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
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

Custom point-to-point links

Point-to-point S-links

~100 servers

~1500 servers

High-Level Trigger Farm

100 kHz / ~200 GB/s ~2000 links

Custom electronic components

PCs (COTS)

Frontend

ReadOut Driver

ReadOut System

Ethernet

HLTPU

HLTPU

HLTPU

HLTPU

HLTPU

Frans Schreuder

FELIX - TWEPP17 Santa Cruz
Upgrade for Phase-I

Custom electronic components including FELIX cards

PCs (COTS)

GBT links* or FULL mode links

PCs

40 Gb Ethernet, Infiniband

---

* GBT: GigaBit Transceiver with Versatile Link
Upgrade for HL-LHC

GBT, LpGBT* or FULL mode links

COTS network technology

~20,000 links

less than 10 TB/s

*LpGBT: Low power GBT

2024

2015 2017 2019 2021 2023 2025

Run 2 Run 3 Run 4

Custom electronic components including FELIX cards

PCs (COTS)
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 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

- **BNL-711 from BNL**
  - Xilinx Kintex Ultrascale XCKU115
  - 48 optical links (MiniPODs)
  - FELIX Phase-1 prototype
  - TTC input ADN2814
  - Si5345 jitter cleaner
  - PCIe Gen3 x16 (2x8 with bridge)
  - Version 2.0 currently tested

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

- **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

**Central Router**
- To-Host data decoding, data boundaries detection, header and packet trailer attachment
- From-Host data encoding, parsing and transmission.

**TTC**
- TTC fmc wrapper
- TTC Decoder
- TTC data fan-out

**Wupper**
- PCIe engine
- DMA engine
- DMA control
- DMA write
- XILINX PCIe End Point
- Interrupt controller

**Firmware clocks**
* Exact output frequency depends on the mmcm input. Outputs are phase aligned to an input clock.

**Housekeeping**
- GBT link data emulator
- GBT-GBT FPGA wrapper

**External TTC board**
- TTC clk
- mmcm

**Optical connectors**
- GBT link data emulator
- GBT link data emulator

**Configuration registers**
- Control and monitor

**Interrupt controller**
FELIX FULL mode firmware block diagram

- **CLK&RST**: mmcm
- **Wupper core**: DMA engine
- **TTC fmc wrapper**: TTC Decoder
- **Central Router**: From-Host data encoding, parsing and transmission.
- **GBT- FPGA wrapper**: GBT link data emulator
- **FM link data emulator**: to-Host path
- **External TTC board**: Busy
- **House-keeping**: Busy
- **GBT link data emulator**: from-Host path
- **Busy**: TTC data
- **Config**: TTC data fan-out
- **Configuration registers**: control and monitor.
- **From-Host data decoding, data boundaries detection, header and packet trailer attachment**: XILINX PCIe End Point
- **To-Host data decoding, data boundaries detection, header and packet trailer attachment**: Interrupt controller
- **TTC clk**: mmcm firmware clocks
- **Wupper PCIe engine**: DMA read
- **Board oscillator**: DMA control
- **Wupper PCIe engine**: DMA write
- **Optical connections**: GBT_NUM
- **Xilinx PCIe**: Interrupt controller

*Exact output frequency depends on the mmcm input. Outputs are phase aligned to an input clock.*
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

- 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

- 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
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

- FEASIC
- E-link
- GBTx
- FPGA
- Calibration
- DSC
- FE configuration
- Event readout
- COTS network switch
- Event readout
- Event readout
- Event readout
- Event readout

- Timing
- Busy

- up to 320 Mb/s
- link up to ~10 Gb/s
- network up to 40Gb/s or 100Gb/s

- FELIX data flow overview
Integration workshop setup

- KC-705VLDB
- MiniFELIX
- gFEX
- MiniFELIX FLX-711
- TTC
- HTG-710
- VLDB
- hw_server
- KC-705
- MiniFELIX
FELIX software: status update

- **Low-Level Software Tools**
  - FLX Card Drivers
  - The flx-tools Suite
  - The f-tools Suite
  - E-Link Configuration

- **Development and Debug**
  - pepo, fel, …

- **Test Software**
  - The flx-tools Suite
  - FLX Card API, flx-init, flx-config, flx-info, …
  - fec, fic, fupload, felink, …
  - elinkconfig

- **Production Software**
  - FELIX Core Application
  - FELIX TDAQ Integration
  - FELIX Monitoring
  - NetIO
  - FELIX Discovery
  - Test clients

- **FELIX Software Map**
  - felix-mon, felix-web
  - netio-cat, …
  - felix-client, felix-dcs, fatcat
Register map automation – Jinja 2
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
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.

Summary

- 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 ATLAS Sub-Detectors Front End developers.

Status:

- 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 system testing and integration: firmware to control of GBTxASIC via IC port.
- Extend the system testing and integration: firmware to control of GBTxASIC via EC port.
- A C++ API and an OPC server and client are in progress to fully support the GBTxASIC.

- Increase overall system reliability.
- Increase the number of input channels supported to 24 for GBT mode.
- Extend system testing and integration: firmware to control of GBTxASIC via IC port.
- Extend the system testing and integration: firmware to control of GBTxASIC via EC port.
- A C++ API and an OPC server and client are in progress to fully support the GBTxASIC.
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](https://github.com/simpway/GBT_KC705)
Integration test: Clock stability

- **Goal:** assess 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

```python
when REG_STATUS_LEDS => register_map_control_s.STATUS_LEDS
<= register_write_data_40_s(7 downto 0); -- Board GPIO Leds

when REG_CR_TH_UPDATE_CTRL => register_map_control_s.CR_TH_UPDATE_CTRL
<= "1"; -- See Central Router Doc

when REG_CR_FH_UPDATE_CTRL => register_map_control_s.CR_FH_UPDATE_CTRL
<= "1"; -- See Central Router Doc

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

when REG_TIMEOUT_CTRL => register_map_control_s.TIMEOUT_CTRL.ENABLE
<= register_write_data_40_s(32 downto 32); -- 1 enables the timeout trailer generation for ToHost mode

register_map_control_s.TIMEOUT_CTRL.TIMEOUT
<= register_write_data_40_s(31 downto 0); -- Number
```

--- # GENERATED code BEGIN #4 ---

```python
for register in registers if register.is_in_group('Bar2') %
  if register.is_write or register.is_trigger %
    if not register.is_first %
      when register.write_data(7 downto 0) =>
        \$StyleSheet\$[register.write_data(7 downto 0)]
    \%
```

--- # GENERATED code END #4 ---
FELIX application:

- 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.
- Decode into variable sized packets for transmission over network.
- Count processed blocks, transfer rates, etc.
- Meta-information, for example event ID, is extracted and matched against a routing table.
- Route.
- Distribute load among multiple systems, handle automatic failover in case of system failures.
- Asynchronous message service 'netio' for data exchange with network hosts.