Data Acquisition with the Front-End LInk eXchange (FELIX) card

Mengqing Wu

22-02-2022
ESIPAP 2022
Lab tutorial

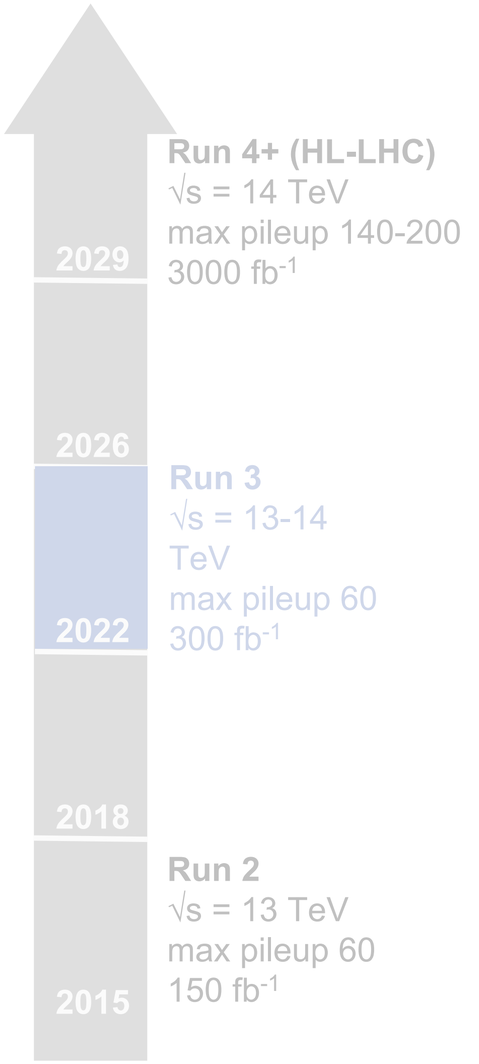
Radboud University Nikhef FELIX ATLAS EXPERIMENT esipap
European School of Instrumentation in Particle & Astroparticle Physics

Today's plan

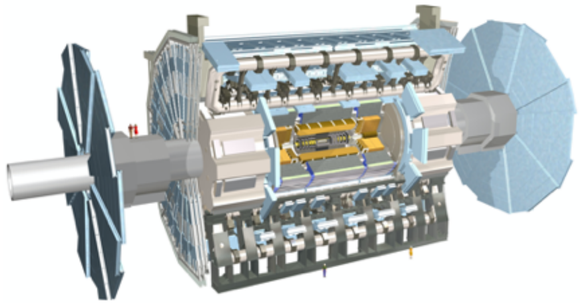
- First of all, a brief overview of the ATLAS Trigger DAQ system from Run2 to Run4+
- Followed by a virtual lab tour: FELIX lab (so-called playground) at NIKHEF Amsterdam
- At the end, a live demonstration will be carried out on data taking using the Run3 FELIX card with a data emulator

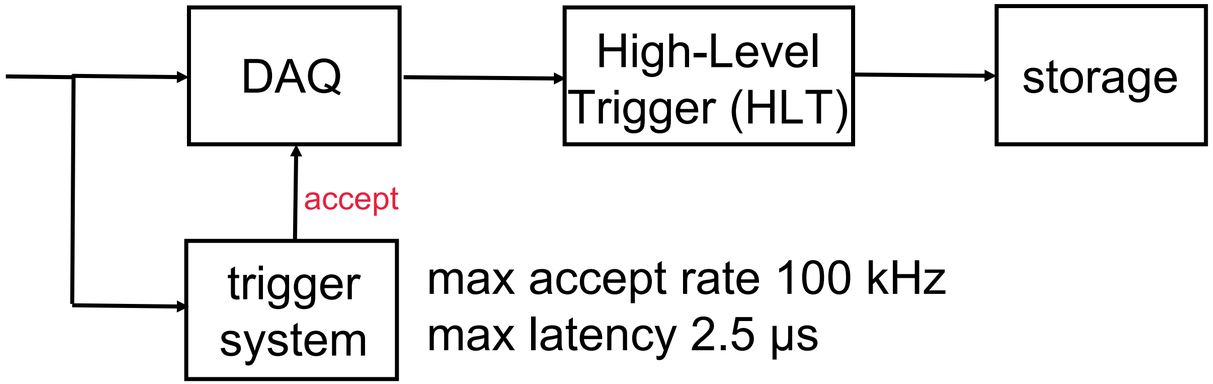
- Acknowledgement:
thanks a lot to the FELIX firmware coordinator Frans Schreuder's help on preparing this tutorial!

ATLAS Trigger and Data Acquisition System

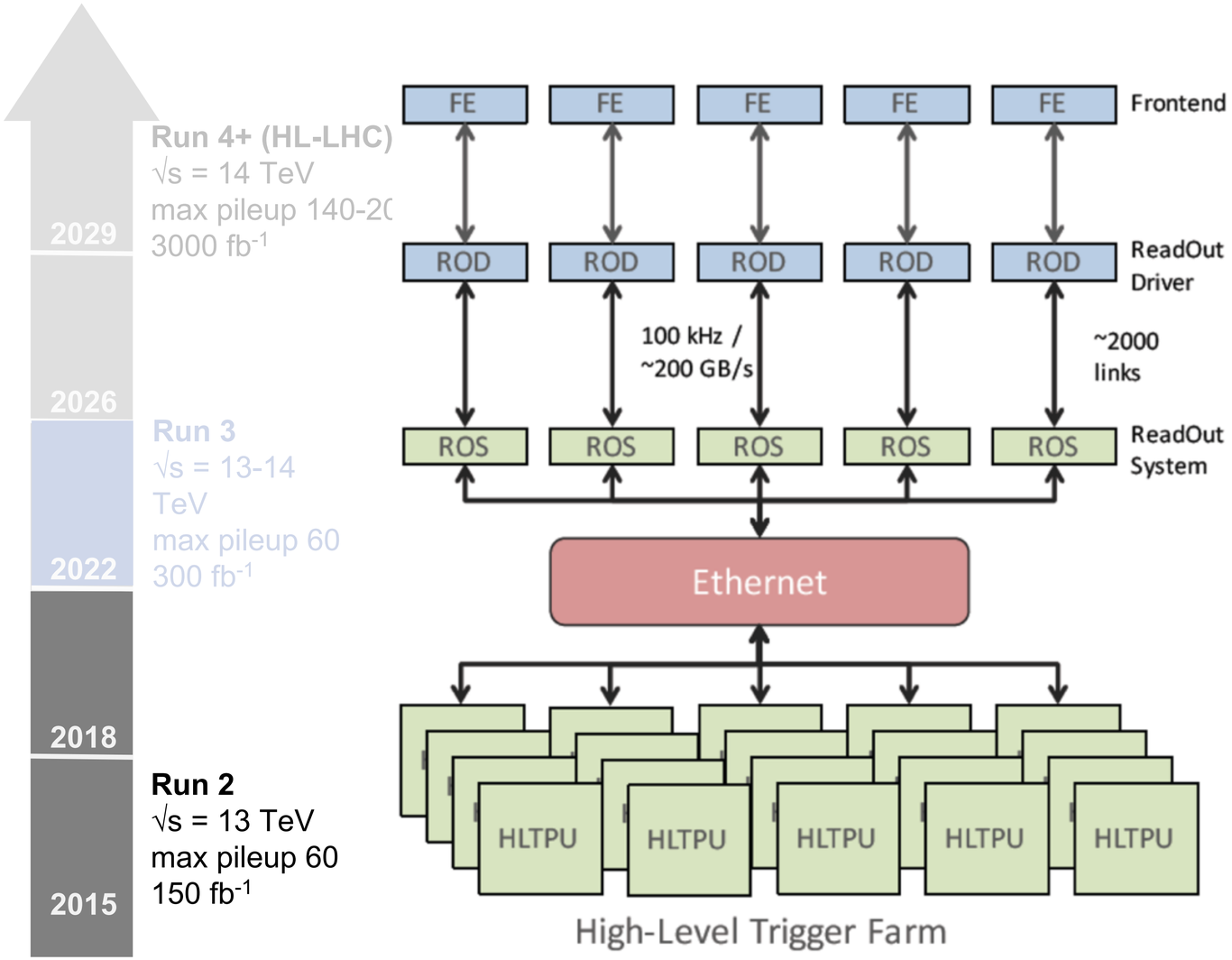


- The Large Hadron Collider collides proton bunches at a 40 MHz rate
- ATLAS detects the collision products
- The maximum event data rate for permanent storage is (currently) 1 kHz
- The trigger systems select interesting events



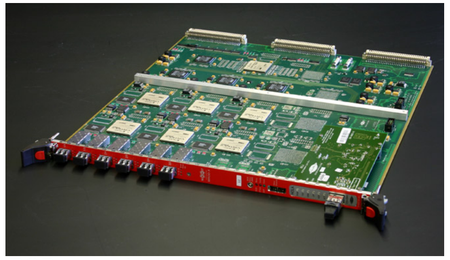


DAQ architecture in Run 2 (2015-2018)

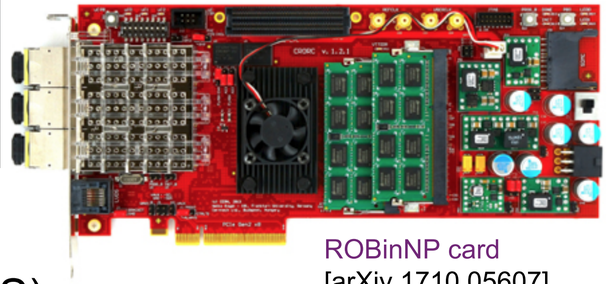


Readout Driver (ROD)

- several types of links from Front End (FE) to ROD
- VME boards, of about a dozen flavours developed and maintained by detectors
- Connected via point-to-point optical link to ROBinNP cards



ROD example

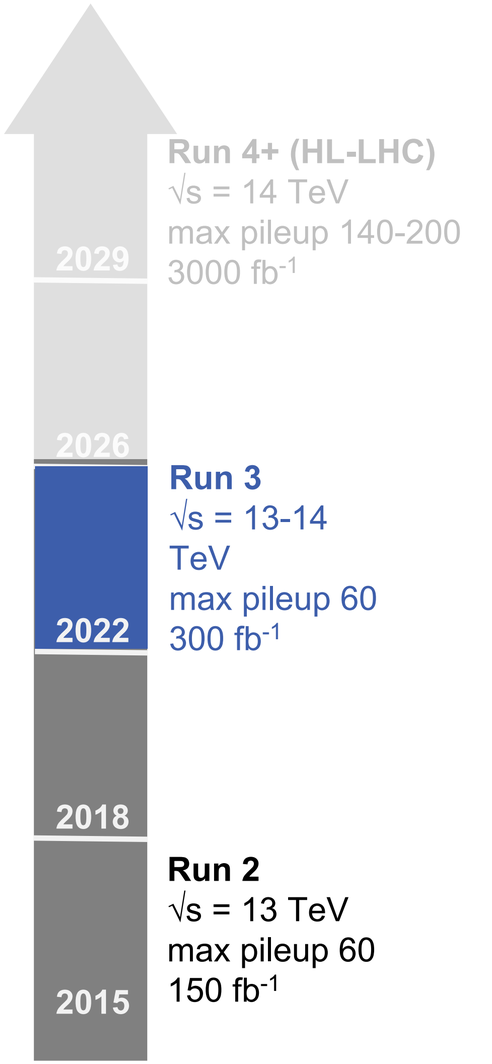


ROBinNP card
[arXiv 1710.05607]

Readout System (ROS)

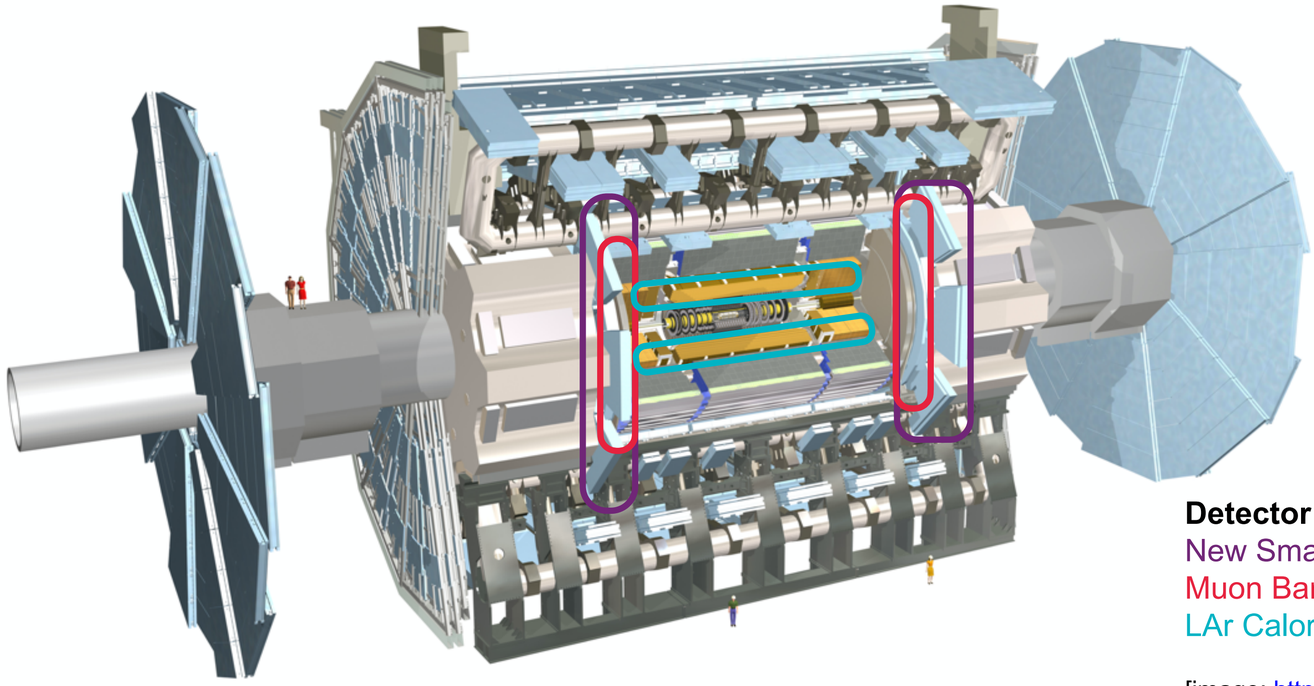
- commodity computers hosting the ROBinNPs
- transfers data to the High-Level Trigger farm over the network

Upgrades for the Run 3



In Run 3: **Now**

- similar conditions to Run 2, higher energy expected
- new detector and trigger systems to improve background rejection
- new readout system: FELIX and SW ROD

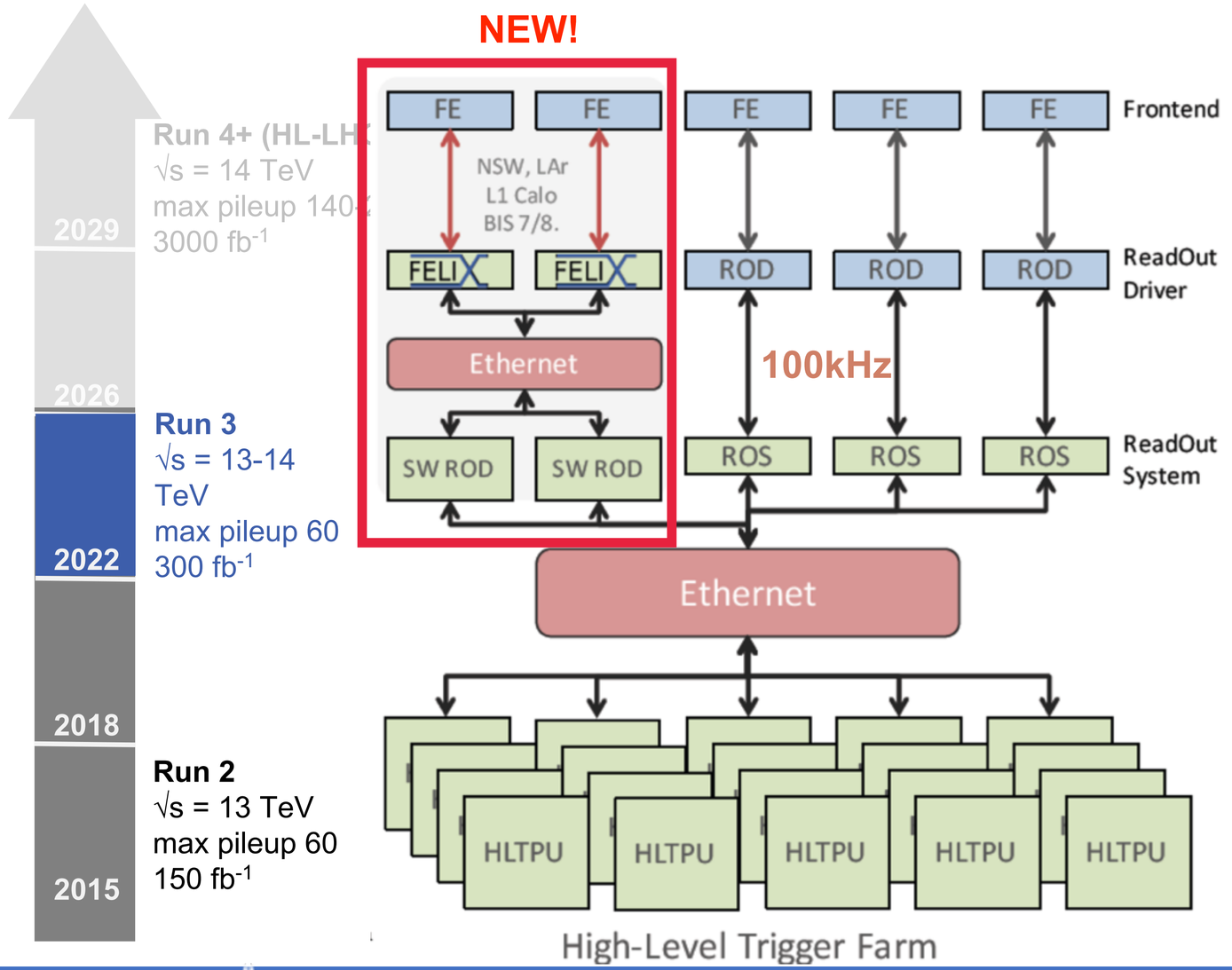


Detector upgrades:
New Small Wheels
Muon Barrel (BIS7/8)
LAr Calorimeter readout

[image: https://cds.cern.ch/record/1095924/]

Upgrades for the Run 3

Today's topic!



Front-End LInk EXchange (FELIX)

- Custom FPGA based PCIe cards hosted on commercial servers
- Interaction with FE includes readout, configuration, trigger, clock distribution, Slow Control, BUSY
- ~100 cards, 60 host servers

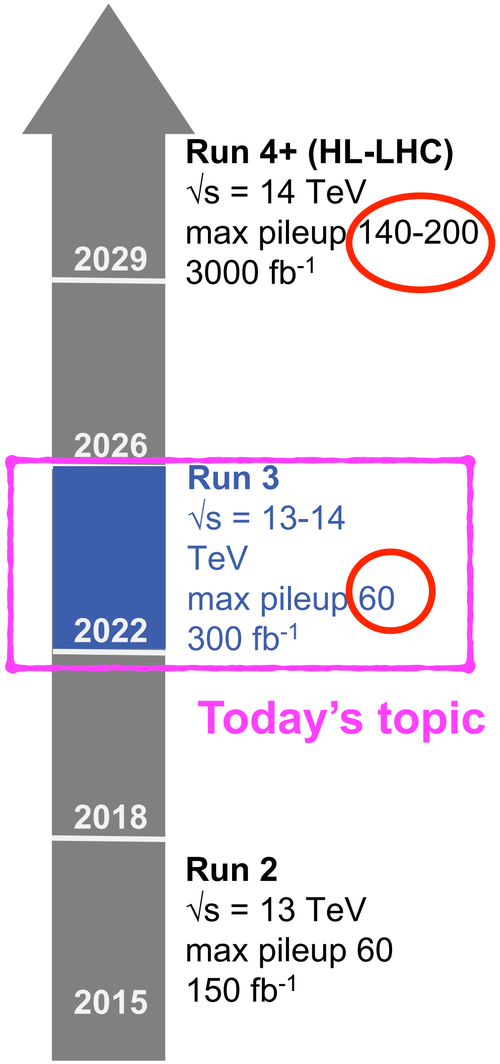
Software Readout Driver (SW ROD)

- Software running on commercial computer
- Builds and aggregates events, detector-specific data processing
- ~30 servers

FELIX + SW ROD

- Reduced hardware/firmware effort
- Data transport decoupled from data processing
- Industry standard data networks earlier in the readout chain
- Aggregation of many links into a single higher speed network link

The preliminary DAQ architecture for the Run 4+



Run 4+ (HL-LHC)
√s = 14 TeV
max pileup 140-200
3000 fb^{-1}

Run 3
√s = 13-14 TeV
max pileup 60
300 fb^{-1}

Today's topic

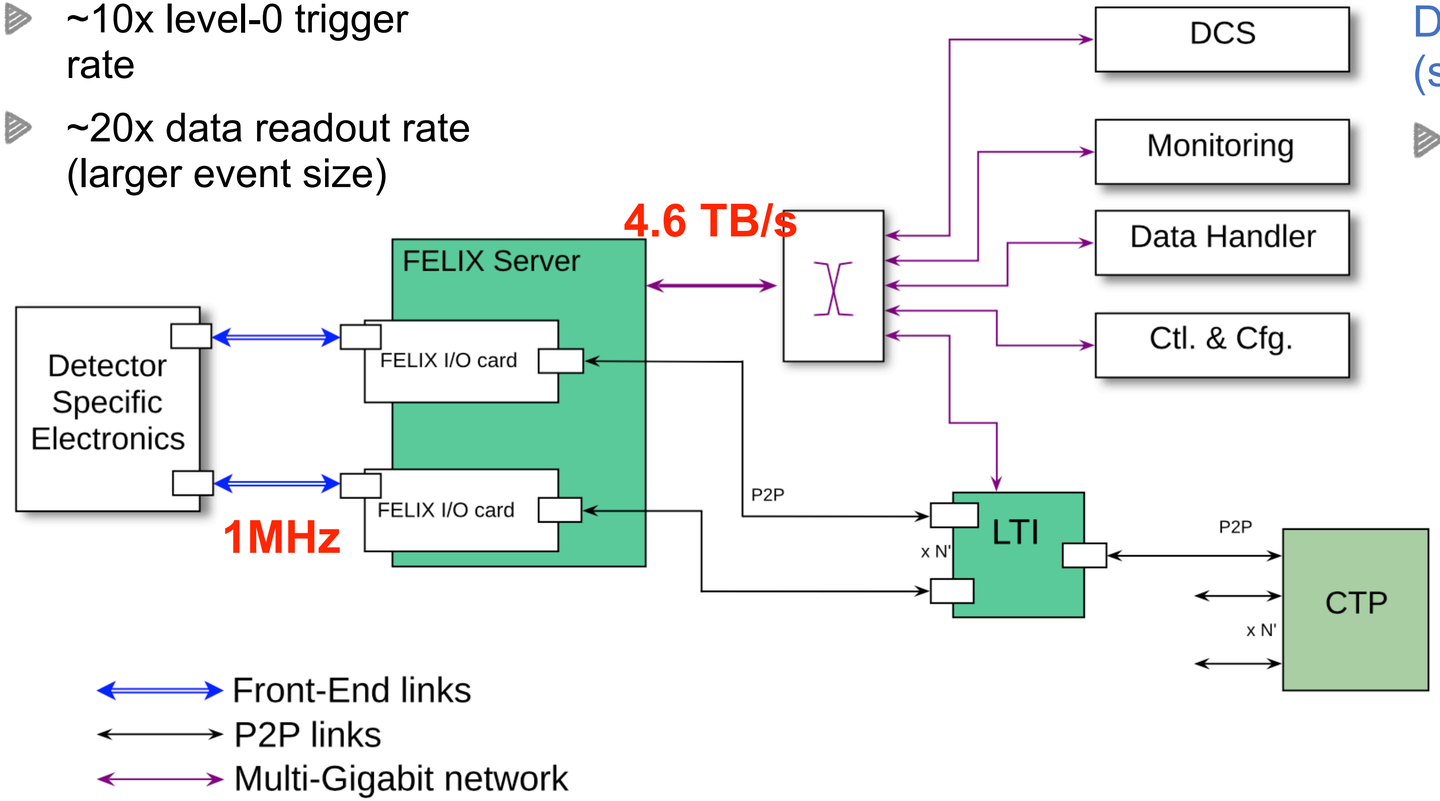
Run 2
√s = 13 TeV
max pileup 60
150 fb^{-1}

Challenges compared to Run2/3

- ~3x mean number of interactions per crossing
- ~10x level-0 trigger rate
- ~20x data readout rate (larger event size)

FELIX

- For all sub-detector systems
- New implementation under development

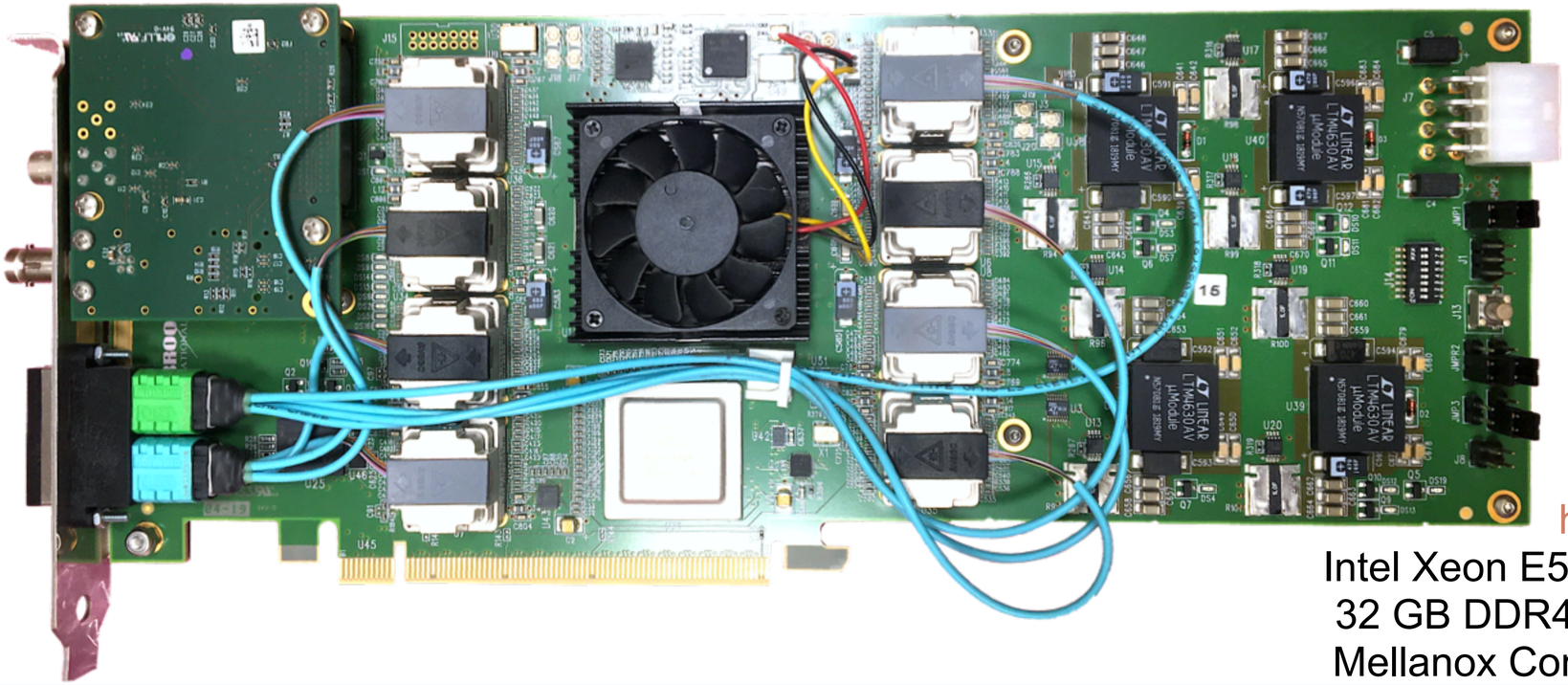


Data Handler (successor of SW ROD)

- More dynamic - to accommodate potential features such as Partial Event Building at data handler level

FELIX Phase1 / Run3 hardware: The FLX-712 card

- FPGA: Xilinx Kintex UltraScale XCKU115
- 4 or 8 MiniPODs to support 24 or 48 bidirectional optical links
- 16-lane PCIe Gen3 (two 8-lane Endpoints with a switch)
- Flash and Micro-controller to support firmware update
- Interfacing to the Timing Trigger and Control (TTC) system
- BUSY output



hosted on FELIX PC

Intel Xeon E5-1660 v4 @ 3.2GHz

32 GB DDR4 2667 MHz memory

Mellanox Connect-X 25/100 GbE

Working modes & firmware

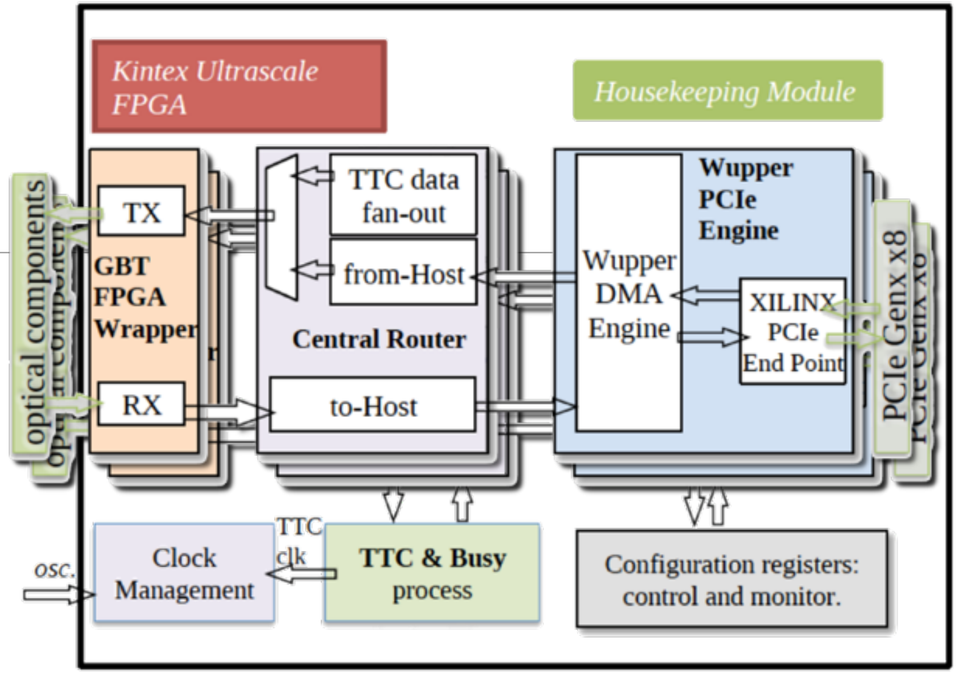
FULL mode

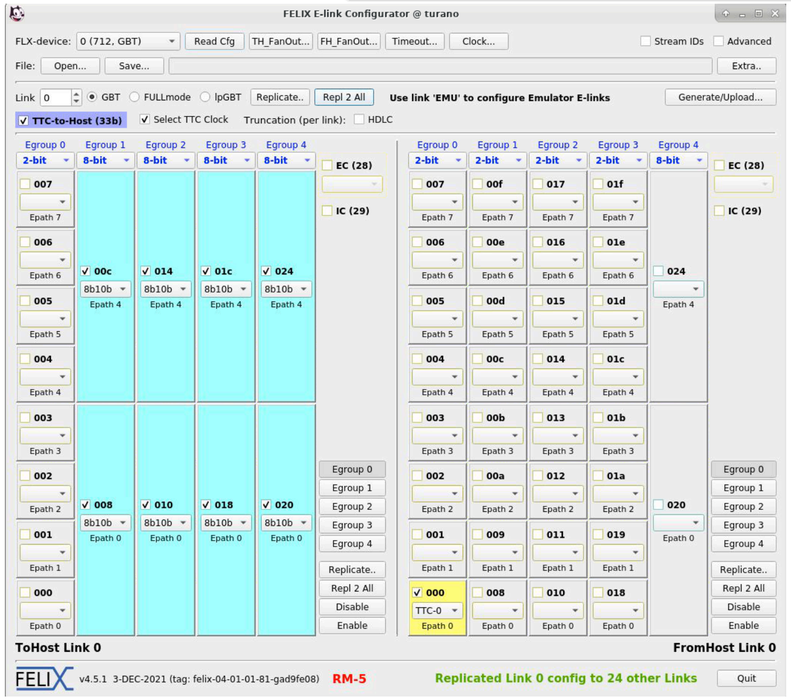
- 24 links, 9.6 Gb/s each, 8b10b encoded
- FPGA-to-FPGA connection
- 12 links saturate the PCIe bandwidth if bandwidth is fully occupied
- To front end: 4.8 Gb/s GBT with E-Links for TTC distribution and configuration

GBT (GigaBit Transceiver) mode

- 24 links, 4.8 Gb/s each conforming to GBT protocol used by the radiation tolerant GBTx ASIC *
- Each link aggregates up to 40 E-Links with configurable bandwidth (80/160/320 Mb/s) + 2 additional E-Links (80Mb/s) for Internal and External Control
- Each E-Link carries a signal with either 8b10b or HDLC encoding, or TTC distribution to the front end electronics
- Heavy utilisation of FPGA resources for configurable logic

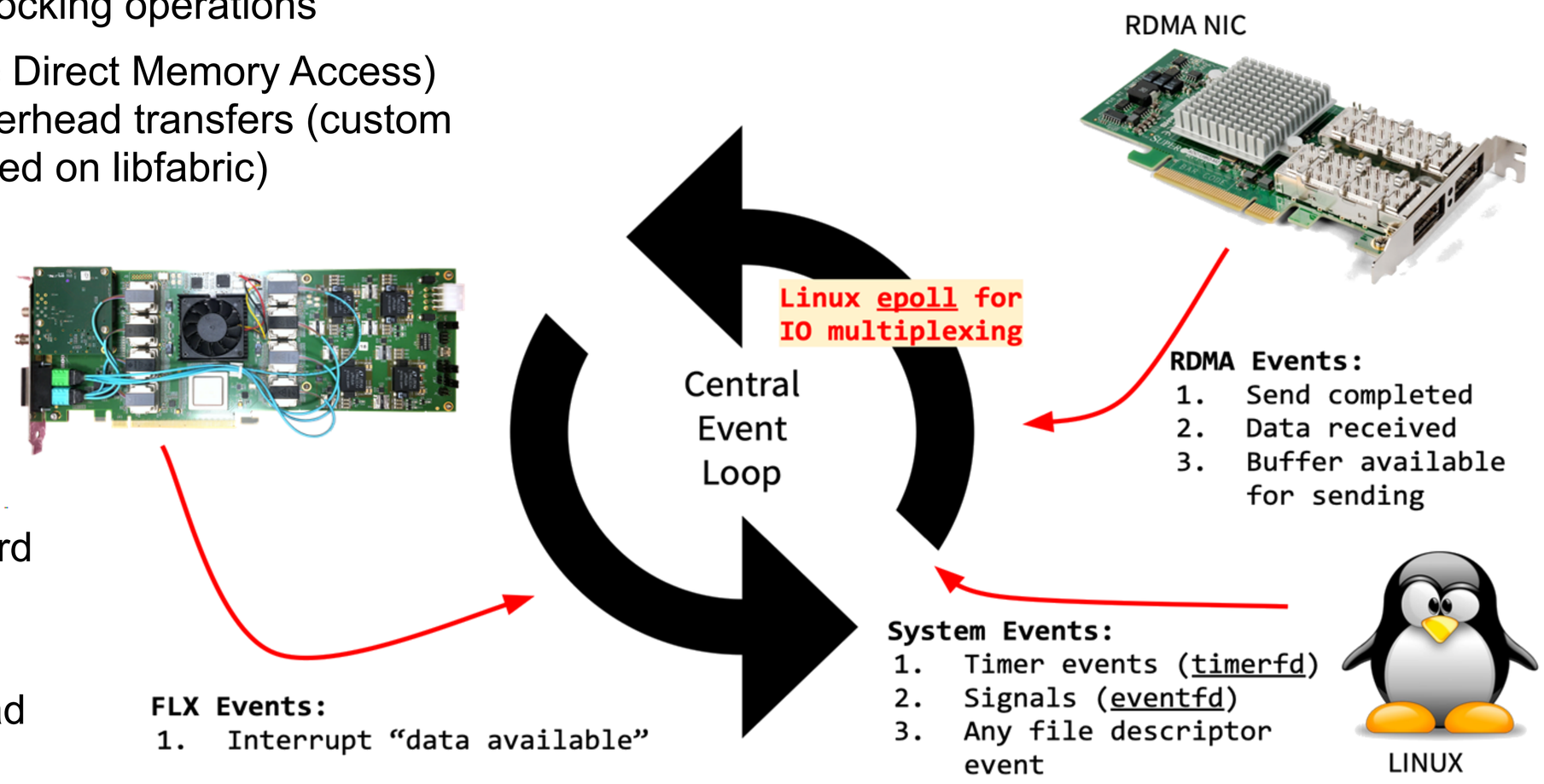
* doi: 10.5170/CERN-2009-006.342

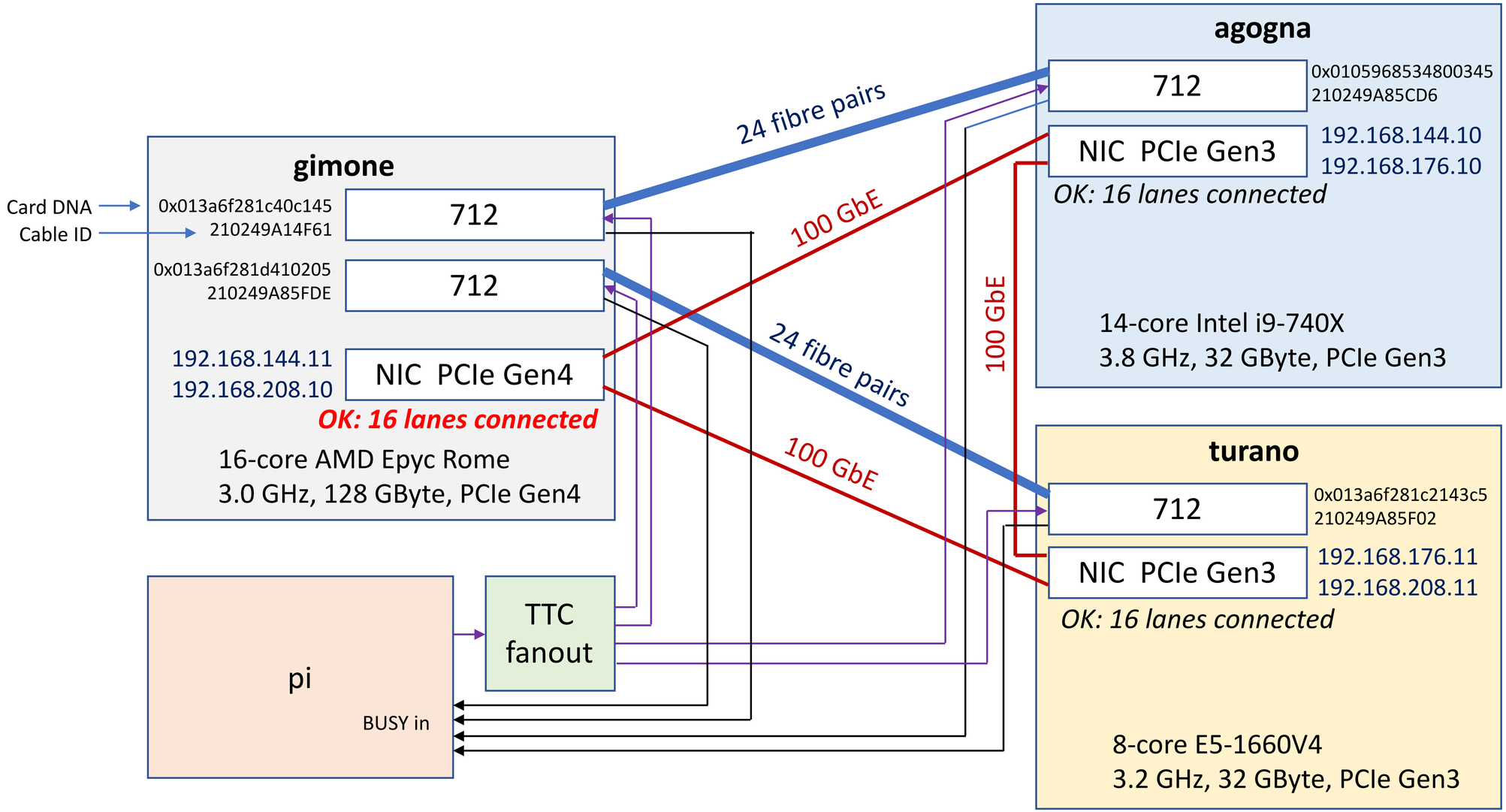




The Felix-star software

- Interrupt-driven event loop architecture
- Asynchronous non-blocking operations
- Uses RDMA (Remote Direct Memory Access) technology for low overhead transfers (custom netio-next library, based on libfabric)



- Single thread, two processes per card
- Allows coalescing of small messages to decrease I/O overhead

Our Today's test setup