FELIX: first operational experience with the new ATLAS readout system and perspectives for HL-LHC

Joaquin Hoya
on behalf of the ATLAS TDAQ Collaboration
The Large Hadron Collider (LHC) collides proton bunches at a **40MHz** rate. 
- ATLAS detects the collision products and selects (trigger) physics events of interest. 
The Run 3 expected avg. event data rate for permanent storage is ~3 kHz.
- New detector and trigger systems installed for Run 3 to improve background rejection.

### New in Run 3

**Muon System**
- New Small Wheels (NSW)
- Inner Barrel RPCs (BIS7/8)

**Calorimeters**
- Liquid Argon (LAr) digital readout

**Trigger and DAQ**
- L1Calorimeter Trigger (L1Calo)
- FELIX & Software Readout Driver (SWROD)

---

**Pileup** = number of interactions per LHC bunch crossing

---

J. Hoya - 11/05/23

---
FELIX and ATLAS TDAQ in Run 3 (2022-2025)

**FELIX: Front-End Link eXchange** ([https://atlas-project-felix.web.cern.ch](https://atlas-project-felix.web.cern.ch))

**Run 3:**
- Same as Run 2 for most sub-detectors.
- Legacy ROD and ROS architecture is being replaced with FELIX and SW ROD. It includes NSW, LAr, L1Calo and BIS78.

**FELIX** is a **data router** that works as an interface between on-detector systems and commodity computing.
- The data being routed includes readout, configuration, trigger, clock distribution, monitoring.
- **FELIX** system consists of commodity servers with PCIe cards. Used for **data routing** only.
- **SWROD** is in charge of **data processing**, aggregation, and monitoring. Hosted by commodity computers.

The introduction of FELIX brings down the number of custom components in the system, reducing design and maintenance efforts. **COTS** earlier in the readout chain.

**GBT**: synchronous serial protocol at 4.8 Gb/s
**FM**: 8b/10b RX link at 9.6 Gb/s (FULL Mode)
FELIX Hardware

The FLX-712 FELIX card

- FPGA Xilinx Kintex UltraScale XCKU115, 16-lane PCIe Gen3.
- 8 MiniPODs to support up to 48 bidirectional optical links (most commonly: 4 MiniPODs/24 links).
- Interface to Timing, Trigger and Control (TTC) systems. BUSY output.
- Flash memory to store firmware.

~300 boards produced, for ATLAS, ProtoDUNE, ATLAS tracker upgrade, and others.

FELIX Firmware

Two main flavours:

**FULL**: interface to other FPGA-base systems
- Up to 24 channels per FLX-712, 9.6 Gb/s each

**GBT**: interface to GBTX
- GBTX is a radiation-hard ASIC [1]
- On-detector data stream aggregator
- Supports 24 x 4.8 Gb/s bi-directional GBT links
- Each GBT link carries multiple data streams (E-links) of configurable bandwidth

---

FELIX Software

Readout application

Felix-star transfer data between the FLX-712 card and network peers
- Interrupt driven central event loop architecture.
- Asynchronous non-blocking architecture.
- Single thread, two processes per card.
- Two data transfer approaches: zero-copy, data coalescence.
- Custom network library based on libfabric [1].
- Uses Remote Direct Memory Access (RDMA) technology for low overhead transfers.

felix-star runs as daemon on FELIX servers
- Each FELIX server hosts up to two FLX-712 cards

FELIX server:
- Intel Xeon E5-1660 v4 @ 3.2GHz.
- 32 GB DDR4 2667 MHz memory.
- Mellanox Connect-X 25/100 GbE).

Network events:
1. Send completed
2. Data received
3. Buffer available for sending

System events:
1. Timer events (timer fd)
2. Signals (event fd)
3. Any file descriptor event

FLX-712 events:
1. data available
2. busy state

Run 3 software architecture scalable for Run 4.
FELIX Performance in ATLAS

64 FELIX PCs, 105 FLX-712 cards installed in ATLAS in 2022.

- Application control and monitoring based on Supervisor [1]
  - automatically start and restarts felix-star applications, and can be monitored and controlled via a web interface.

- Monitoring integrated in the ATLAS infrastructure:
  - Integration into the ATLAS-wide ErrorReporting System (ERS) [4].

---

[2] doi: 10.1051/epjconf/202024501020
[3] https://go2.grafana.com
FELIX Performance: LAr
LDPB (LAr Digital Processing Blade) -- FELIX in FULL mode

- FELIX design max throughput 128 Gb/s
- Network card 100 Gb/s
- LDPB is the system with largest Throughput!
- Stable performance at ~100 kHz.
- Message size that can be up to 22 kB.
  - Only system where a true zero-copy approach is used (no data coalescence).

- FELIX software does not copy messages in network buffers.
- Network cards sends messages directly from their fragments in the FELIX DMA buffer.
**FELIX Performance: L1Calo**

Level-1 Calorimeter trigger -- FELIX in FULL mode

FELIX (in full mode) being used by:
- gFEX (GlobalFeatureExtractor)
- eFEX (ElectronFeatureExtractor)
- jFEX (Jet Feature Extractor)
- TREX (Tile Rear Extension)

Plots for gFEX:
- Throughput ~8Gb/s in a high rate run.
- Stable performance at ~100 kHz.
- L1Calo uses a feature called "streams":
  - 16 links per FLX-712 using up to 9 streams, each carrying data at 100 kHz.
- Avg. message size: 3kB
- L1Calo uses buffered mode.

Throughput ~8 Gb/s
Rate ~96 kHz
FELIX Performance: NSW
New Small Wheel -- FELIX in GBT mode

- The NSW has the largest number of E-links
  - ~200 per FELIX card
- Each E-link providing data at ~100 kHz.
- Avg. message size: 40B

One major challenge in SW during last year was the late packet arrival:

- All messages were delivered but with a latency up to 100ms (could exceed SWROD time window)
- The leading cause was CPU saturation, reaching 100% at high rate (>80 kHz)
- Performance optimizations deployed since earlier this year led to messages delivered on time!
HL-LHC $t\bar{t}$ event in ATLAS ITK at $<\mu>=200$
ATLAS DAQ in Run 4

2029+

Run 4 conditions
- 1 MHz L1 trigger rate \(\rightarrow \times 10 \text{ Run 3}\)
- Up to 200 avg. interactions per bunch-crossing \(\rightarrow \times 3 \text{ Run 3}\)
- 4.6 TB/s data throughput \(\rightarrow \times 20-30 \text{ Run 3}\)

FELIX requirements
- Readout of all sub-detectors
- ~14000 optical links with bandwidth up to 25 Gb/s
- support for new detector-specific functionalities

Data handler – evolution of SWROD, under development.
Future FELIX cards

Prototypes, firmware and software upgrades

A new FELIX card is necessary to support
● increased maximum link bandwidth (10 → 25 Gb/s).
● new timing/trigger interface (will receive data at 9.6 Gb/s).

Prototypes
● FLX-181 and FLX-182 prototypes.
● Xilinx XC(V)M1802 FPGA up to 24 links 25 Gb/s.
● new FPGAs, 4+ generation PCIe, new optical transceivers (FireFly™).

Firmware Upgrades to support
● Additional data encoding types.
● Higher link and PCIe interface speed.
● Larger buffers in computer memory.

Software Upgrades
● Same architecture as in Run 3 but different deployment scheme:
  ● Run 3: only 2 readout applications per card.
  ● Run 4: up to 8 readout applications.

FLX-181
FLX-182
Integration with new systems

**Inner Tracker (ITk)**

New all-silicon inner tracker
- Increased acceptance up to $|\eta|<4$ and pile-up rejection.
- Comparable/better tracking performance at much higher pile-up conditions (~200).

FELIX is being used in the ITk **Pixel** and **Strips** production and testing.

**ITk Strips**
FELIX Strips firmware functional
Configuration and readout via FELIX.

**ITk Pixel**
FELIX Pixel firmware successfully tested.

![ITk Pixel demonstrator @ SR1, CERN](image)

New flavours in addition to GBT and FULL mode:
- lpGBT (evolution of GBT)
- PIXEL & STRIP (custom lpGBT)
- Interlaken (64b/67b encoding)
Summary

- FELIX is a data acquisition component for the ATLAS experiment to interface detector electronics and commodity computing.
- Run 3:
  - FELIX was used instead of the legacy readout architecture for new sub-detector systems, reducing the amount of custom hardware in the data taking path.
  - FELIX firmware and the software are mature and used for data taking.
  - Good performance for all the new systems (NSW, LAr and L1Calo).
- Run 4:
  - FELIX will readout all sub-detectors.
  - Hardware prototypes under development.
  - Firmware under development. Early builds successfully tested using Run 3 hardware.
  - Run 3 software architecture scalable for Run 4.
  - FELIX is already part of the early production and testing of some of the new Run 4 detectors.
Backup Slides
FELIX Performance

L1Accept vs Data rates

LDPB

NSW

L1Calo

Data rate

L1A rate

Data rate
FELIX: Front-End LInk eXchange

- FELIX is a **data router** that works as an interface between on-detector systems and commodity computing.
- ATLAS-wide effort to harmonize detector readout systems.
- Designed to cope with the expected higher data volumes and event processing complexity.
- The data being routed includes readout, configuration, trigger, clock distribution, monitoring, BUSY and TTC signals.
- The firmware is modular and flexible, with a routing module between the custom serial links and PCIe interface.
- The software includes drivers, low-level tools, test software and routing software.
- First-generation FELIX cards are in use during Run 3.

*TTC refers to the Trigger, Timing and Control systems*
FELIX Firmware

Two main flavours:
- **FULL**: to interface the FELIX to other FPGA-base systems
  - Up to 24 channels per FLX-712, 9.6 Gb/s each
- **GBT**: to interface to GBTX
  - GBTX is a radiation-hard ASIC developed at CERN [1]
  - Used as on-detector data stream aggregator
  - GBT firmware supports 24 x 4.8 Gb/s bi-directional GBT links
  - Each GBT link carries multiple data streams (e-links) of configurable bandwidth

<table>
<thead>
<tr>
<th>Mode</th>
<th>Message size</th>
<th>Rate per link</th>
<th>(e)links per card</th>
<th>Total message rate per card</th>
<th>Total data rate per card</th>
<th>Use case</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>4800 bytes</td>
<td>100 kHz</td>
<td>12</td>
<td>1.2 MHz</td>
<td>46 Gbps</td>
<td>LAr</td>
</tr>
<tr>
<td>GBT</td>
<td>40 bytes</td>
<td>100 kHz</td>
<td>192</td>
<td>19.2 MHz</td>
<td>7.5 Gbps</td>
<td>NSW</td>
</tr>
</tbody>
</table>

FELIX Software
Felix-star architecture
FELIX Software

Readout application

- server publishes links/e-links, clients subscribe
- two data transfer approaches: zero-copy, data coalescence

- user-friendly API hides the complexity of network library for client applications

API functions

- subscribe(elink_number)
- unsubscribe(elink_number)
- send_data(elink_number)

API callback hooks

- on_message_received(elink_number)
- on_connection_established(elink_number)
- on_disconnection(elink_number)
FELIX users in ATLAS Run 3

Muon Spectrometer [GBT mode]
- New Small Wheels (NSW)
  - sTGC (Small-strip Thin Gap Chamber)
  - MicroMegas (Micro Mesh Gaseous Structure)
- BIS78 (Barrel Inner Small MDT sector 7/8)

L1 calorimeter trigger [FULL mode]
- gFEX (Global Feature Extractor)
- jFEX (Jet Feature Extractor)
- TREX (Tile Rear Extension)
- ROD, Hub for eFEX (Electron Feature Extractor)

Liquid Argon Calorimeter [48-ch GBT / FULL mode]
- LTDB (LAr Trigger Digitizer Board, custom GBT)
- LDPB (LAr Digital Processing Board, FULL)

Tile Calorimeter test system [FULL mode]
Buffer slices not present in FULL mode.
Performance on FELIX testbed at CERN

Data acquisition at 1 MHz trigger rate

- 32 links, 260 Mbps each.
- 1 MHz random trigger rate.
- Stable transfer rate, no errors at ×10 design trigger rate!

![Diagram showing data acquisition setup and performance metrics](image-url)