ATLAS DAQ in Run 3
Introduction

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

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

- **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

~100 machines

100 kHz / ~200 GB/s

~1800 links

Ethernet

High-Level Trigger Farm

~1500 nodes

Custom electronic components

Commodity hardware

José Zhang, Argonne National Laboratory

04/13/2016
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)

~2000 Multi-core PCs with rolling replacement

Jinlong Zhang, Argonne National Laboratory
DAQ (Run 3)

Point-to-point
(GBT and additions)

40 Gb Ethernet

Custom electronic components

Commodity hardware

High-Level Trigger Farm

Jinlong Zhang, Argonne National Laboratory

04/13/2016
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

<table>
<thead>
<tr>
<th>FELIX in Run 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link from detectors</td>
</tr>
<tr>
<td>FELIX I/O card</td>
</tr>
<tr>
<td>FELIX PC servers</td>
</tr>
<tr>
<td>FELIX NICs (40 GBE)</td>
</tr>
<tr>
<td>~1000</td>
</tr>
<tr>
<td>~50</td>
</tr>
<tr>
<td>~25</td>
</tr>
<tr>
<td>~50</td>
</tr>
</tbody>
</table>

Jinlong Zhang, Argonne National Laboratory
TTCfx
Custom FMC accepting TTC input
Outputing TTC clock and CH A/B info

Hitech Global HTG-710
2 CXP cages
Virtex-7 X690T
PCIe Gen 3 x 8 lanes

Jinlong Zhang, Argonne National Laboratory

Xilinx VC-709
4 SFP+ connectors
Virtex-7 X690T
PCIe Gen 3 x 8 lanes

Mellanox Dual-port 40 Gb
Ethernet or Infiniband

Jinlong Zhang, Argonne National Laboratory

Development

TTC FMC

Central Router

GBT I/F x24

FPGA Card

64 Gb/s

PCIe Gen-3

Large buffers per group of E-links

64 Gb/s

PCIe Gen-3

PC memory

DMA

FELIX Application

CPU

DMA

MSI-X

Device Driver

NIC

Optical Links

2 – 4 40-Gb/s ports

Mellanox Dual-port 40 Gb
Ethernet or Infiniband
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 established 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

Jinlong Zhang, Argonne National Laboratory
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