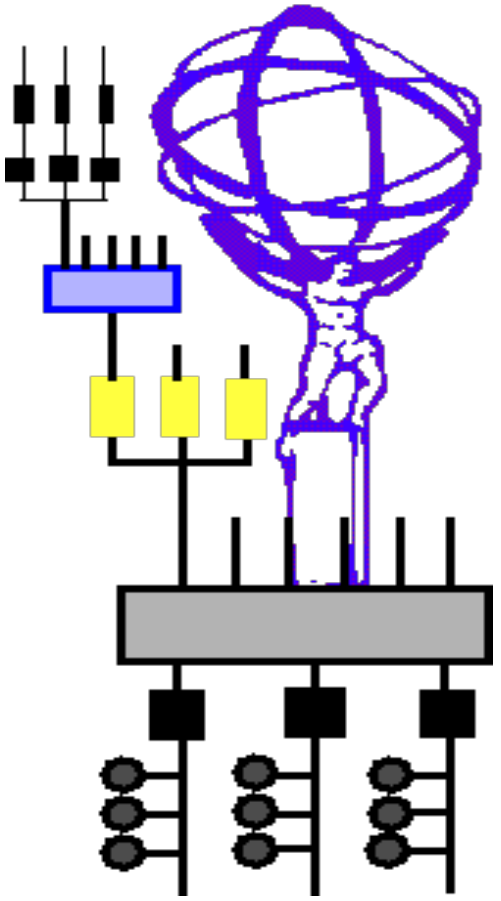
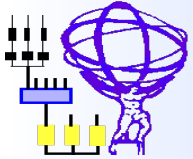


ATLAS DAQ/HLT Architecture for Phase II

W.Vandelli – CERN Physics Department/ATD



Outline

- ATLAS DAQ/HLT Architecture Evolution from Run 1 to Run 4
- New Detector Readout Architecture
- FELIX Project
- Projected HLT Farm Size

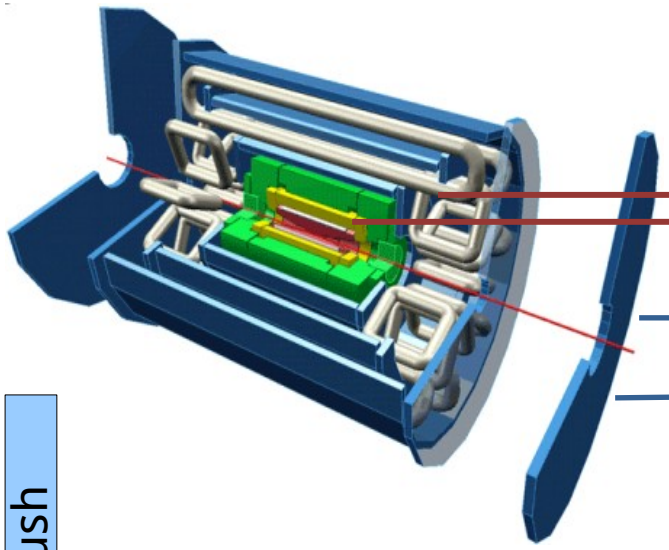
DAQ/HLT Upgrade Plans

- Driving Parameters

	# of Trigger levels	Level-xRate (kHz)		Event Size (MB)	Network BW (GB/s)	Storage GB/s	Storage kHz
Run 1	3	Lvl-1 HLT	75 ~0.4	~1	10	0.5	~0.4
Run 2	2	Lvl-1 HLT	100 1	~2	50	1	1
Run 3	2	Lvl-1 HLT	100 1	~2	50	1	1
Run 4	3	Lvl-1 HLT	400 10	~5	2000	25	10

- Notes
  - Storage bandwidth includes compression factors
  - Run 3: No major upgrades in the DAQ systems currently foreseen
  - **Run 4: Two stage hardware trigger** (Level-0 and Level-1)

ATLAS Run 1



Muon detector data

Calorimeter detector data

20 MHz

Level-1 trigger

O(2 μs)

FE Electronics

75 kHz

1574 Readout Links

Push

Pull

Detector Readout

O(50 ms)

~20 x 10 GE links

300 x 1 GE links

Level-2

Multi-layered data network

6 kHz

~40 x 1 GE links

<400 Hz>

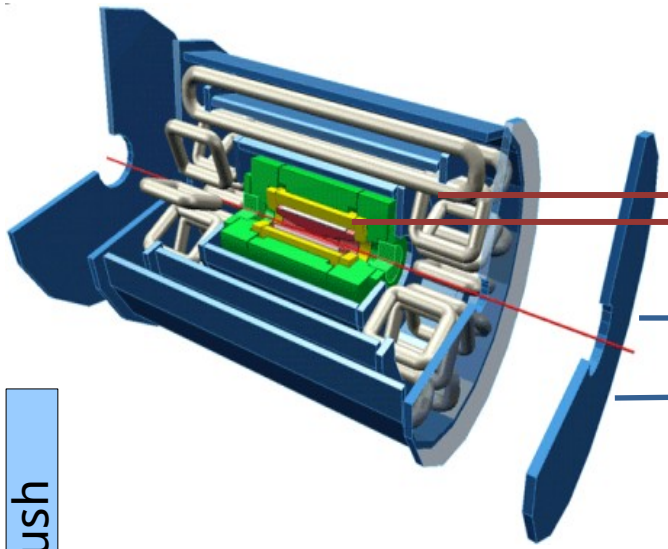
O(s)

Full Event building

Event Filter

Mass storage

ATLAS Run 2



Muon detector data

Calorimeter detector data

40 MHz

Level-1 trigger

O(2 μs)

FE Electronics

100 kHz

2000 Readout Links

Push

Pull

Detector Readout

400 x 10 GE links

RoI-based trigger-driven data-collection

Multi-layered data network

<1 kHz>

25 kHz equivalent

70 x 10 GE links

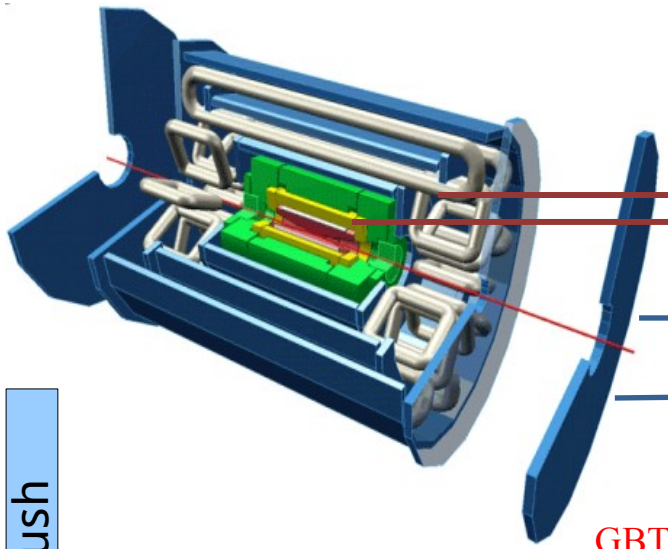
O(200 ms)

RoI based Event building

High-Level Trigger (Combined L2 & EF)

Mass storage

ATLAS Run 3



Muon detector data

Calorimeter detector data

40 MHz

Level-1 trigger

O(2.5 μs)

FE Electronics

100 kHz

GBT Links for new Detectors

2000 Readout Links

Detector Readout

400 x 10 GE links

Multi-layered data network

<1 kHz>

25 kHz equivalent

70 x 10 GE links

O(200 ms)

RoI based Event building

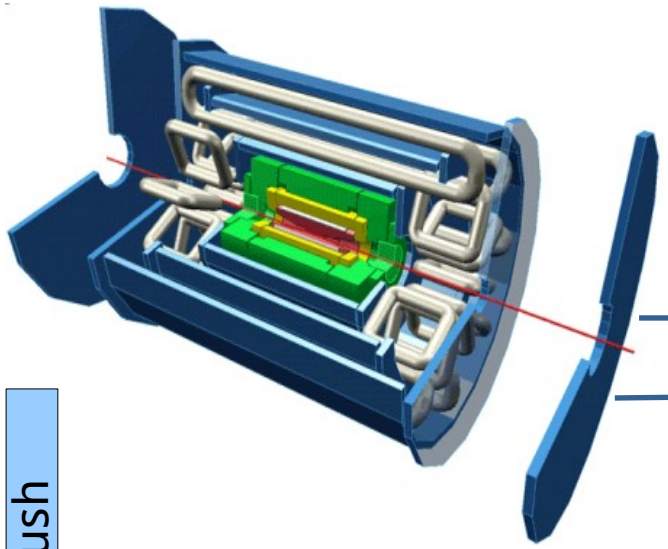
High Level Trigger

Mass storage

Push

Pull

ATLAS Run 4



Push

Pull

FE Electronics

GBT

8000 Readout Links

Level-1 trigger

O(30 μs)

400 kHz

Detector Readout

Multi-layered data network

10 kHz

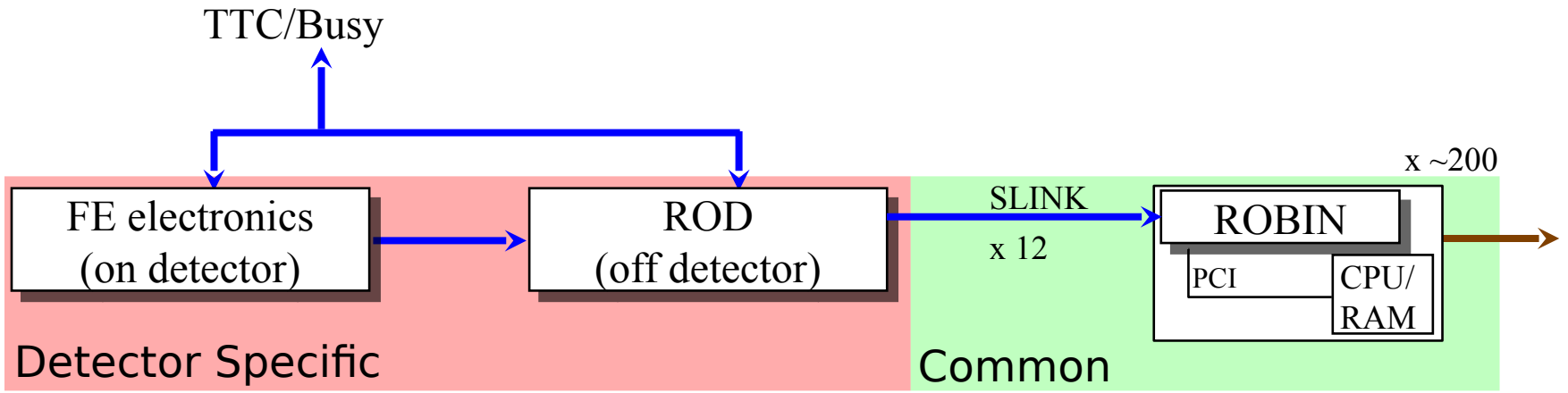
O(s)

RoI based Event building

High Level Trigger

Mass storage

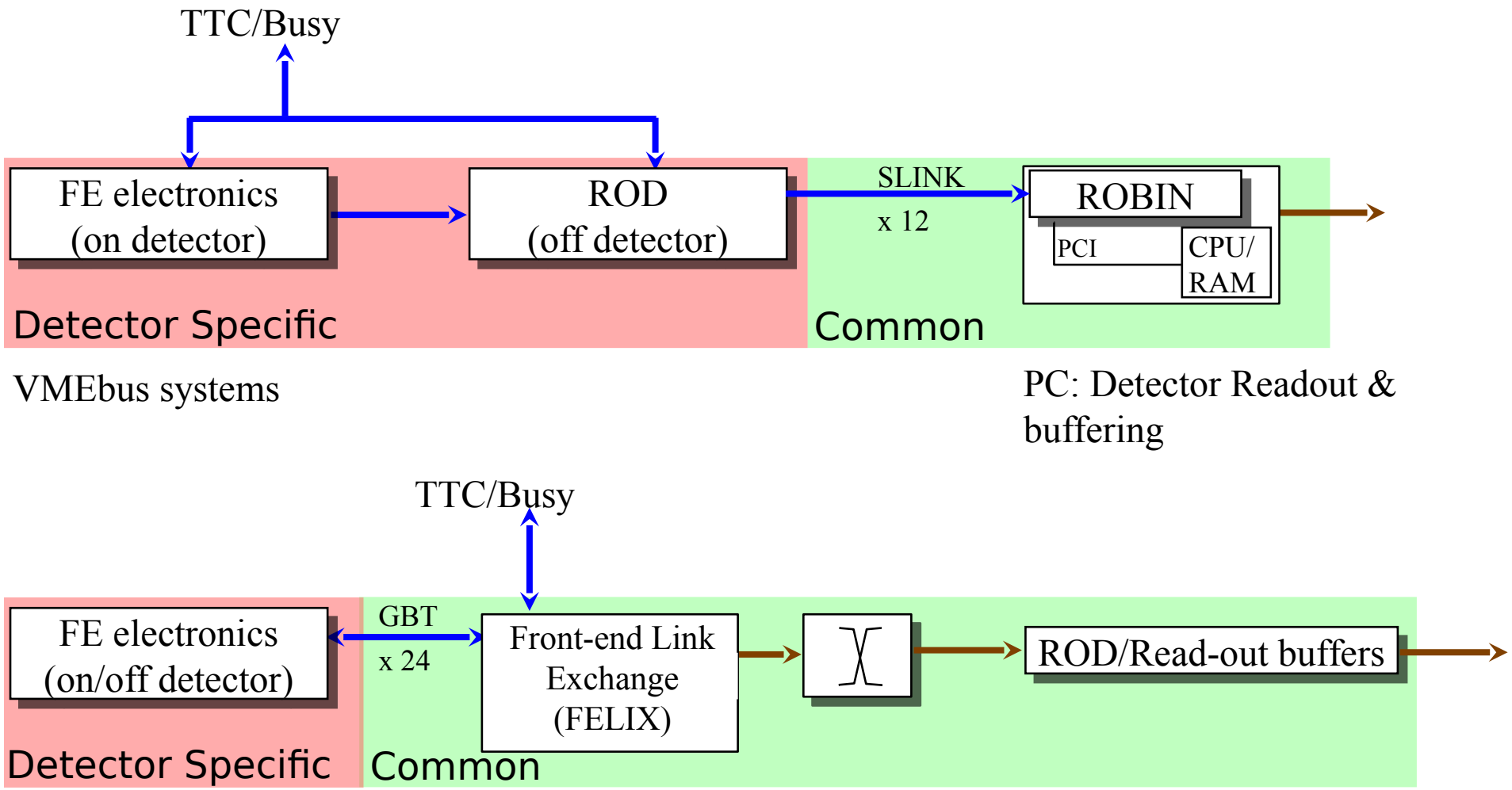
Detector Readout Run 1-2



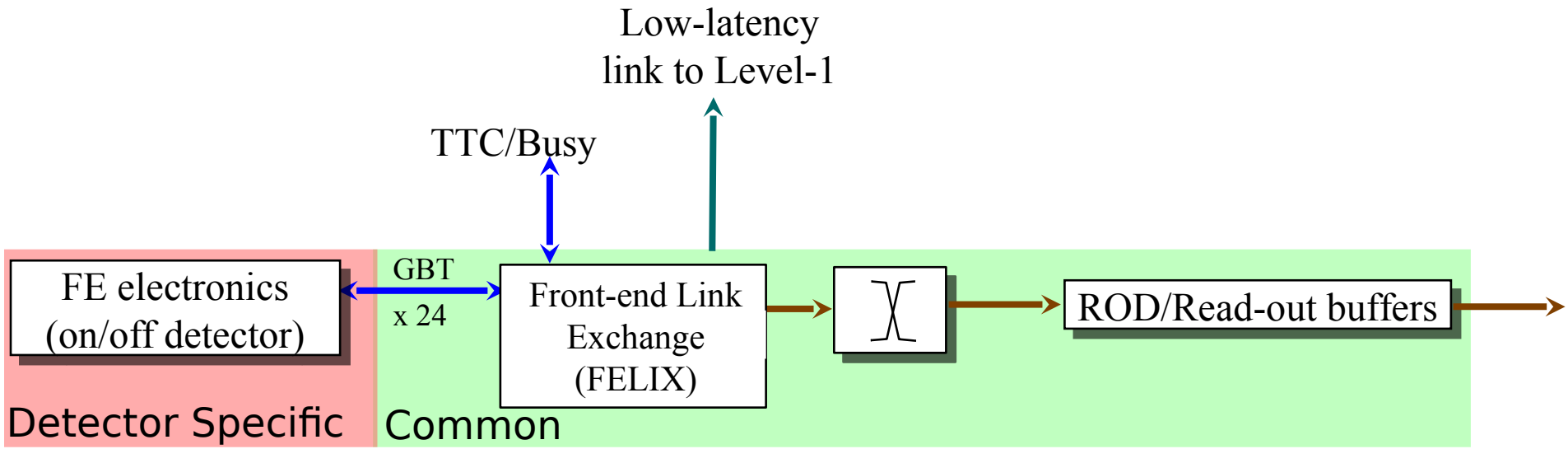
VMEbus systems

PC: Detector Readout & buffering

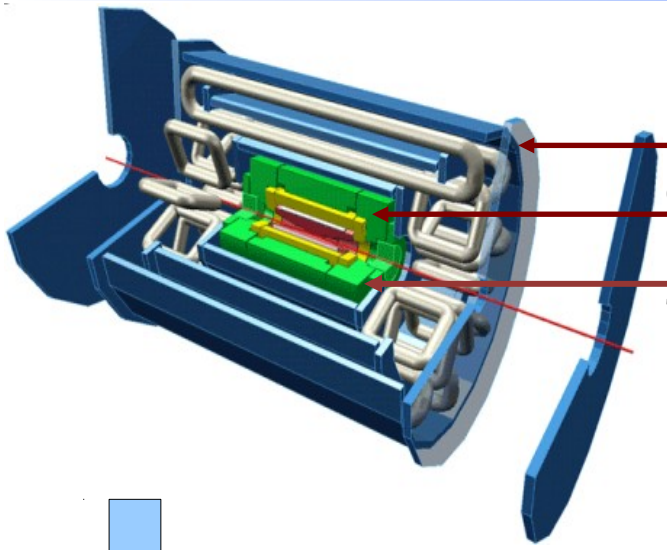
Detector Readout Run 3



Detector Readout Run 4



Architecture in Run 4



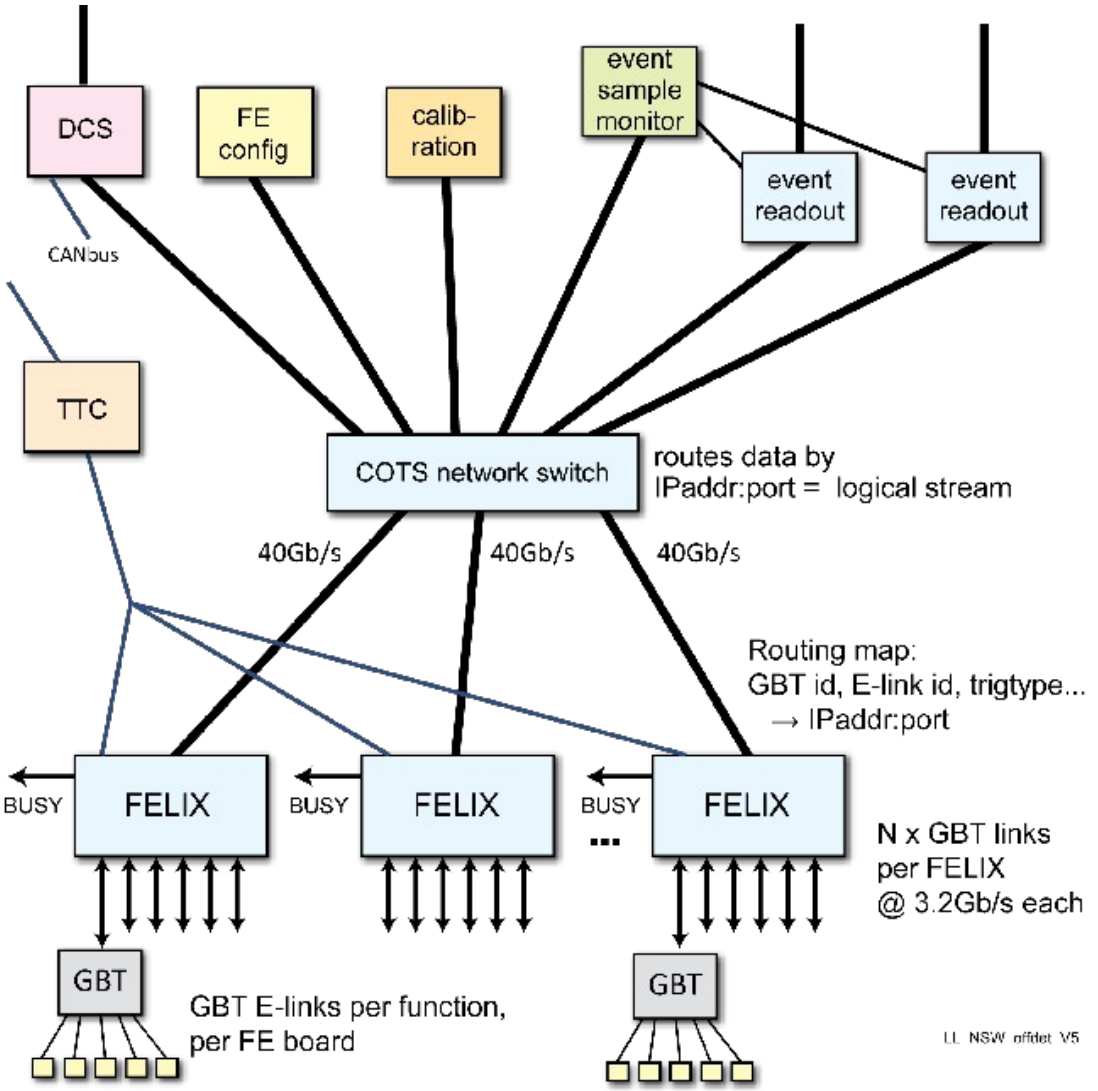
Muon
Calorimetry
Tracking
~8k GBT
FE Electronics
Level-1 trigger
O(30 μs)
Low-latency link to Level-1
~200
FELIX
L0: 1000 kHz
L1: 400 kHz
Push
Monitoring, Control, Configuration, DCS
200 x 200 Gbps links
~ N
"ROD"/Readout
Multi-layered data network
10 kHz
Pull
High Level Trigger
O(s)
Mass storage

Run 4 Readout: Advantages

- Higher level of commonality between detectors
  - FELIX is a common object providing functionality today implemented in detector-specific back-end custom elx (ROD)
- Performance-scalability built-in
  - programmable FE-DAQ connectivity
  - e.g. able to de-populate DAQ buffers whether needed to accommodate higher requirements, like increasing L1 rate
- Increased use of COTS components
  - all ROD-like functionality (including data processing) could most likely be implemented in standard computers by Phase-II
- Capability to disentangle ROD-like functions from H/W implementation
  - different granularity for monitoring, control, data handling ...
  - DCS and DAQ traffic separation

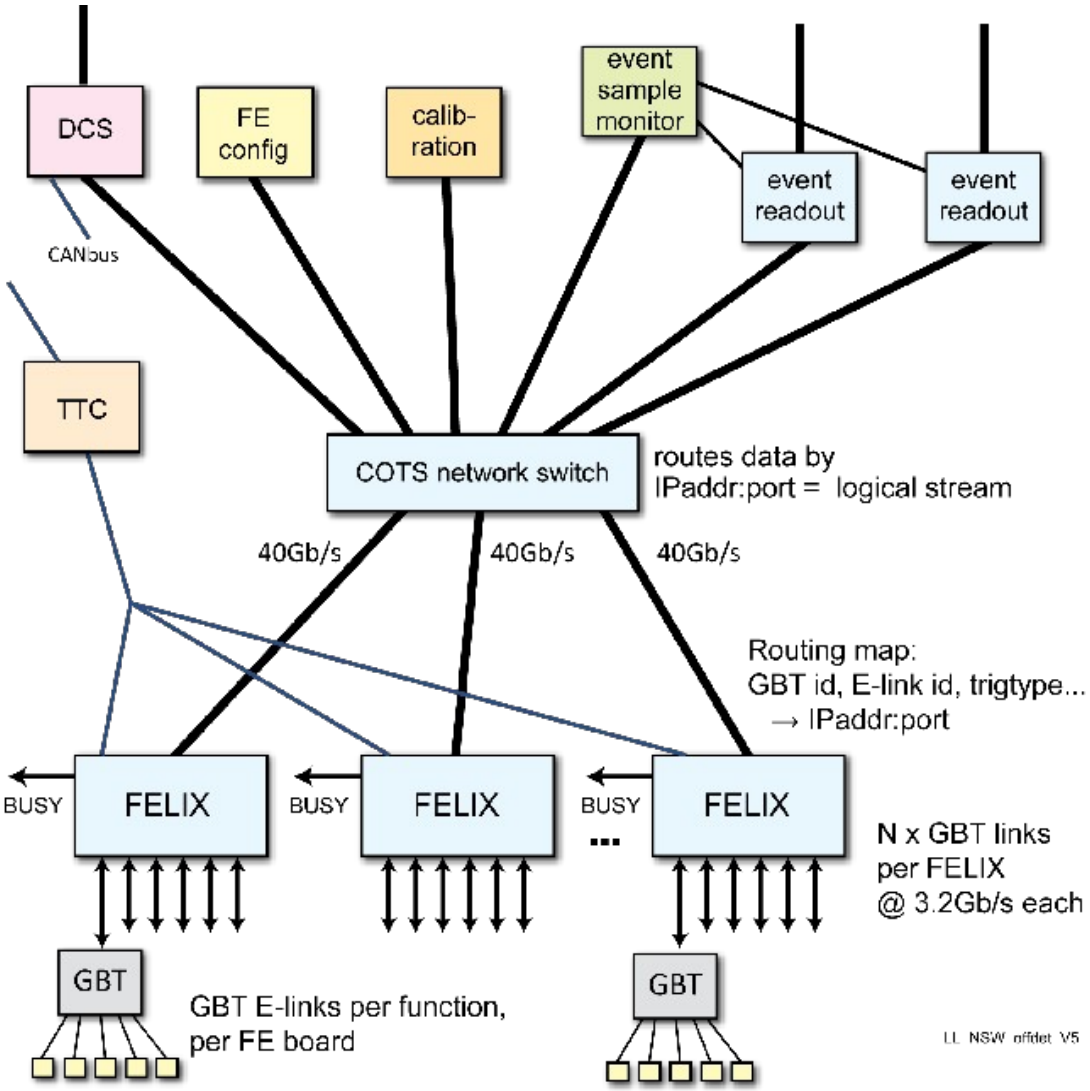
FELIX Functionalities

- **Configurable** data routing device
  - route data **by E-link and/or streams**
  - **advanced routing** based on L1 ID or Trigger type
    - allow for intermediate data-aggregation step
  - data duplication and sampling for monitoring
- Handling of high-level switched protocol
  - Infiniband/Ethernet/...
  - QoS for different traffic types



FELIX Functionalities

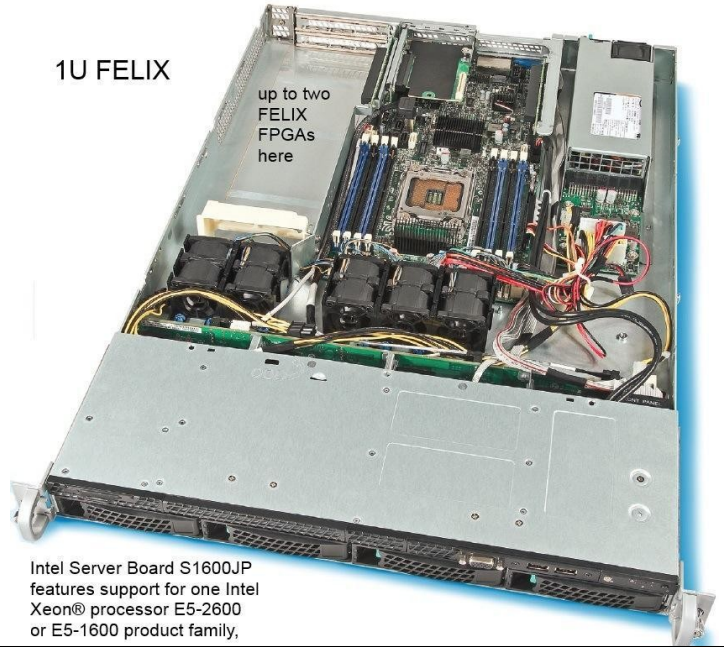
- TTC handling
  - busy handling (no back pressure on GBT)
  - isolates FE electronics from TTC evolution
  - may provide L1 and ECR counting
    - same implementation across ATLAS
- Dedicated low latency path for critical data
  - e.g. Level 1 trigger input data for L0/L1 trigger architecture
- Command synchronisation for calibration
  - programmable sequences
  - synchronisation through TTC



FELIX Demonstrator

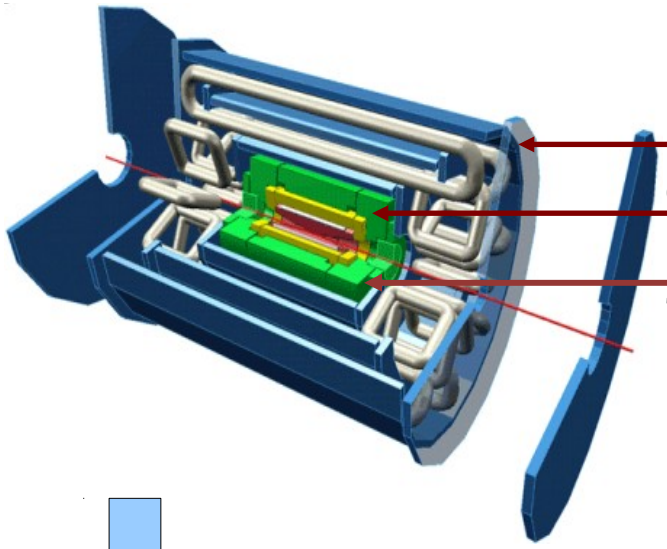
- Hitech Global HTG-710
  - 2 CXP cages: can connect 24 bi-directional links
  - With Virtex-7 X690T FPGA (for PCIe Gen3 8 lanes)
- Intel server board S1600JP
  - 3 PCIe Gen3 slots
- **Scheduled for Q1'15**
  - final design review in Q2'15

- Show feasibility, investigate technology
- Support Phase-I detector developments
  - forward compatibility with Phase II



Not final solution/technology

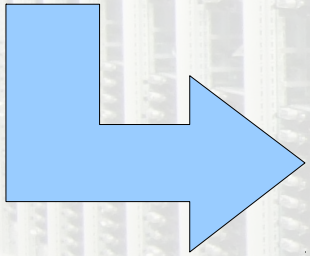
Architecture in Run 4



Muon
Calorimetry
Tracking
~8k GBT
FE Electronics
Level-1 trigger
O(30 μs)
Low-latency link to Level-1
~170
FELIX
L0: 1000 kHz
L1: 400 kHz
170 x 200 Gbps links
Monitoring, Control, Configuration, DCS
Push
Pull
~ N
"ROD"/Readout
Multi-layered data network
10 kHz
High Level Trigger
O(s)
Mass storage

HLT farm size: compute power

- Processing time extrapolations to PU = 200
  - Run 4 rejection requirements similar/better than in Run 1-2 → 1k/100k vs 10k/400k
- A factor **O(50)** in HLT compute power needed wrt to Run 1
- Moore's law on a ~10 years period predicts a **factor 100 increase**



Compute power requirements within expected technology envelope → **HLT farm of similar size wrt to Run 1**

BUT

Software will have evolve to be at least as efficient as today on future technologies (GPGPU, Many-cores, ARM64, ...)

- Assume a similar packaging → **~50 racks**

HLT farm size: network connectivity

- (Temporary) Disregard RoI approach
  - worst case scenario network requirements
- 5MB@400kHz → **~20 Tbps**
- Reasonable to assume
  - 100 Gbps per CPU socket (computing unit)
  - established (>)400 Gbps technology
    - Infiniband EDR x12 → 300 Gbps

Run 2(*)	# of 10 Gbps links
ROS	400
HLT	100

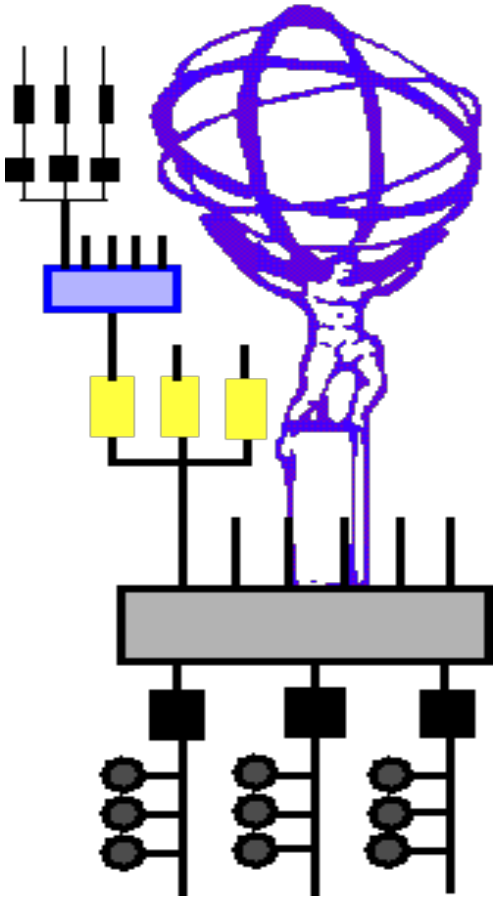
Run 4	# of 400 Gbps links
FELIX	200
HLT	50

- Total **number of ports ~unchanged**
- Network topology and link speeds mix & match depend on compute power packaging

(*) Including redundancy

Outlook

- ATLAS Trigger & DAQ **Technical Design Report** planned in **Q4 2017**
  - cover Level-0/1, DAQ, HLT and Trigger
- Intermediate milestone with the **Initial Design Review** in **Q1 2016**
- IDR preparation include
  - RoI-based vs full event building
  - use of distributed file systems
    - or other data-management abstraction layers
  - processing time estimates
  - technology tracking
    - especially for FELIX
  - studies on event building traffic and flow control



Extra

ATLAS Level-0 – Level-1