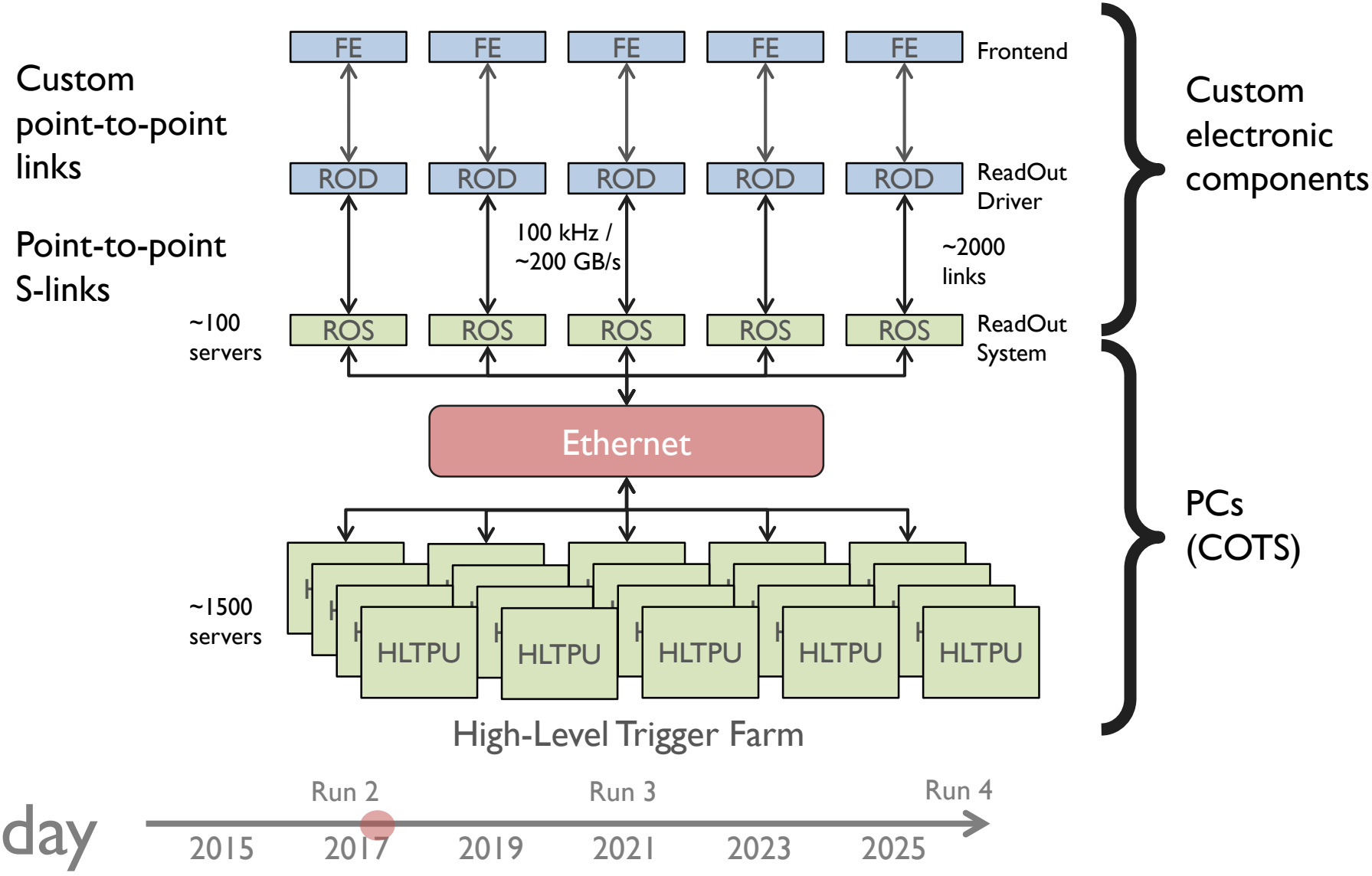
SCIPP
UC SANTA CRUZ

CERN

Outline

- ATLAS DAQ Today, Phase I upgrade and HL-LHC
- FELIX Hardware
- FELIX Firmware: 2 modes of operation
- System integration and testing
- Integration with different ATLAS and non-ATLAS subdetectors

ATLAS DAQ Today

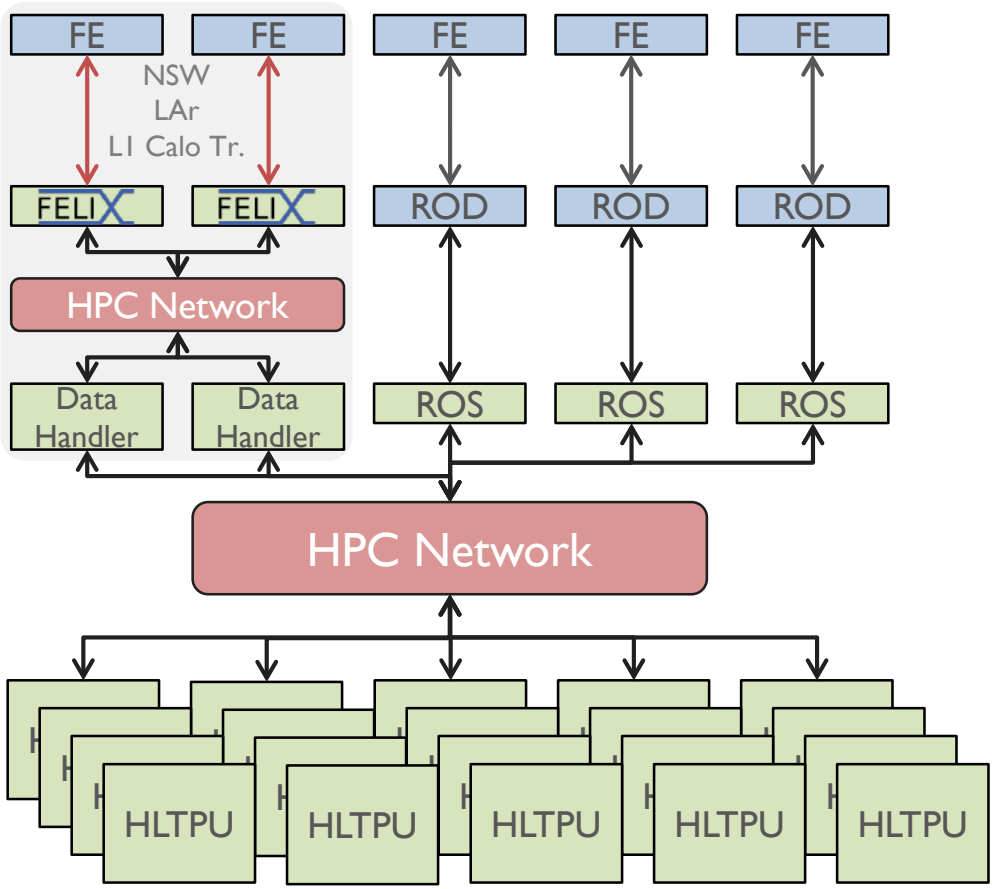


Upgrade for Phase-I

GBT links* or FULL mode links

PCs

40 Gb Ethernet, Infiniband

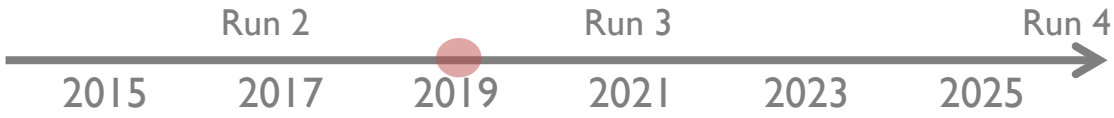


Custom electronic components including FELIX cards

PCs (COTS)

* GBT: GigaBit Transceiver with Versatile Link

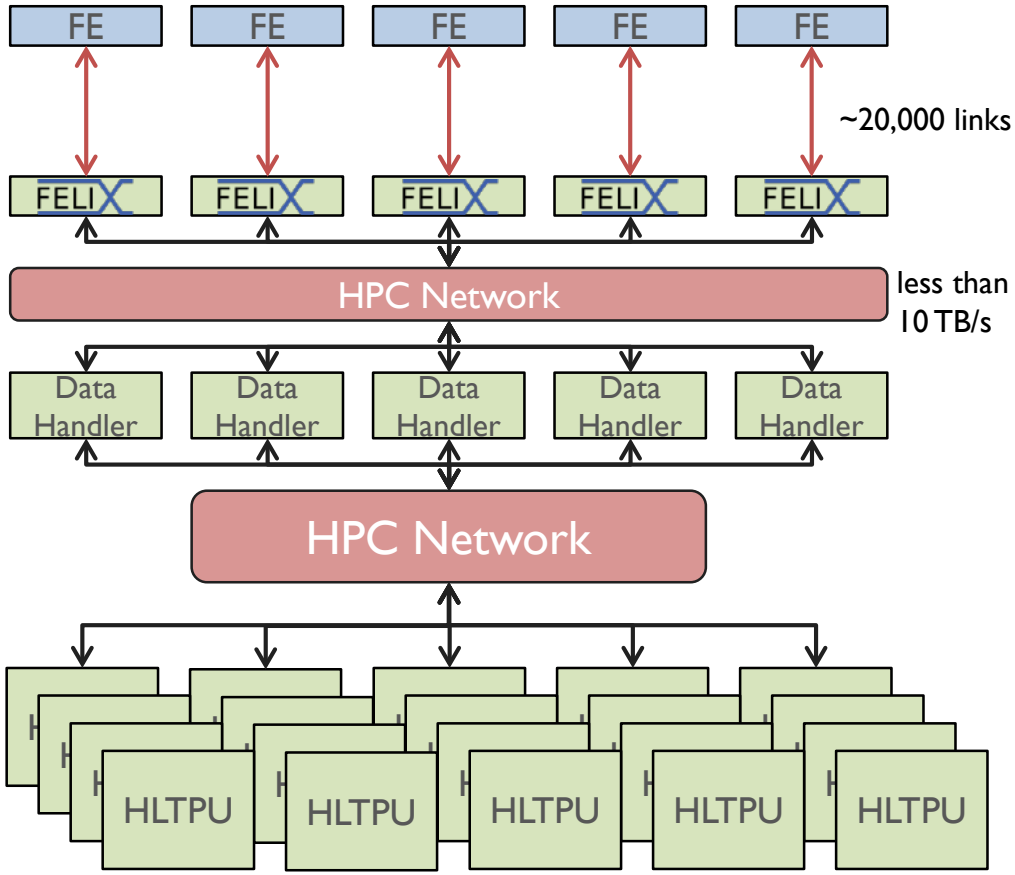
2019



Upgrade for HL-LHC

GBT, LpGBT* or FULL mode links

COTS network technology



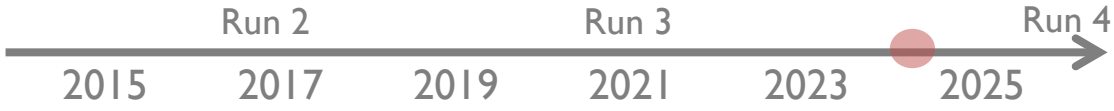
High-Level Trigger Farm

Custom electronic components including FELIX cards

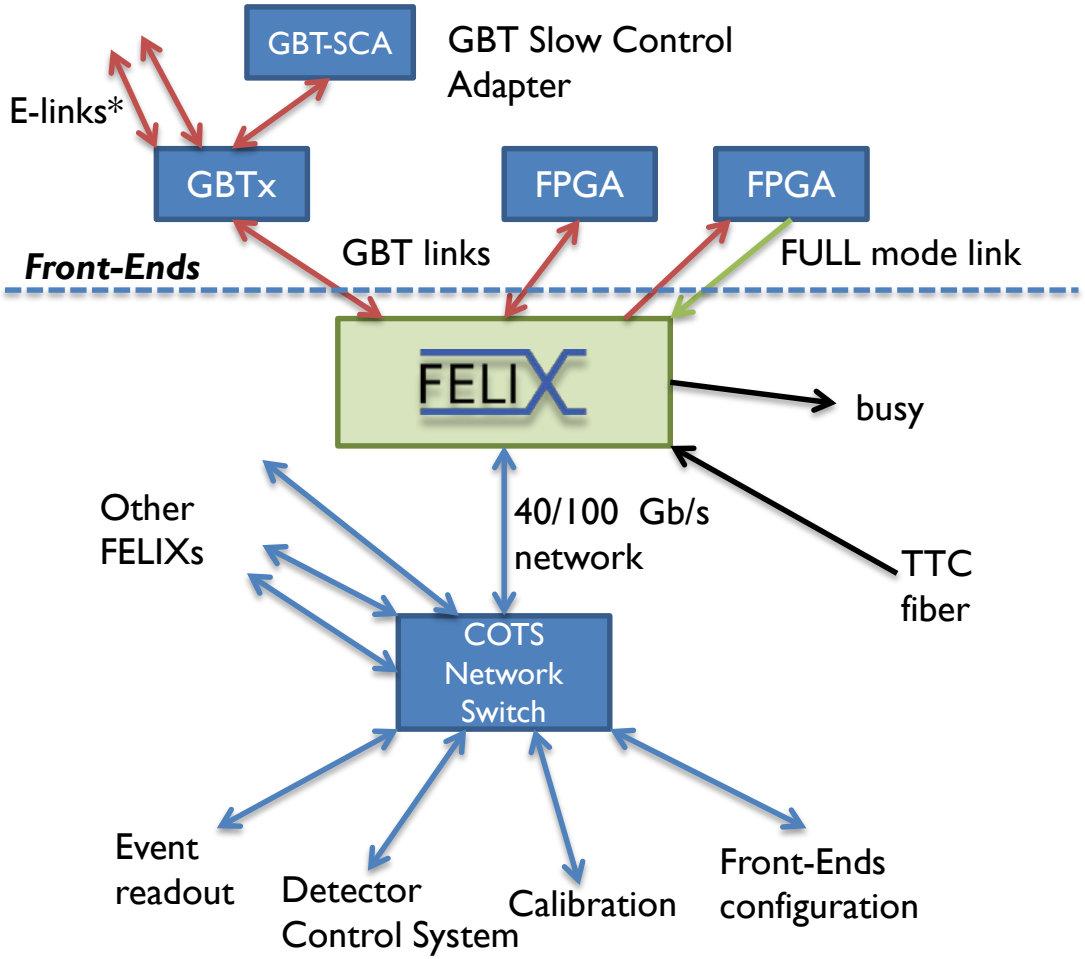
PCs (COTS)

*LpGBT: Low power GBT

2024



FELIX functionality



*** 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.

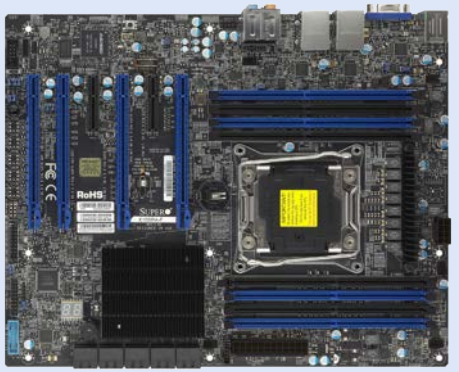
- Scalable architecture
- Routing of event data, detector control, configuration, calibration, monitoring
- Connect the ATLAS detector Front-Ends to the DAQ system, for both the to and from FE directions
- Configurable E-links in GBT Mode
- Detector independent
- TTC (Timing, Trigger and Control) distribution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

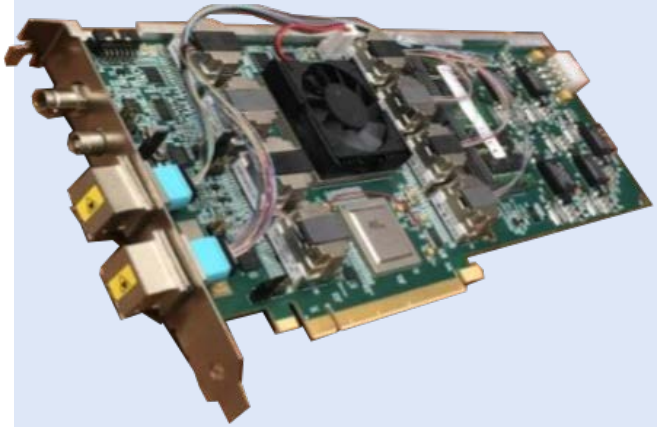
FELIX server PC components

- **VC-709 from Xilinx**
  - Virtex7 X690T FPGA
  - FLX-709 or miniFelix
  - 4 optical links (SFP+)
  - Intended for FE development support
  - PCIe Gen3 x8

- **TTCfx (v3) mezzanine card**
  - TTC input
  - ADN2814 for TTC clock-data recovery
  - Si5345 jitter cleaner

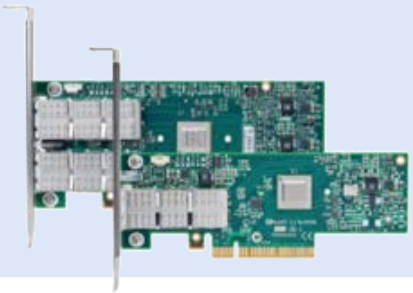
- **SuperMicro X10SRA-F used for development**
  - Broadwell CPU, e.g. E5-1650V4, 3.6GHz
  - PCIe Gen3 slots



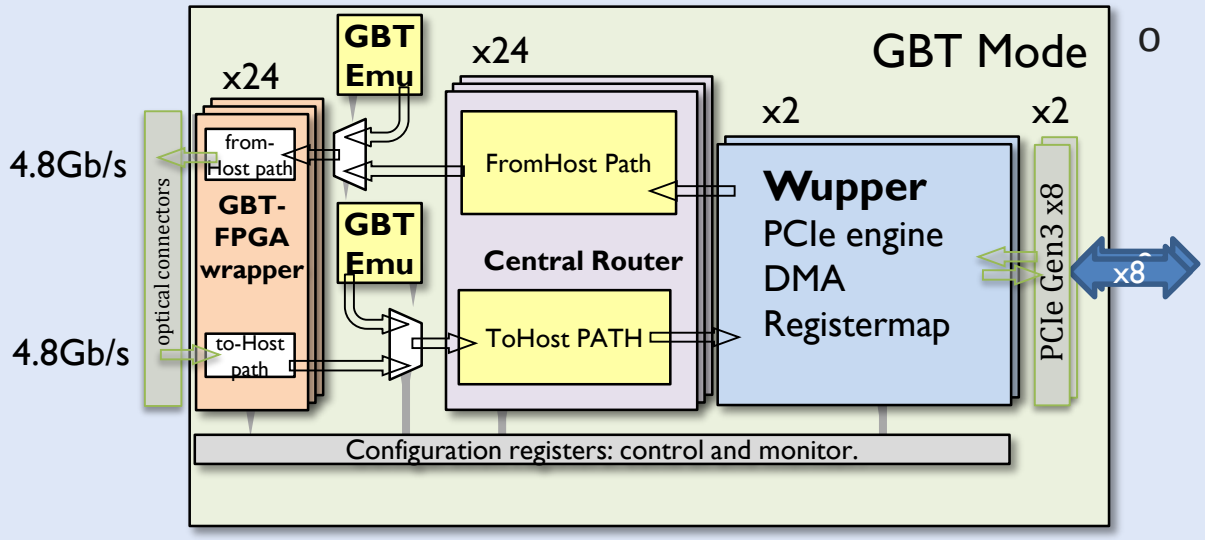


- **BNL-711 from BNL**
  - Xilinx Kintex Ultrascale XCKU115
  - 48 optical links (MiniPODs)
  - FELIX Phase-I prototype
  - TTC input ADN2814
  - SI5345 jitter cleaner
  - PCIe Gen3 x16 (2x8 with bridge)
  - Version 2.0 currently tested

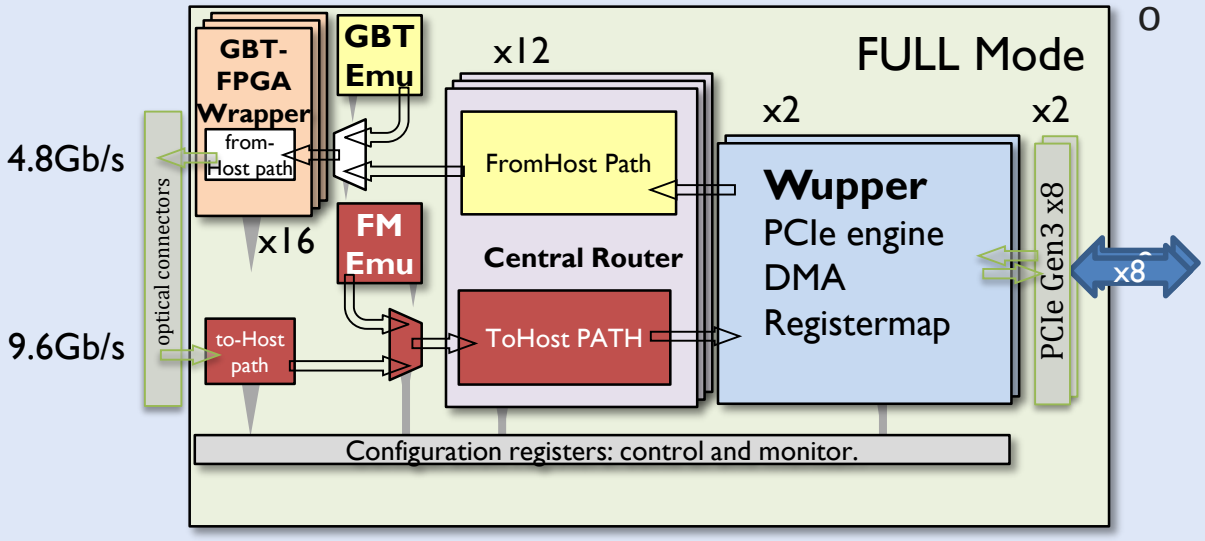
- **Mellanox ConnectX-3**
  - 2x FDR/QDR Infiniband
  - 2x 10/40 GbE



FELIX: modes of operation

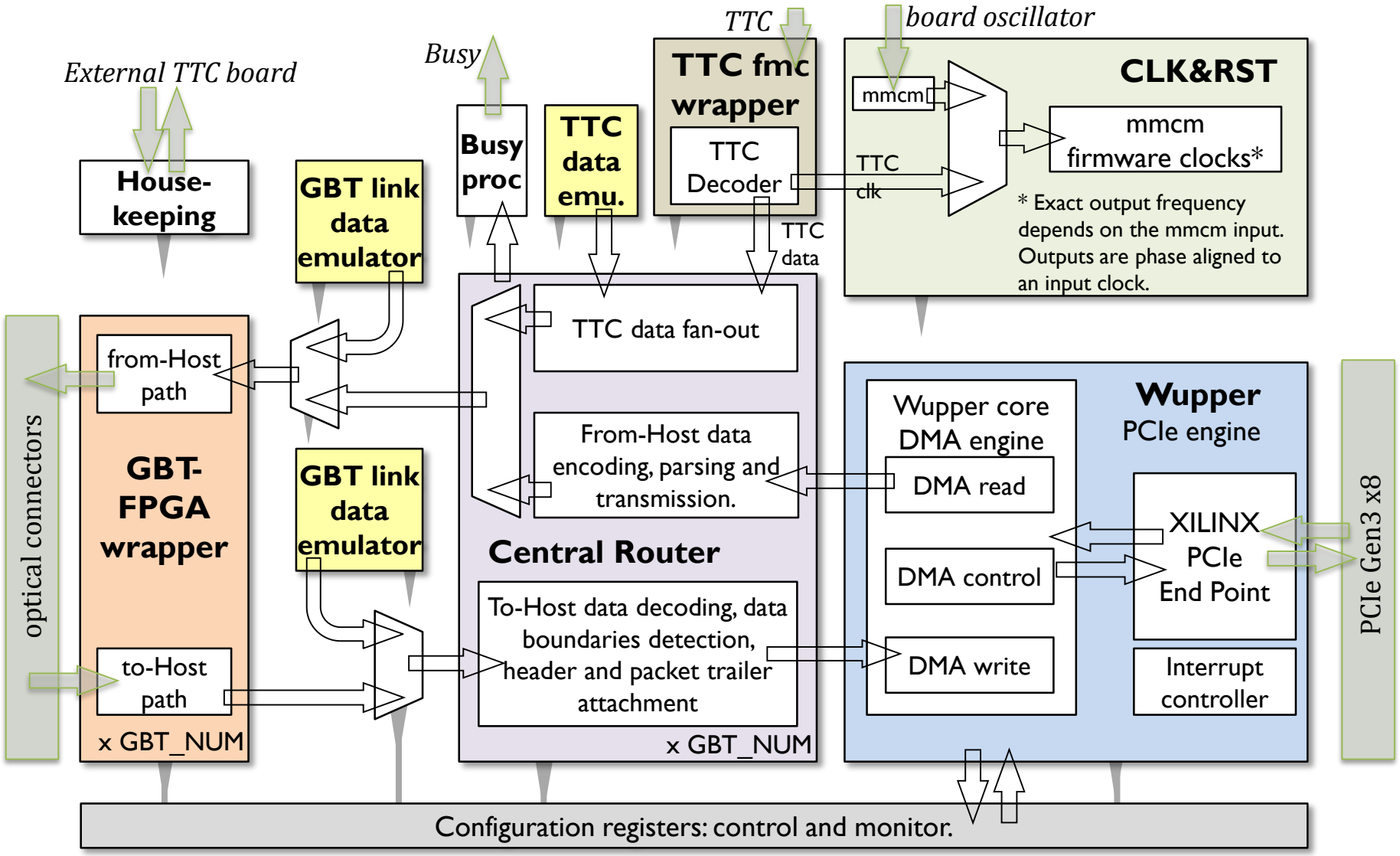


- **GBT mode**
  - Line rate: 4.8 Gb/s
  - Up to 24 bidirectional optical links
  - 3.2 Gb/s payload with FEC or 4.48 Gb/s payload
  - Routes TTC information
  - Optical link divided in E-Links
  - Communicate with GBTx & GBT-SCA

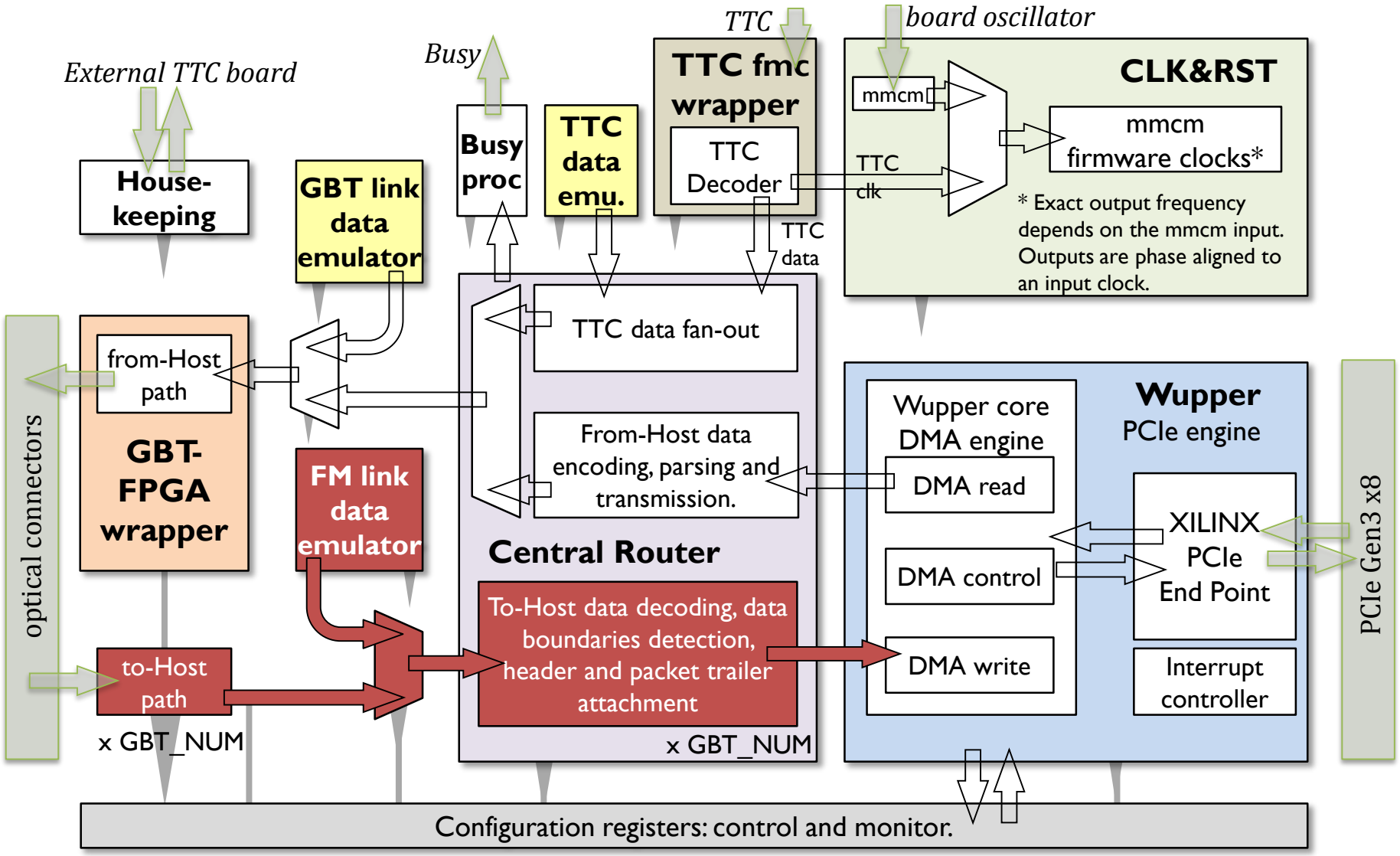


- **FULL mode**
  - Line rate: 9.6 Gb/s
  - Up to 12 bidirectional optical links
  - Routes TTC information
  - 7.68 Gb/s payload: 8B/10B encoding
  - CRC
  - BUSY-ON and OFF
  - 4.8 Gb/s GBT links to FE

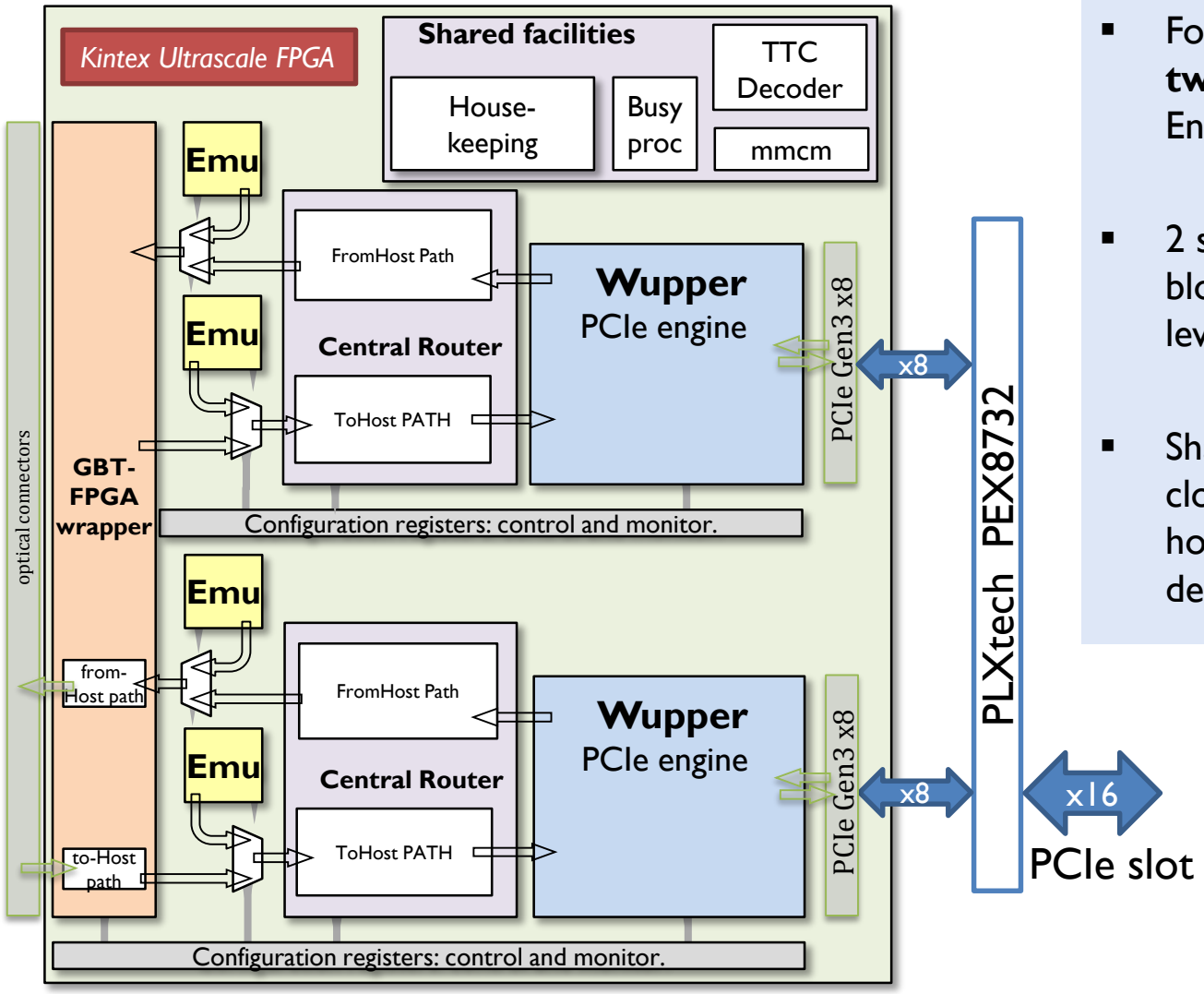
FELIX GBT firmware block diagram



FELIX FULL mode firmware block diagram

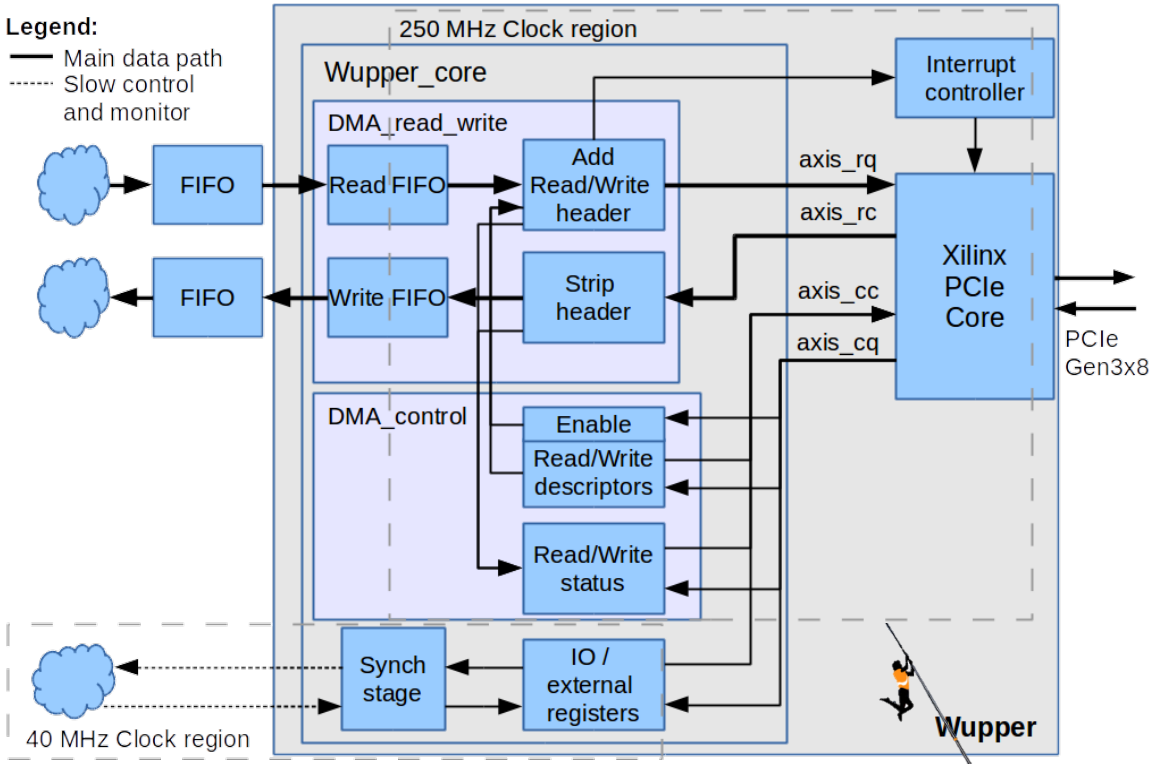


FELIX firmware on PCIe x16



- For BNL-711: the PC will see **two** independent Gen3 x8 lanes EndPoints, as Xilinx PCIe devices.
- 2 sets of an identical firmware block are instantiated in the top level design.
- Shared facilities in grey (include clock resources, the housekeeping module, TTC decoder and busy logic).

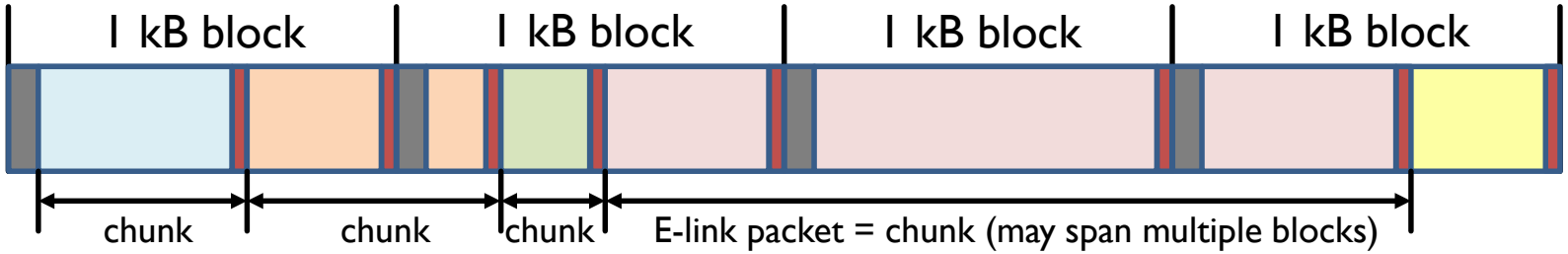
Wupper: PCIe engine for FELIX



- 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

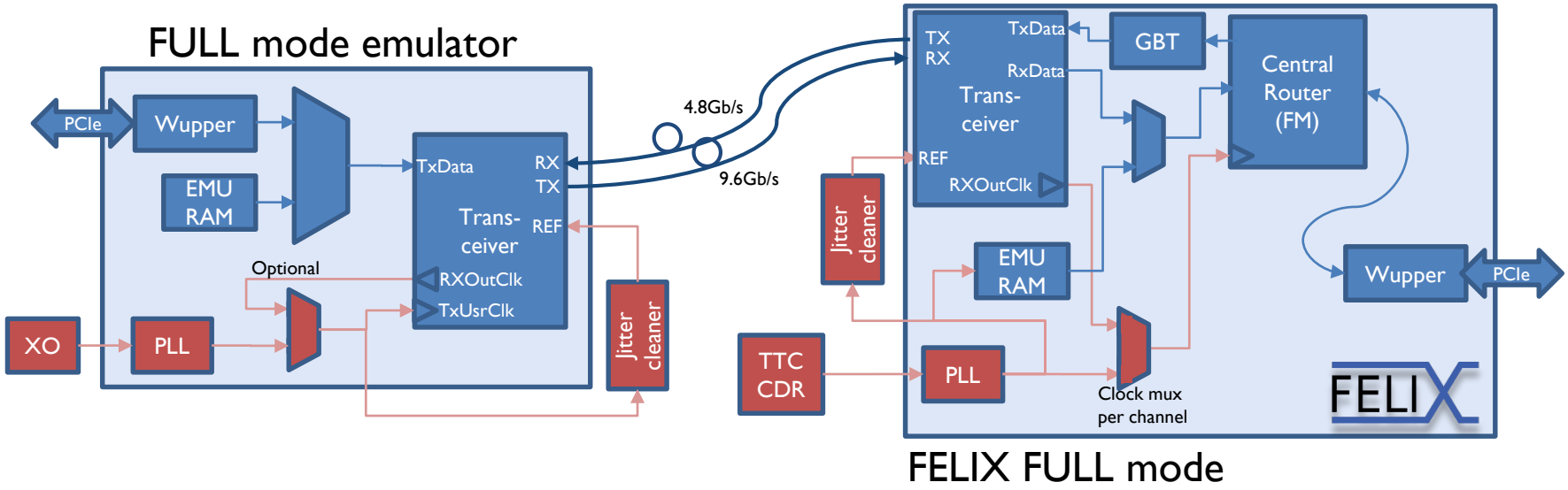
Data format



- 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.

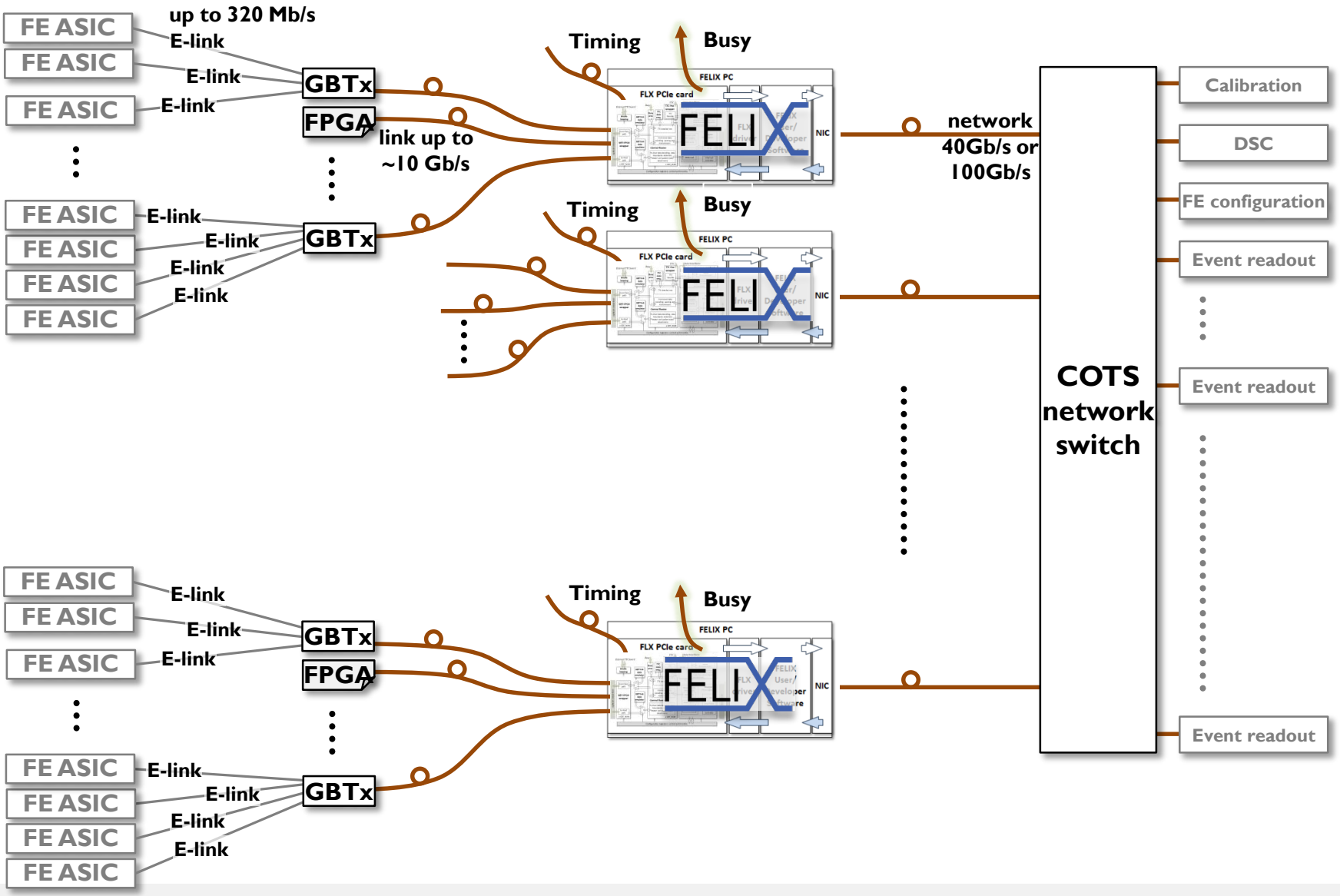
- Block header: (32 bits)
  - E-link ID
  - Block sequence
  - Start of block symbol
- Fragment trailer (16 bits)
  - Fragment type
    - First, last, both, middle, null
  - Flags
    - Error, truncation, timeout, CRC error
  - Fragment length
    - 10 bits

FULL mode chain and clocking

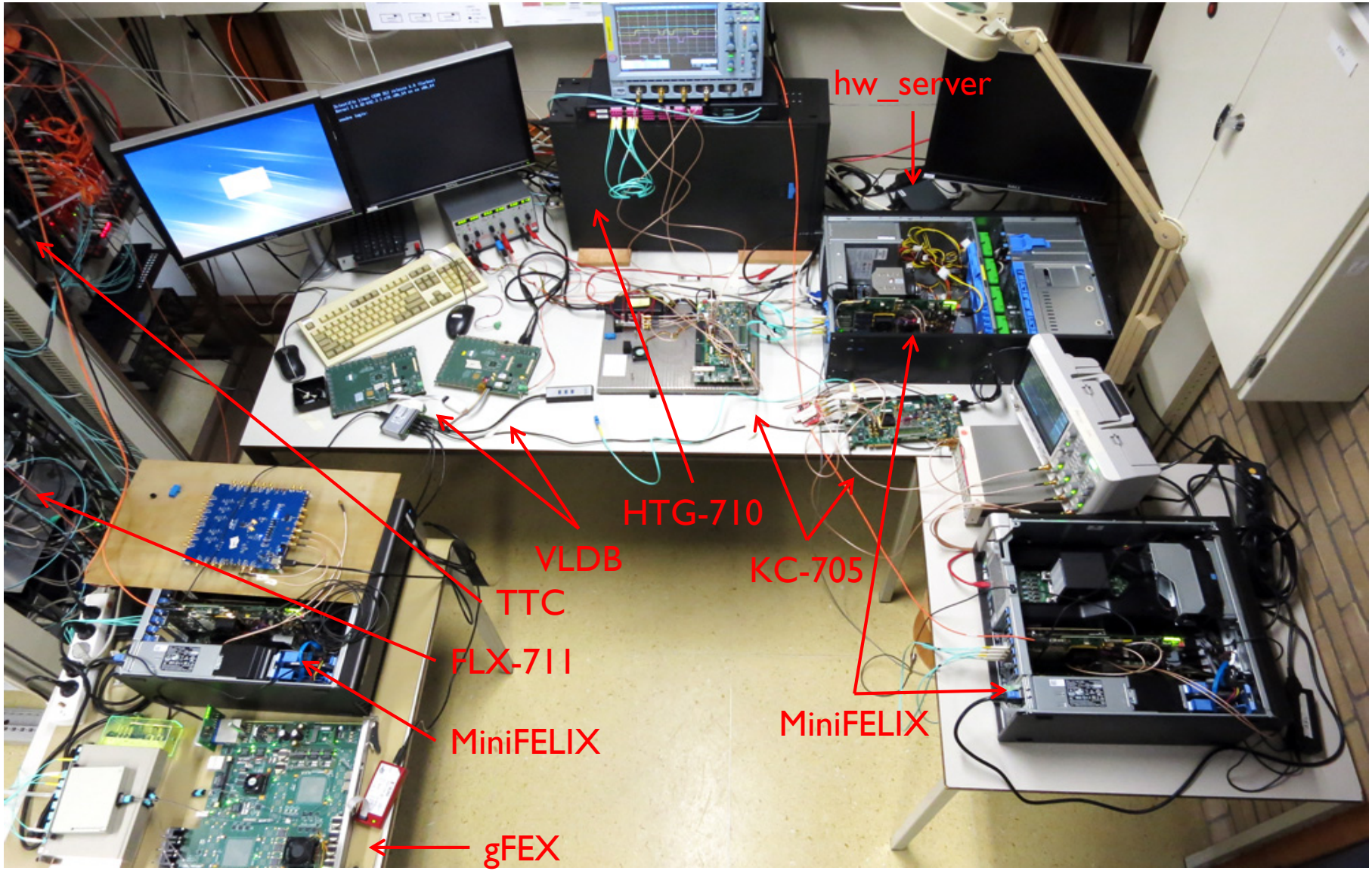


- 9.6Gb/s Link tested with **32-bit PRBS31** generator and checker.
  - No error occurred for ~72 hours run. **BER < 1E-15**.
- Complete design tested with different FPGA based emulated data generators (gFEX, VC709, VC707)
  - No errors occurred for several TB of data transmitted
- Optional RX clock recovery for TX in emulator
- Clock recovery in FELIX, or local clock for internal emulator

FELIX data flow overview



Integration workshop setup



FELIX software: status update

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

Register map automation – Jinja 2

YAML:

```yaml
MINI_EGROUP_CTRL:
  number: 24
  format_name: MINI_EGROUP_CTRLS
  entries:
    - name: EC_TOHOST
      format_name: EC_TOHOST_{index:02}
      type_name: EC_TOHOST
      desc: Configures the ToHost Mini egroup in EC mode
      type: W
      bitfield:
        - range: 3
          name: BIT_SWAPPING
          desc: 1: two input bits are swapped
          default: 0
        - range: 2..1
          name: ENCODING
          desc: Configures encoding of the EC channel
          default: 0x2
        - range: 0
          name: ENABLE
          desc: Enables the EC channel
          default: 1
    - name: EC_FROMHOST
```

Jinja

WupperCodeGen

C++

```
...
/* MINI_EGROUP_CTRL */
{ REG_EC_TOHOST_00,
  "Full Register",
  0x02700,
  REGMAP_REG_READ|REGMAP_REG_WRITE,
  REGMAP_ENDPOINT_0|REGMAP_ENDPOINT_1
},
{ REG_EC_FROMHOST_00,
  "Full Register",
  0x02710,
  REGMAP_REG_READ|REGMAP_REG_WRITE,
  REGMAP_ENDPOINT_0|REGMAP_ENDPOINT_1
},
{ REG_MARK_FOR_DELETION_00,
  "Full Register",
  0x02720,
  REGMAP_REG_READ,
  REGMAP_ENDPOINT_0|REGMAP_ENDPOINT_1
},
{ REG_EC_TOHOST_01,
  "Full Register",
  0x02730,
  REGMAP_REG_READ|REGMAP_REG_WRITE,
  REGMAP_ENDPOINT_0|REGMAP_ENDPOINT_1
},
...
```

LaTeX

MINI_EGROUP_CTRL						
0x2700	0,1	EC_TOHOST_00				
			BIT_SWAPPING	3	W	0: two input bits of EC e-link are as documented, 1: two input bits are swapped
			ENCODING	2:1	W	Configures encoding of the EC channel
			ENABLE	0	W	Enables the EC channel
0x2710	0,1	EC_FROMHOST_00				
			BIT_SWAPPING	5	W	0: two output bits of EC e-link are as documented, 1: two output bits are swapped
			ENCODING	4:1	W	Configures encoding of the EC channel

VHDL

```
...
  --** MINI_EGROUP_CTRL
  constant REG_EC_TOHOST_00                : std_logic_vector(19 downto 0) := x"02700";
  constant REG_EC_FROMHOST_00              : std_logic_vector(19 downto 0) := x"02710";
  constant REG_EC_TOHOST_01                : std_logic_vector(19 downto 0) := x"02730";
  constant REG_EC_FROMHOST_01              : std_logic_vector(19 downto 0) := x"02740";
...
  when REG_EC_TOHOST_00    => register_map_control_s.EC_TOHOST_00.BIT_SWAPPING    <= register_write_data_40_s(3 downto 3);    -- 1: two input bits are swapped
                              register_map_control_s.EC_TOHOST_00.ENCODING       <= register_write_data_40_s(2 downto 1);    -- Configures encoding of the EC channel
                              register_map_control_s.EC_TOHOST_00.ENABLE         <= register_write_data_40_s(0 downto 0);    -- Enables the EC channel
  when REG_EC_FROMHOST_00  => register_map_control_s.EC_FROMHOST_00.BIT_SWAPPING  <= register_write_data_40_s(5 downto 5);    -- 1: two output bits are swapped
                              register_map_control_s.EC_FROMHOST_00.ENCODING     <= register_write_data_40_s(4 downto 1);    -- Configures encoding of the EC channel
                              register_map_control_s.EC_FROMHOST_00.ENABLE       <= register_write_data_40_s(0 downto 0);    -- Configures the FromHost Mini egroup in EC mode
  when REG_EC_TOHOST_01    => register_map_control_s.EC_TOHOST_01.BIT_SWAPPING    <= register_write_data_40_s(3 downto 3);    -- 1: two output bits are swapped
                              register_map_control_s.EC_TOHOST_01.ENCODING       <= register_write_data_40_s(2 downto 1);    -- Configures encoding of the EC channel
                              register_map_control_s.EC_TOHOST_01.ENABLE         <= register_write_data_40_s(0 downto 0);    -- Enables the EC channel
  when REG_EC_FROMHOST_01  => register_map_control_s.EC_FROMHOST_01.BIT_SWAPPING  <= register_write_data_40_s(5 downto 5);    -- 1: two output bits are swapped
                              register_map_control_s.EC_FROMHOST_01.ENCODING     <= register_write_data_40_s(4 downto 1);    -- Configures encoding of the EC channel
                              register_map_control_s.EC_FROMHOST_01.ENABLE       <= register_write_data_40_s(0 downto 0);    -- Configures the FromHost Mini egroup in EC mode
...
```

Integration test with front-ends

- ATLAS sub-detector test setups, currently implementing FELIX
  - *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

Integration test with front-ends

- *Non-ATLAS detectors connected to FELIX*
  - 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 readout solution
  - The current most noticeable group is the ProtoDUNE collaboration (vertical slice):
    - **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 readout by a FELIX system and long term stability testing is now being target

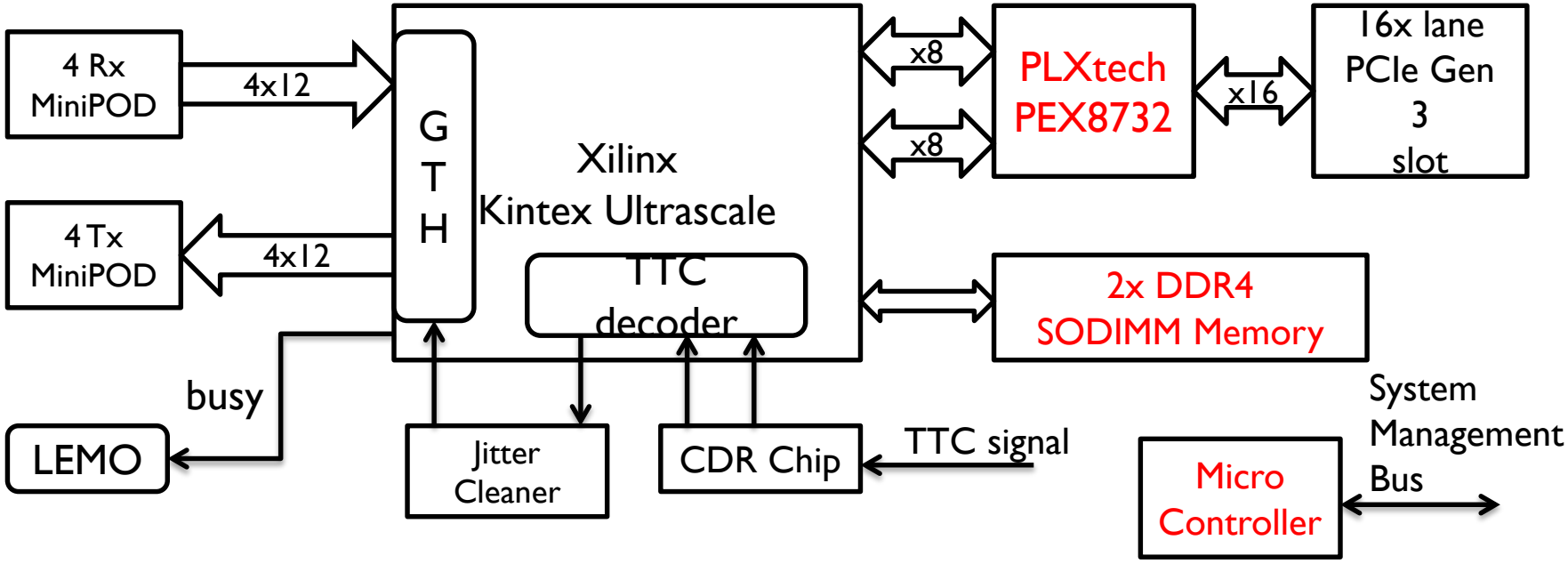
Summary

- 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-711 (BNL 16 lane card)
- Ongoing efforts
  - Increase overall system reliability.
  - Increase the number of input channels supported to 24 for GBT mode
  - Extend the 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 in 2018, installation in 2019

Thank you for your attention

Features of BNL-711

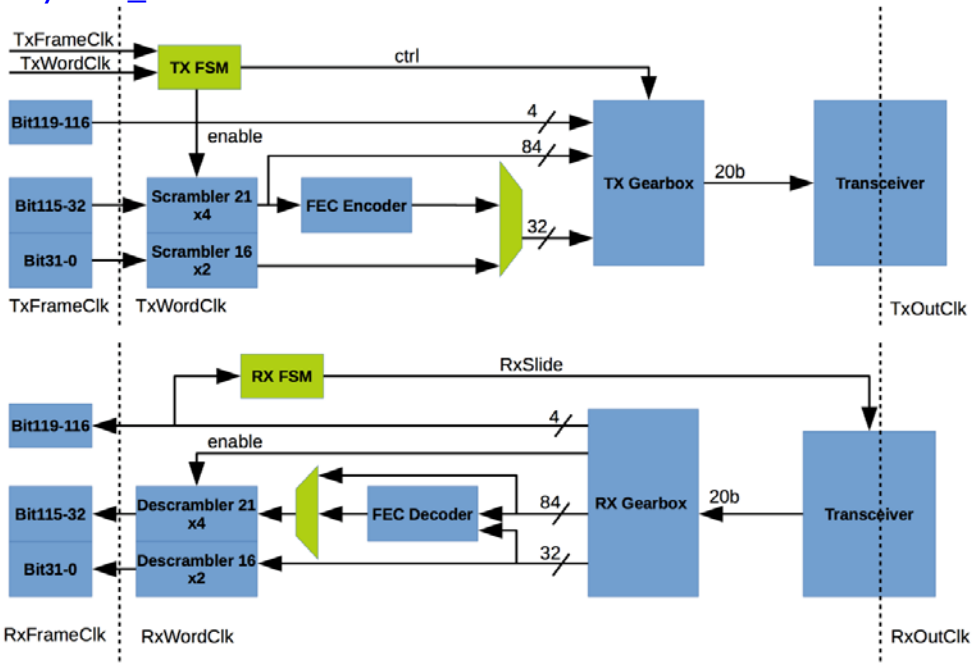
FELIX base line hardware platform: PCIe FPGA board gen3 x16, "BNL-711"



- Developed at BNL also as the DAQ platform for the LTDB (Liquid Argon Trigger Digitizer Board) production test platform
- PLXtech PEX8732 to handle PCIe Gen3 x16 lanes (max 128 Gbps) interface to host
- 48-ch MiniPOD TX & RX, up to 14Gb/s per link
- 2x SODIMM DDR4 interfaces (not used in FELIX, removed in v2.0)
- Integrated TTC interface, busy output, and on-board jitter cleaner
- Micro-Controller (Atmega 324A) for FPGA firmware update and version control

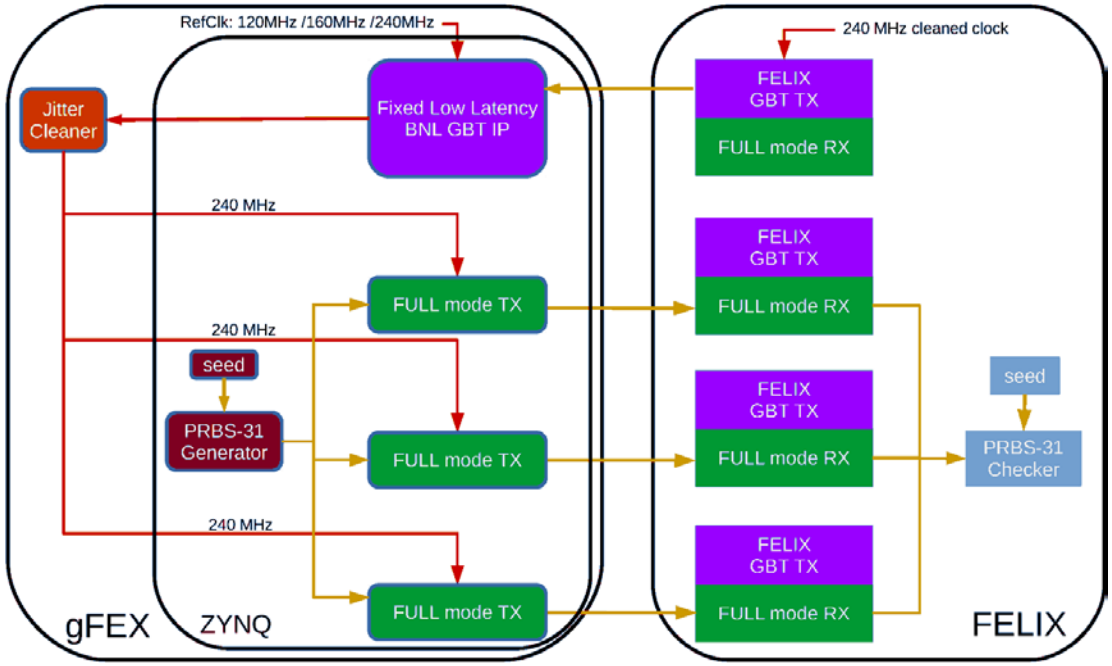
Transceiver wrapper for FELIX

- Supports 4.8 Gb/s **GBT mode**, and 9.6 Gb/s **FULL mode** defined 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.
  - Some blocks like *Rx side frame alignment*, *Tx side time domain crossing* are redesigned.
  - The single channel example design for KC705/VC709 can be found at https://github.com/simpway/GBT_KC705



Integration test: Clock stability

- Goal: asses the quality of the clock distributed by FELIX with an indirect measurement



- Use a BERT (Bit Error Rate Tester) at 9.6 Gbps (bit time 100 ps) as an indirect measurement of the clock stability (overall clock has to cope with that)
- Measurement done with both local and TTC clock
- No errors occurred for ~72 hours run. **BER < 1E-15**.

Register map automation – Jinja 2

```
2791        ---- ## GENERATED code BEGIN #4 ----
2792        ---------------------------------------
2793        when REG_STATUS_LEDS                    => register_map_control_s.STATUS_LEDS
            <= register_write_data_40_s(7 downto 0);    -- Board GPIO Leds
2794        when REG_CR_TH_UPDATE_CTRL              => register_map_control_s.CR_TH_UPDATE_CTRL
            <= "1";                                     -- See Central Router Doc
2795        when REG_CR_FH_UPDATE_CTRL              => register_map_control_s.CR_FH_UPDATE_CTRL
            <= "1";                                     -- See Central Router Doc
2796        when REG_FH_IC_PACKET_RDY               => register_map_control_s.FH_IC_PACKET_RDY
            <= register_write_data_40_s(23 downto 0);   -- Rising edge indicates the complete packet can be
            read
2797        when REG_TIMEOUT_CTRL                   => register_map_control_s.TIMEOUT_CTRL.ENABLE
            <= register_write_data_40_s(32 downto 32);  -- 1 enables the timout trailer generation for
            ToHost mode
2798                                                    register_map_control_s.TIMEOUT_CTRL.TIMEOUT
                                                        <= register_write_data_40_s(31 downto 0);   -- Number
```

```
1153        ---- ## GENERATED code BEGIN #4 ----
1154        ---------------------------------------
1155    {% for register in registers if register is in_group('Bar2') %}
1156    {% for bf in register.bitfield if bf.is_write or bf.
        is_trigger %}
1157    {% if loop.first %}
1158            when {{"%-34s"|format(register.prefix_name|
                prepend('REG_'))}} => register_map_control_s.{{
                "%-30s"|format(bf.dot_name)}} <= {{"%-40s"|format
                (bf|vhdl_value('register_write_data_40_s')|semi
                )}} {{bf.desc|vhdl_comment(153)}}
1159    {% else %}
1160                {{"%-37s"|format(" ")}}
                    register_map_control_s.{{"%-30s"|format(bf.
                    dot_name)}} <= {{"%-40s"|format(bf|
                    vhdl_value('register_write_data_40_s')|semi
                    )}} {{bf.desc|vhdl_comment(153)}}
1161    {% endif %}
1162    {% endfor %}
1163    {% endfor %}
1164        ---------------------------------------
1165        ---- GENERATED code END #4 ##  ----
1166        ---------------------------------------
```

FELIX software: data path

FELIX application:

Read blocks from file descriptors	→ DMA transfers and reads 1KByte blocks of data, encoded for the transfer over PCIe. Every block contains accumulated data from one E-link or FULL Mode link.
Block decoding	→ Decode into variable sized packets for transmission over network.
Statistics	→ Count processed blocks, transfer rates, etc.
Extract Meta Information	→ Meta-information, for example event ID, is extracted and matched against a routing table.
Route	→ Route.
Load Balancing, Error Handling	→ Distribute load among multiple systems, handle automatic failover in case of system failures.
Send to network	→ Asynchronous message service 'netio' for data exchange with network hosts.