Argonne NATIONAL LABORATORY

ATLAS DAQ in Run 3

DAQ@LHC, April 12-14, 2016

U.S. DEPARTMENT OF ENERGY

Introduction

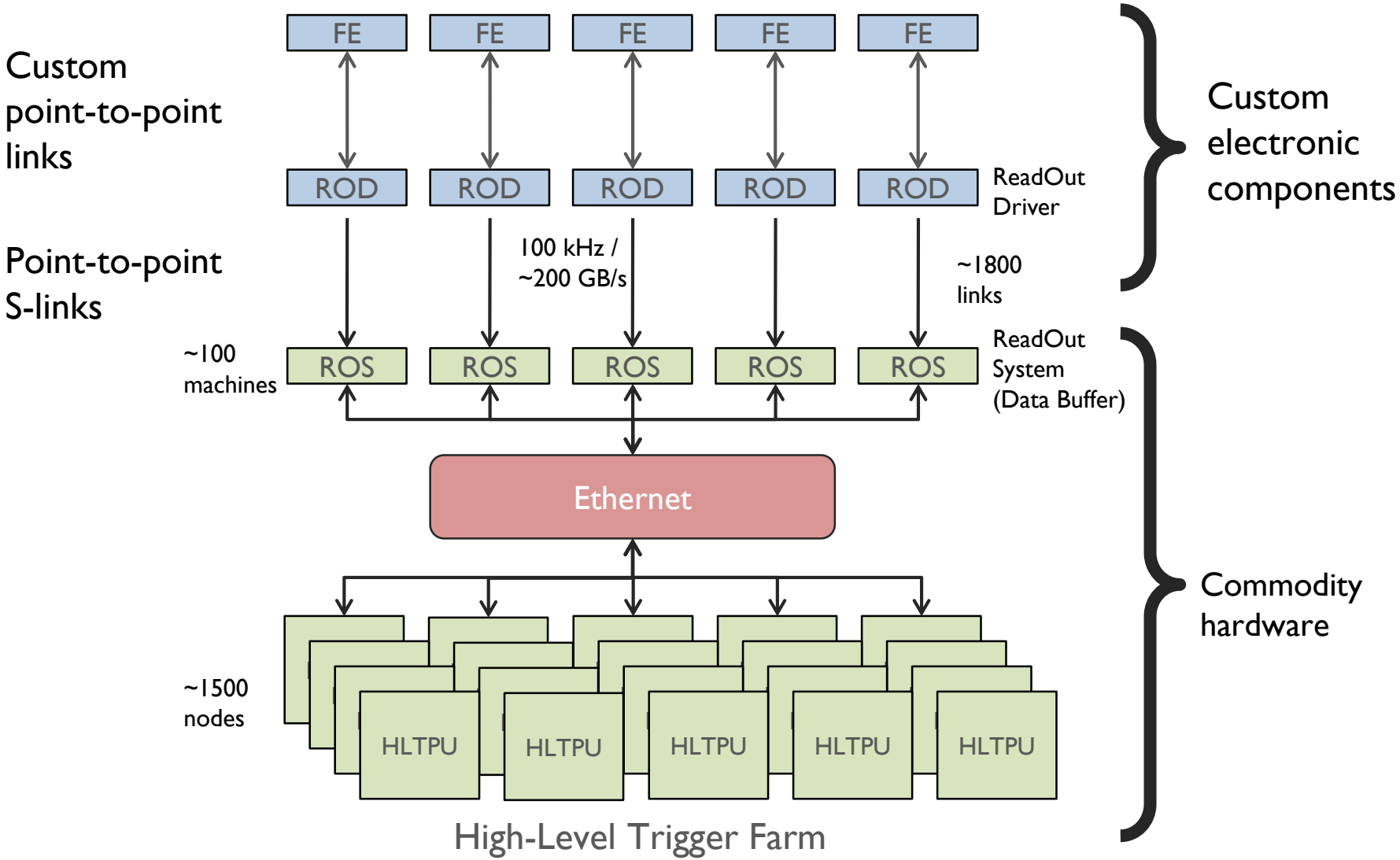
- ATLAS DAQ system evolution
- Front-End LInk eXchange (FELIX)
- Software ReadOut Driver (ROD)
- Summary

ATLAS TDAQ Operating Parameters

	# Trigger levels	Rates (kHz)		Event Size (MB)	Network Bandwidth (GB/s)	Storage	
						GB/s	kHz
Run 1	3	L1 (L2+EF)	75 ~0.4	~1	10	0.5	~0.4
Run 2	2	L1 HLT	100 1	~2	50	1	1
Run 3	2	L1 HLT	100 1	~2	50	1	1
Run 4	3	L0 L1 HLT	1000 400 10	~5	2000	~30	10

- **Evolutional changes from Run 2 to Run 3**
  - **Rolling replacement, Software, network, Readout System (ROS) if needed, … …**

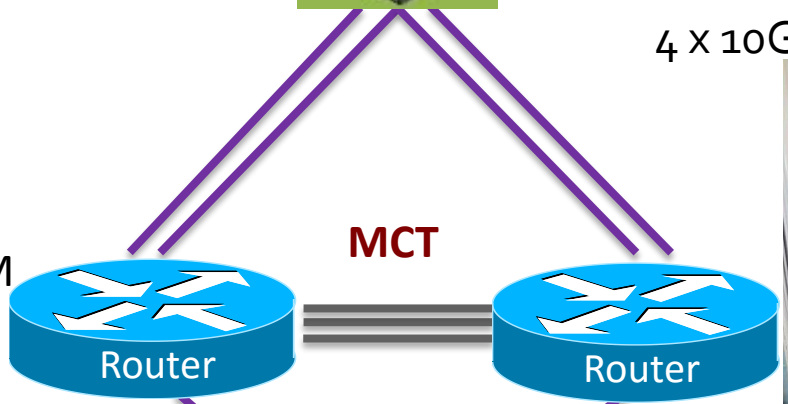
DAQ (Run 2)


Custom point-to-point links
Point-to-point S-links
FE
FE
FE
FE
FE
ROD
ROD
ROD
ROD
ROD
ReadOut Driver
100 kHz / ~200 GB/s
~1800 links
Custom electronic components
~100 machines
ROS
ROS
ROS
ROS
ROS
ReadOut System (Data Buffer)
Ethernet
~1500 nodes
HLTPU
HLTPU
HLTPU
HLTPU
HLTPU
Commodity hardware
High-Level Trigger Farm

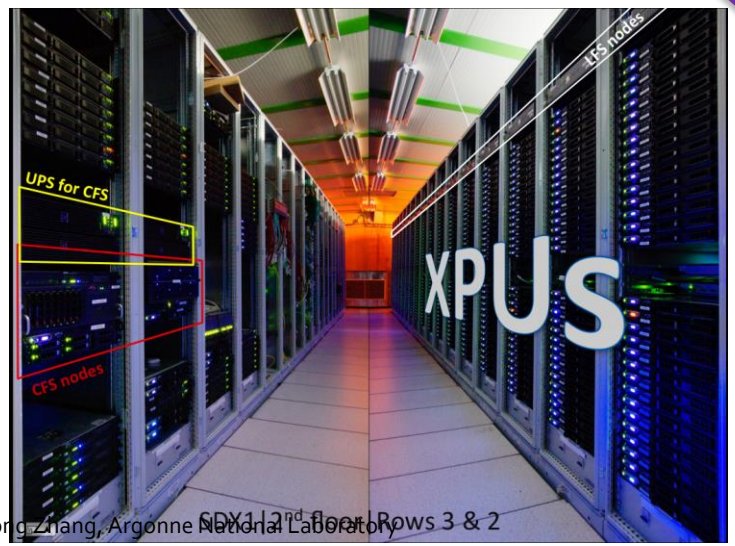
Hardware Components

ROBINNP (ALICE C-RORC)
Xilinx V6 LX240T FPGA
3 QSFP+ transceivers
2 x 4 GB DDR3 1066 MHz RAM
PCIe 8 lane Gen2 interface

ReadOut System (ROS)
~100 PCs, each with
2 ROBINNP
4 x 10GBE



Network (1 GBE, 10GBE)
With
Redundancy &
Multi Chassis Trunking
(Brocade)



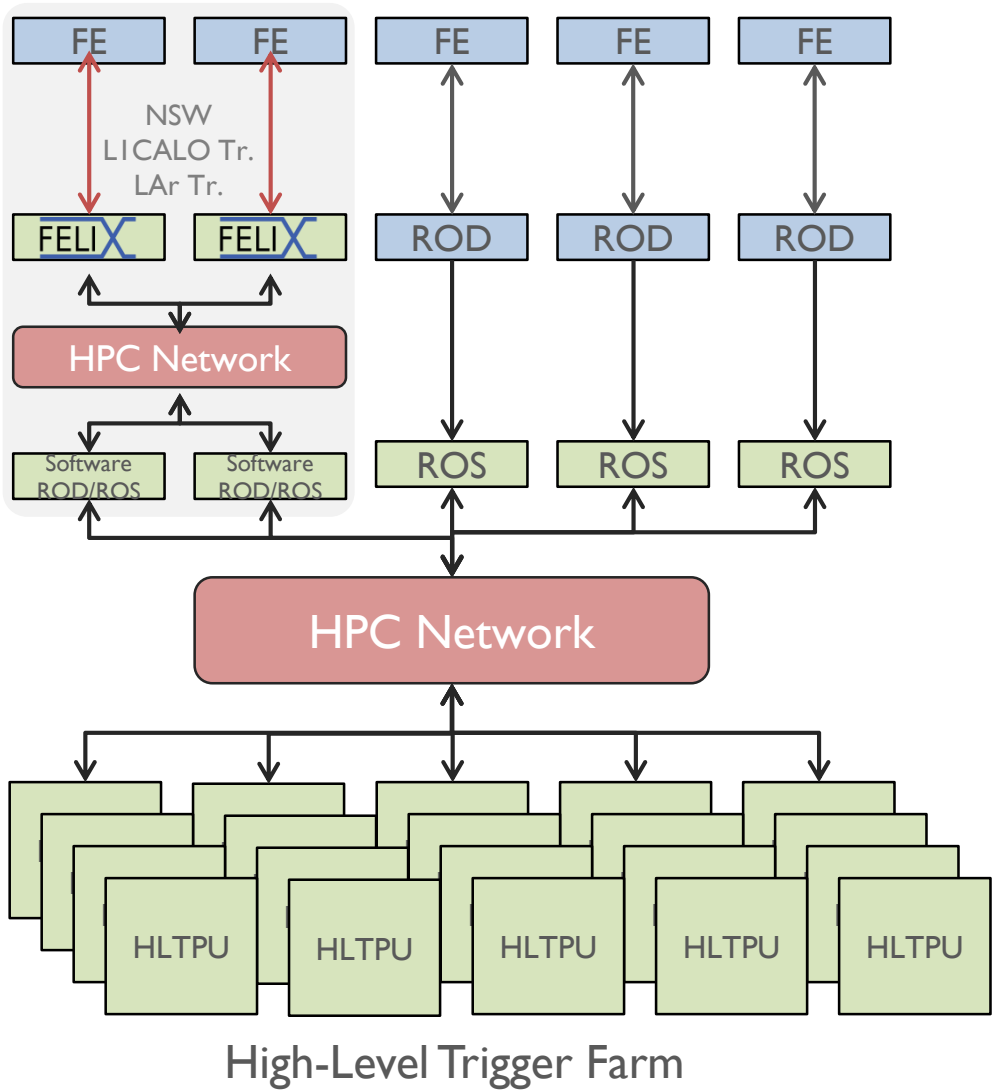
HLT Farm

~2000 Multi-core PCs
with rolling replacement

DAQ (Run 3)

Point-to-point
(GBT and additions)

40 Gb Ethernet



Custom electronic components

Commodity hardware

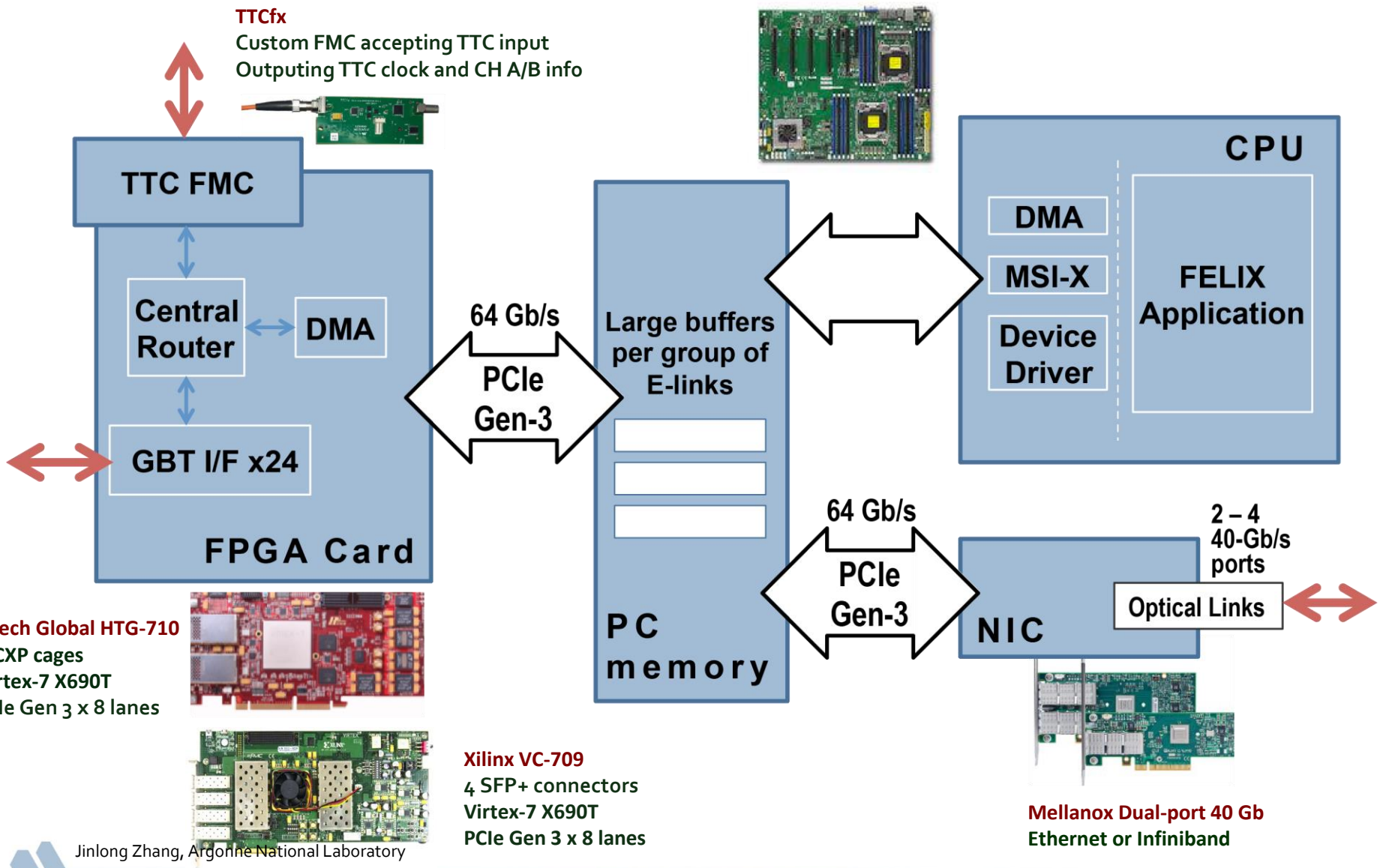
FELIX in Brief

- Router between serial/synchronous links (GBT and other protocols) and high level network links (40 GBE, InfiniBand, etc)
- Detector-agnostic and encapsulating common functionality (Merges and/or splits data streams but leaves content untouched)
- Handling detector configuration and control of calibration procedures
- Ensuring connectivity to the detector FE even in case of unavailability of other components (critical for DCS)
- TTC/BUSY handling for current and Phase-II TTC system

FELIX in Run 3

Link from detectors	~1000
FELIX I/O card	~50
FELIX PC servers	~25
FELIX NICs (40 GBE)	~50

Development



Status

- Firmware
  - GBT implementation and additions
  - TTC interface and TTC/BUSY handling
  - Central router
  - PCIe data transfer
- Software
  - Driver and API
  - Core software application
  - Configuration and testing tools
- Overall System
  - Full chain functionality demonstrated
  - Multi channel design being tested

A Hardware Candidate

- Xilinx Kintex Ultrascale FPGA
  - 64 GTH transceivers (16.375 Gb/s)
- 4 MiniPOD TX and 4 MiniPOD RX
  - 48 bidirectional optical links (14 Gb/s)
- PCIe Gen3 x16 lanes
- Two DDR4 SODIMM up to 16GB
- Onboard: Clock conditioner, TTC receiver, BUSY lemo output

FELIX Network

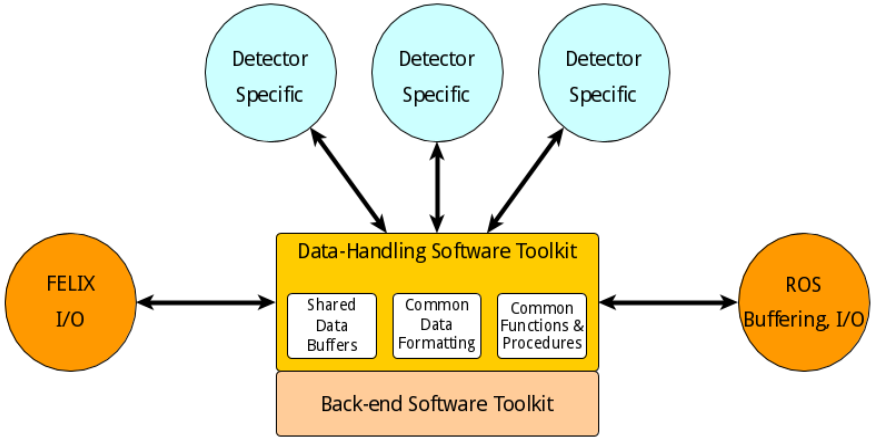
- Multi gigabit network
  - O(100) ports at least at 40 GBE
- Technology choices (to implement a communication library)
  - 40 GBE or 100 GBE
  - FDR or EDR Infiniband
  - Intel Omnipath
- Very different usages of the same physical network
  - Different priorities → Need to find an effective way to do QoS
  - Different traffic patterns → Need to understand them and dimension accordingly
  - Critical traffic to transport (DCS, control) → Need to build a robust network

Network R&D

- Exploring how Data Centre Bridging Ethernet can be beneficial
  - Priority Flow Control and Congestion Notification
  - Towards lossless Ethernet
- Server based networking
  - Packet forwarding performed by the server's CPU
  - Close to unlimited buffers hosted in the server's memory
- Software Defined Networking
  - Trying to identify potential use cases (Network fail-over, dynamically assigning bandwidth, …)
- DAQ network simulation
  - Working framework based on PowerDEVS tool
  - Well stablished methodology for modelling and simulation

Software ROD

- Interface to FELIX
  - Bridge boundaries
  - Upstream transport (e.g. configuration)
  - Downstream transport (e.g. bulk dataflow)
- Heterogeneous ROS implementation
  - Retain the uniform interface between ROS and HLT, and the ROS buffering capability
- Detector specific functionality
  - Provide a platform for detector customized configuration, fragment building and processing code



Status

- Exploratory design with
  - Inter-process communication (zeroMQ, interaction with stateless domain, etc)
  - Fragment Building Thread model
- Testing with
  - Dummy FELIX application and generic fragment objects
  - Realistic FELIX application based on NetIO
- Next step
  - Completion and testing of FELIX interface
  - Detector specific implementation

Summary

- **ATLAS DAQ system stays the same from Run 2 to Run 3 except**
- **FELIX (and related) to be introduced as new readout system for some detector systems**