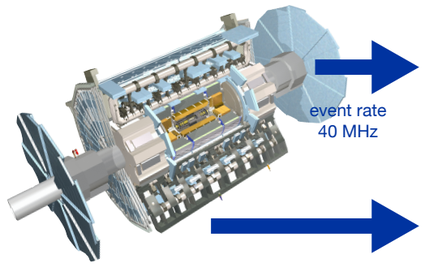
23rd IEEE Real Time Conference, August 01-05 2022 (virtual)

FELIX: the new ATLAS readout system from Run 3 to High Luminosity LHC

Felix Zahn - on behalf of the ATLAS TDAQ Collaboration

ATLAS trigger and data acquisition system (DAQ)



event rate 40 MHz

Level-1 (L1) Trigger

Accept signal
max latency 2.5 µs
max rate 100 kHz

Data acquisition (DAQ) system

100 kHz

High-Level Trigger (HLT)

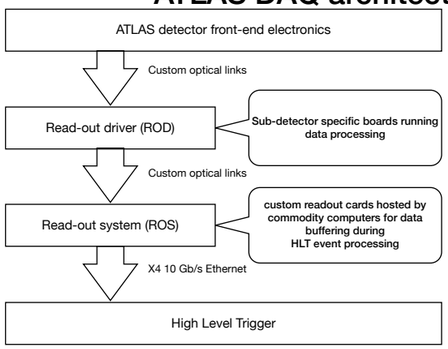
3 kHz

Permanent Storage

- Large Hadron Collider collides protons at a 40 MHz rate
- the maximum event rate for physics to permanent storage is 3 kHz
- two-level trigger to select events

ATLAS RUN 3
LARGE HADRON COLLIDER
est. 2022 at 13.6 TeV

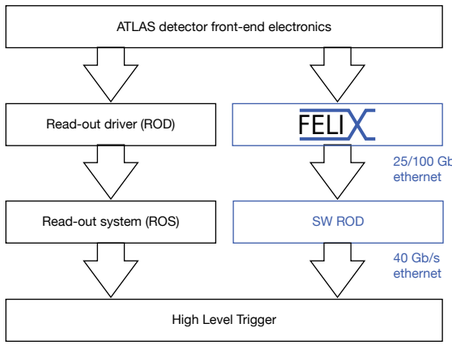
ATLAS DAQ architecture for Run 2



ROD for muon drift tubes
P. Jansweijer, NIKHEF

ROS readout card (RobinNP)
[arXiv 1710.05607]

ATLAS DAQ architecture for Run 3



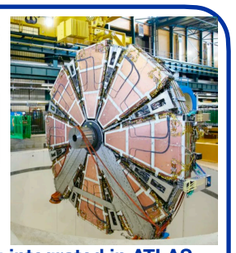
FELIX
- PCIe cards hosted by commodity computers
- only custom component in new architecture
- data routing, no processing

SW ROD
- software in charge of data processing and aggregation, monitoring
- hosted by commodity computers

FELIX operations in ATLAS

Installation and function

- Currently 64 FELIX PCs, 105 FLX-712 cards installed in ATLAS
- FELIX used for readout, control/monitoring and clock & trigger routing for NSW, L1 Calo, LAr, and the Barrel RPC upgrade



[CERN tweet]

Monitoring integrated in ATLAS infrastructure

- operational monitoring [3] with Grafana [4] dashboard
- logs in Error Reporting System [5]
- conditions of PC, FLX-712, and network recorded by computer cluster monitoring

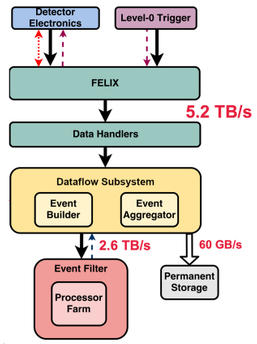
ATLAS DAQ in Run 4

Run 4 conditions

- 1 MHz L1 trigger rate (×10 Run 3)
- up to 200 interactions per bunch-crossing (×3 Run 3)
- 4.6 TB/s data throughput (×20-30 Run 3)

FELIX requirements

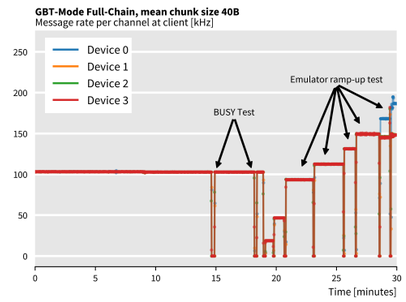
- taking over readout of all sub-detectors
- ~14000 optical links with bandwidth [2.5 or 25] Gb/s
- Interface to new Data Handler
  - replacement of SW ROD in Run 4
- support for new detector-specific functionality
  - e.g. continuous “trickle” reconfiguration of new tracker

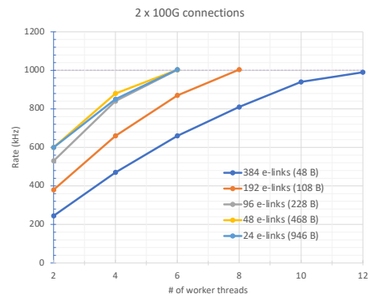


FELIX upgrades

- **Hardware**: new FPGAs, PCIe Gen 4+, new optical transceivers, increased max. link bandwidth (10 → 25 Gbps), new timing/trigger interface
- **Firmware**: to support: additional data encoding types, higher link and PCIe interface speed, more buffers in computer memory
- **Software**: Same architecture as in Run 3 but more buffers and more software instances allow decoupling of different data types or traffic flows

Performance from Run 3 to Run 4



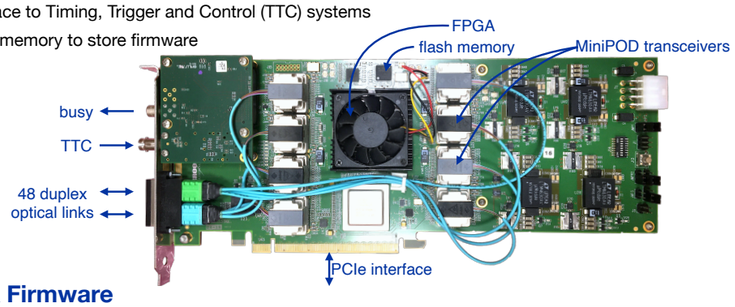


- Theoretically with ×4 buffers (150 → 600 kHz & 120 → 480 kHz trigger rate)
- 1 MHz already reachable e.g. 24 FULL links with 192 byte messages
- Ongoing tests with newer PCs successfully satisfy Run 4 throughput requirements.

FELIX - the new read-out system for Run 3

FLX-712 FELIX card (~300 boards produced, for ATLAS, ProtoDUNE, ITK, etc.)

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



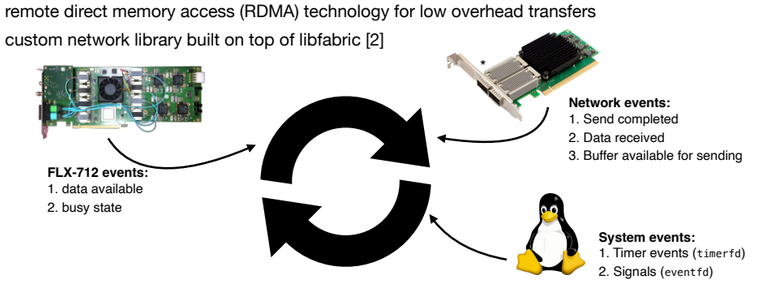
FELIX Firmware

- FULL: 24 links/card, 9.6 Gb/s each in host direction
- 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

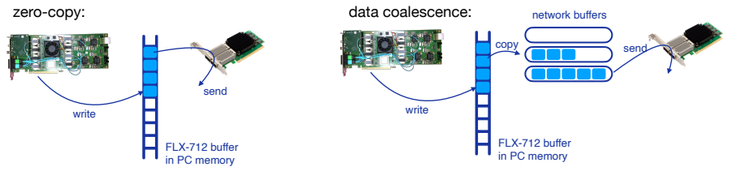
Mode	Msg size	Rate/ link	(e)links/ card	Total msg rate/card	Total data rate/card	Use case
FULL	4800 b	100 kHz	12	1.2 MHz	46 Gb/s	LAr calo
GBT	40 b	100 kHz	192	19.2 MHz	7.5 Gb/s	NSW

FELIX software

- interrupt-driven event-loop architecture / asynchronous non-blocking operations
- remote direct memory access (RDMA) technology for low overhead transfers
- custom network library built on top of libfabric [2]



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



[1] doi: 10.5170/CERN-2009-006.342
* Nvidia/Mellanox ConnectX-5 [image by storagereview.com]
[2] https://ofiwg.github.io/libfabric/
[3] doi: 10.1051/epjconf/202024501020
[4] https://go2.grafana.com
[5] doi: 10.1088/1742-6596/608/1/012004