FELIX: The new detector readout system for the ATLAS experiment

Soo Ryu, on behalf of the ATLAS TDAQ Collaboration
Argonne National Laborator, 9700 South Cass Avenue B109, Lemont, IL 60439, USA
E-mail: sryu@anl.gov

Abstract. After the Phase-I upgrades (2019) of the ATLAS experiment, the Front-End Link eXchange (FELIX) system will be the interface between the data acquisition system and the detector front-end and trigger electronics. FELIX will function as a router between custom serial links and a commodity switch network using standard technologies (Ethernet or Infiniband) to communicate with commercial data collecting and processing components. The system architecture of FELIX will be described and the status of the firmware implementation and hardware development currently in progress will be presented.

1. Introduction
The Large Hadron Collider (LHC) located at CERN, Geneva, Switzerland will cease operations for 18 months starting in 2019 to allow for upgrades to the accelerator and experiments. During this period the ATLAS experiment plans to upgrade several sub-systems. In the upgrade period the ATLAS DAQ system will be renovated with a new concept of bridging the streams of data between various sub-detector electronics and the endpoints of the data acquisition network. The new concept will be realized through the Front-End Link eXchange (FELIX) project, a data routing system that consists primarily of commercial off-the-shelf (COTS) components.

In the present LHC Run phase the ATLAS DAQ system is a tiered data processing system with custom and commodity hardware as depicted in Figure 1. Custom units, known as ReadOut Drivers (RODs), interface detector front-ends with the ReadOut System (ROS) [1] which consists of PCs housing custom and commodity components. Data are buffered at 40.08 MHz within detector front-end electronics and filtered by the first level accept signal sent by the Timing Trigger and Control (TTC) system [2] at 100 kHz. The front-end units then forward data streams to the ROD by means of custom point-to-point links. Upon data arrival, the ROD dispatches the data to the ROS using the S-Link [3] point-to-point optical link developed by CERN. The ROS buffers and transfers data on request to the High-Level Trigger (HLT) computing farms via a commodity network.

In the next phase of LHC operation starting in 2021 the FELIX system will be used to interface both the commodity network and the front-end units for the New Small Wheel muon detector [4] and the trigger readout of Liquid Argon (LAr) Calorimeter [5]. The radiation hard optical link developed by CERN, GBT [6], has been selected for the connection to the front-end units. Using industrial standard network protocols such as TCP/IP, Infiniband, or Omni-path the FELIX system will support connections to various COTS devices. The functions
Figure 1. ATLAS DAQ design in (a) LHC Run 2 (2015-2018) and (b) LHC Run 3 (2021-2023)

implemented in the ROD such as data aggregation, compression and buffering are moved into FELIX and the software implementation of ROD (SWROD).

The FELIX system will bring several advantages to the ATLAS DAQ system. Since the FELIX system will be implemented with commodity hardware the DAQ system reduces reliance on custom hardware. The FELIX system also expands the network architecture of the DAQ system, which is more scalable and easier to maintain.

The functional requirements of FELIX are:

- LHC global timing and trigger information from the TTC system should be forwarded by FELIX to the detector front-end units with a low, fixed latency.
- FELIX should provide 24 physical connections to the several front-end units which may accommodate multiple detector channels.
- Each physical link should be able to convey up to 40 data streams from different detector channels and each data stream should be distinguishable at the network endpoints.
- FELIX should support control, calibration and monitoring of sub-detectors.

In this paper we introduce the hardware platform for FELIX development in Section 2. The firmware design and updated features are discussed in Section 3. The status of the integration activities with the ATLAS front-end units is described in Section 4.

2. FELIX interface card

The FELIX concept has been demonstrated via a development platform based on a server PC with a PCIe Gen-3 interface card equipped with a high performance FPGA. Three PCIe card models have been considered. The card specifications are listed in Table 1. A custom mezzanine card, TTCfx [7], has been designed to interface to the data and clock received from the TTC
The TTCfx recovers the TTC data and LHC global clock coming through a fiber optic connector using circuitry exploiting the clock and data recovery chip ADN2814 [8] and a jitter cleaner chip Si5345 [9]. The recovered 40.08 MHz clock is used throughout the FPGA on the FELIX card. A LEMO connector on the TTCfx provides a signal to assert a busy signal to the ATLAS trigger system.

### Table 1. FELIX interface card candidates

| Model | FPGA       | PCIe Interfaces | # of links | manuf.  |
|-------|------------|-----------------|------------|---------|
| HTG-710 | XC7VX690T-2FFG1761C | Gen3 8 lanes   | 24 | HiTech-Global |
| VC-709 | XC7VX690T-2FFG1761C | Gen3 8 lanes   | 4   | Xilinx Inc.  |
| FLX-711 | XCKU115 FLVF1924 | Gen3 16 lanes  | 48  | custom       |

#### 2.1. FELIX prototype board

One of the candidates for the final design of FELIX card is referred to as FLX-711, which is derived from a test board for LAr Trigger Digitizer Boards (LTDBs [10]). The board is equipped with a 16-lane PCIe Gen3 interface and a Xilinx Kintex UltraScale FPGA with 64 GTH transceivers. As shown in the Figure 2 the FLX-711 has 8 MiniPODs which can host up to 48 bidirectional optical links. Since the Xilinx Integrated Block for PCIe is designed for 8 lanes, a PCIe switch PEX8732 is used to connect two 8-lane endpoints to the 16-lane PCIe slot. This guarantees an achievable bandwidth up to 128 Gb/s. FLX-711 incorporates the same circuitry and connectors used in the TTCfx, which improves stability of the reference clock for GBT link. A software control and management via System Management Bus and automatic FPGA programming with an on card flash memory have been implemented. Several tests were completed for PCIe DMA throughput and optical link speed. The total measured throughput using this double 8-lane PCIe design was measured at up to 101.7 Gb/s. The optical link speed was tested at 12.8 Gb/s during an IBERT [11] test. More details for the tests on this board can be found in Ref. [12].

#### 3. Firmware design

The FELIX firmware consists of four major components as shown in Figure 3: GBT wrapper, Central Router, PCIe Direct Memory Access (DMA) engine and TTC decoder.

The GBT wrapper module consists of a Xilinx GTH hard block and GBT logic block derived from the CERN GBT-FPGA [13] design. The GBT interface block receives 120 bits of data
frame at 40.08 MHz, however the actual user bandwidth is 3.2 Gb/s since the payload length is 80 bits. The remaining 40 bits are reserved for the forward error correction, a frame identifier and the slow control for GBTx ASIC [14] and GBT-SCA [15]. The data frames are encoded/decoded in the GBT logic block and split into 20 bits at 240.48 MHz for input to the GTH interface. To ensure fixed latency performance the reference clocks for the GBT wrapper design are phase-synchronous to the clock recovered from the TTCfx. In addition, crossing clock domains within the FPGA on the path to/from the transceivers is not allowed. The GBT wrapper, like all SERDES units in modern FPGAs, is based upon 8b10b encoding with FIFO buffers. To obtain lower latency the GBT wrapper bypasses the internal FIFO and manually controls the gearbox and frame alignment portions of the SERDES block of the FPGA. For the optimization more details can be found in Ref. [16]

The Central Router routes the data packets between the GBT interface and the PCIe DMA engine with respect to an e-link configuration selected by user. E-link is a logical collection of bits in which the data packet is serialized. The possible e-link data widths are 2, 4, 8 and 16 bits. A 120 bit wide data frame from the GBT interface is made up of e-links and each e-link is connected to a dedicated 2 kB FIFO. The Central Router has dedicated data managers on both sides to handle the data depending on its direction. Data received from each GBT link is read by a data manager interfacing with the GBT wrapper in the 120 bit wide data frame at 40.08 MHz. The e-link processor (e-link proc) stacks data packets in the 2 kB FIFO at 40.08 MHz until it is filled to 1 kB. A trailer which carries the definition of the packet is added when the packets are written to the FIFO. Under the control of the data manager interfacing with the PCIe engine the data in a 256 bit-wide buffer are multiplexed to another FIFO with the same width connected to the PCIe interface. The Central Router passes the trigger information from the TTC decoder to the GBT interface. The Central Router identifies BUSY conditions for each e-link as they arrive and sends a bit-map of BUSY states to the TTC decoder block.

The PCIe DMA engine, Wupper [17], provides the data path between the Central Router and the Xilinx PCIe Gen3 hard block. Wupper provides a FIFO that has the same width (256 bits) as the Xilinx AXI4-Stream interface and runs at 250 MHz, thus the maximum throughput is 64 Gb/s. Wupper handles a set of DMA descriptors, with an address, a read/write flag, the transfer size and an enable line, which are mapped as normal PCIe memory or IO registers. Transactions to/from the server PC are controlled by looking at the address and transfer direction defined in the DMA descriptors. A status register in each DMA descriptor is used by software to detect the pending or processed requests.

A TTC decoder receives the serial bit stream and 160.32 MHz of clock out of an ADN2814 in the TTCfx. The logic decodes the serial bit stream into two interleaved data streams at 40.08 MHz, each of which carries the first level accept and the serialized commands. The decoding of

Figure 3. FELIX firmware modules and connectivity
the command is done after the deserializer. The first level accept and deserialized commands are forwarded to the GBT wrapper via the Central Router in a 10 bit-wide data bus that is time-multiplexed as required for the width of each e-link. The status of the internal counters to mark the bunch crossing and event number are passed down to the host PC via the Central Router. The logic also aggregates the inhibit signals from multiple channels of the front-end units and asserts a BUSY signal to the entire TTC network through the LEMO output on the TTCfx.

3.1. Clocking and system latency
FELIX is required to transfer global timing and trigger information to the sub-detector systems with fixed, reproducible latency. This goal has been achieved by making the clocks driving the functional blocks in the FPGA phase synchronous to the 40.08 MHz of the TTC clock. The clock scheme of FELIX is depicted in Figure 4. The phase of the 160.32 MHz clock obtained from the ADN2814 on the TTCfx is aligned to that of the 40.08 MHz TTC clock. The TTC decoder receives TTC data and the 160.32 MHz input clock directly from the ADN2814 and aligns the phase of the decoded data to the phase of the 40.08 MHz clock which is derived from the input clock. Then several clocks are synthesized with the clock of the TTC decoder with equal phase: 40.08 MHz, 80.16 MHz, 160.32 MHz and 320.64 MHz. These clocks are used to drive the Central Router and the GBT logic block.

The 240.48 MHz reference clock of the GTH hard block is also derived from the TTC clock. The GTH hard block has dedicated clock and data recovery circuitry which recovers the clock used in the transmitter of the other GTH hard block. Using the circuitry in the GTH the 240.48 MHz clock is recovered at the GTH receiver on the front-end unit with equal phase and frequency, which makes the FELIX and the detector front-end units exist in the same clock domain.

The phase noise in the reference clock of the GTH hard block had been a critical issue for firmware development. The TTC clock has phase noise which interferes with stable data recovery in the GTH hard block. In order to reduce the phase noise we prepared a dedicated clock route incorporated with a jitter cleaner, Si5345.

The latency of the functional blocks has been measured and is shown in Table 2. The latency
Table 2. Latency of FELIX functional blocks for TTC data, measured using the existing Level-1 trigger system.

| Segment                              | Latency (ns) | Total   |
|--------------------------------------|--------------|---------|
| ADN2814 & TTC decoder                | ~100         |         |
| Central Router & Fan-out             | 100          | 260-281 |
| GBT TX & GTH TX                      | adjustable (60-81 with 6 steps), here: ~68 |         |
| FrontEnds GTX RX & GBT RX            | ~53          | ~53     |

Figure 5. A diagram for FULL mode implementation

measured with each segment is constant in time. The overall latency of FELIX can be selected between 260 ns and 281 ns.

3.2. Firmware implementation for higher throughput

Due to the requirement of more than twice of the GBT link bandwidth from LAr and Tile calorimeter front-end units [18], a new transmission mode was implemented with a GTH hard block and a simpler protocol using 8b10b encoding, which is called the “FULL mode” as depicted in Figure 5. Note that FULL mode is independent of the GBT standard and is not phase synchronous with the TTC clock. The line rate of the FULL mode link is 9.6 Gb/s, but due to the overhead of 8b10b encoding the actual user bandwidth is 7.68 Gb/s. The GTH transmitter at the front-end unit and receiver in FELIX are operated at 240 MHz × 32-bits and the data packet is serialized accordingly. The packet is identified and assembled by detecting the packet alignment symbols in the Central Router. The utilization of FULL mode requires a mechanism to control the data flow to cope with the data congestion between Central Router and PCIe DMA engine. The data flow will be managed by forwarding a dedicated symbol to the front-end units. The specifics of the mechanism are to be determined. As an example of a possible implementation, if the link in the direction toward the front-end unit is using GBT mode one of the K-characters in 8b10b encoding will be used as the flow control symbol and sent on the e-link to the front-end units. The performance tests of FULL mode are in progress alongside the implementation for each hardware platform.

3.3. Limiting factors in the FPGA implementation

Several limitations have been found during firmware development. The utilization rate of FPGA resources is proportional to the number of GBT channels and the most of LUTs are occupied by the Central Router. For instance the full FELIX implementation with 4 GBT links consumes
34% of LUTs in the Virtex7 690T which has 433K LUTs. We estimated 28% and 5% of LUTs are occupied by the Central Router and the GBT wrapper, respectively. Based on the 4-channel results, extrapolation to 8 channels is at the conventional 75% utilization limit for LUTs and builds greater than 8 channels will clearly be impossible due to the lack of resources. In order to resolve this we introduced the Xilinx ultrascale FPGA which accommodates 663K LUTs as a main chipset of the FLX-711. Efforts to reduce the resource requirements of the Central Router continue in parallel. The PCIe Gen3 throughput is another limitation in this development. For FULL mode, the total bandwidth needed for 24 links is 184.3 Gb/s which is above the physical speed limit of PCIe Gen3 with 16 lanes (126 Gb/s). However, the design will be scaled to maximise the number of links which can be serviced given current constraints.

4. Integration tests with front-end units
For the upcoming ATLAS Phase-I upgrade (2019) integration tests with detector front-end units have been in progress since the second quarter of 2016. There are two tests with the front-end units, one with the LAr Trigger Digitizer Board (LTDB) and the other with the Control and Readout ITK Board (CaRIBOu) [19]. The LTDB is a readout board to digitize the input analog signals from the detectors and transmit the digitized data to FELIX. The test is done with a prototype of LTDB which has five GBTx and five GBT-SCA chips. The objectives of this test are: 1) demonstrate the ability to transmit compatible data between GBTx on the LTDB and the GBT wrapper in FELIX, and 2) LAr detector control and monitoring data exchange using GBT-SCA. During the test LTDB was able to recover the phase synchronous TTC clock and trigger information with fixed latency. With the recovered clock LTDB was also able to receive control commands and send board status information to FELIX using GBT-SCA.

CaRIBOu is a modular test system for pixel detector sensor R&D studies for ATLAS Phase-II upgrade (2023). After the Phase-II upgrade FELIX is planned to interface all detector front-end units. The purposes of the test are to demonstrate the ability to transmit compatible data between the GBT-FPGA in CaRIBou and the GBT wrapper in FELIX and verify the data integrity during the long period of time. The test setup is shown in Figure 6, which consists of a control and readout (CaR) board, several front-end chip carrier boards and a FELIX interface board (ZC706). The trigger signals encoded in a customized data format are sent to the Local Trigger Interface (LTI) board through an ethernet connection. LTI emulates the TTC system and generates several reset signal types as well as the Level-1 accept signals. LTI transmits the trigger and reset signals to the FELIX card via GBT link so that FELIX is synchronized to the LTI clock and starts data taking from the ZC706 board through the GBT link. FELIX showed no bit errors over the test duration of 12 hours.

5. Summary and prospects
As a part of the ATLAS DAQ upgrade in 2019 the FELIX system has been developed as a part of a new DAQ architecture concept. FELIX is a PC based data routing system with a FPGA based PCIe card at its core. With several hardware platforms being evaluated the FELIX firmware and software development has matured sufficiently to support testing with prototype ATLAS detector front-end units. To meet the high throughput requirement from several detector systems, a new data transmission mode, FULL mode, has been implemented. A new hardware platform FLX-711 was manufactured and tested to fully satisfy the designed performance requirements. Integration tests evaluating the implementation and performance with two detector front-end units have been completed. Continuing tests with additional front-end units are planned in the upcoming years. FELIX development is on track for a successful deployment in 2019.
Current ZC706 firmware supports to interface one DUT.

System clock & TTC commands are from LTI emulator.

For this test platform, an Ethernet cable connects one RJ45 from HSIO to LTI emulator. The emulator extracts clock and commands from it.

This makes FELIX & CaRIBOu system to be synchronized with the telescope readout.

VHDCI-FMC adapter boards are not shown.

Telescope CaRIBOu Back-End VHDCI cable 3 ~50m

Figure 6. FELIX - CaRIBOu test setup. FELIX served as a back-end of the LTI Emulator and the CaRIBOu system.

Reference
[1] Crone G et al 2010 The ATLAS ReadOut System - Performance with first data and perspective for the future Nucl. Instr. Meth. Phys. Res. A. 623 534-536
[2] Timing, Trigger and Control Systems for LHC Detectors, LEB Status Report/RD12 Status Report on the RD-12 Project, http://ttc.web.cern.ch/TTC/RD12status99.pdf
[3] Bij E et al 1997 S-LINK, a data link interface specification for the LHC era IEEE Trans. Nucl. Sci. 44 398-402
[4] Coliban R M et al 2016 The Read Out Controller for the ATLAS New Small Wheel JINST 11 C02069
[5] Schwinhorst R et al 2016 The Phase-1 upgrade of the ATLAS first level calorimeter trigger JINST 11 C0108
[6] Moreira P et al 2007 The GBT, a Proposed Architecture for Multi-Gb/s Data Transmission in High Energy Physics TWEPP2007 3-7 September 2007, pg. 332.
[7] Levinson L TTCfx a TTC FMC card for FELIX https://espace.cern.ch/ATLAS-FELIX/FE_support/Documentation/FELIX_TTC_FMC-V3.pdf
[8] Analog Devices Inc. ADN2814: Continuous Rate 10 Mb/s to 675 Mb/s Clock and Data Recovery IC with Integrated Limiting Amp http://www.analog.com/static/imported-files/data_sheets/ADN2814.pdf
[9] Silicon Labs Inc. Si5345/44/42 Rev D Data Sheet - 10 Channel, Any-Frequency, Any-Output Jitter Attenuator/ Clock Multiplier http://www.silabs.com/documents/public/data-sheets/Si5345-44-42-D-DataSheet.pdf
[10] ATLAS Collaboration 2013 ATLAS Liquid Argon Calorimeter Phase-I Upgrade Technical Design Report CERN-LHCC-2013-017; ATLAS-TDR-022; http://cds.cern.ch/record/1602330
[11] XILINX Inc. IBERT for UltraScale GTH Transceivers v1.3 http://www.xilinx.com/support/documentation/ip_documentation/ibert_ultrascale_gth/v1_3/pg173-ibert-ultrascale-gth.pdf
[12] Anderson J et al 2016 FELIX: a PCIe based high-throughput approach for interfacing front-end and trigger electronics in the ATLAS Upgrade framework JINST 11
[13] Marin M B et al 2015 The GBT-FPGA core: features and challenges JINST 10 C03021
[14] CERN GBT Project: GBTX Manual https://espace.cern.ch/GBT-Project/GBTX/Manuals/gbtxManual.pdf
[15] Caratelli A et al 2015 The GBT-SCA, a radiation tolerant ASIC for detector control and monitoring applications in HEP experiments JINST 10 C03034.
[16] Chen K et al 2016 Optimization on fixed low latency implementation of GBT protocol in FPGA arXiv:1608.08916v3
[17] PCIe Gen3x8 DMA for virtex7 http://www.opencores.org/project/virtex7_pcie_dma
[18] Gomes A et al 2016 The new front-end electronics for the ATLAS Tile Calorimeter Phase 2 Upgrade JINST 11 C02015.
[19] Liu H et al 2016 Development of a modular test system for the silicon sensor R&D of the ATLAS Upgrade https://arxiv.org/abs/1603.07950