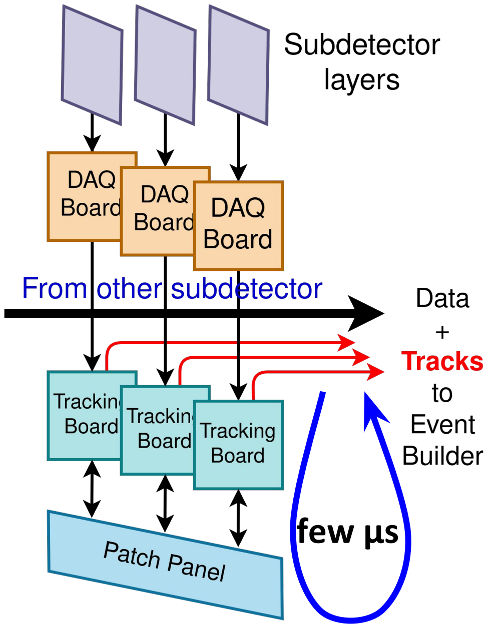
An optical network for accelerating real-time tracking with FPGAs

Federico Lazzari on behalf of the LHCb-RTA project

Connecting The Dots - 20-30 April 2020

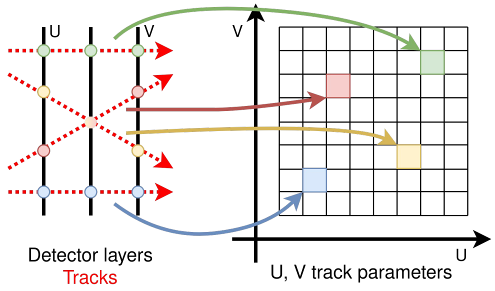
Introduction

- Computational needs for data acquisition and reconstruction keep increasing, and it is worth exploring new solutions based on heterogeneous computing.
- Modern FPGAs have the capability to perform highly parallel processing, with high throughputs and low latencies
- **This allows us to develop a tracking system that can operate at the very first level of processing in a high-rate environment like the LHC**
  - Can instrument just the desired subdetectors
  - Tracks could replace the hits data in real time, reducing data volume before even performing Event Building, and providing trigger primitives
- However, reconstructing tracks requires to combine data from several different layers, that are typically read out separately by the DAQ
- We need a distributed system, with a way to exchange data quickly and effectively between FPGA modules (see figure)
- The “Artificial Retina” is a highly-parallel architecture allowing us to do just this [1]

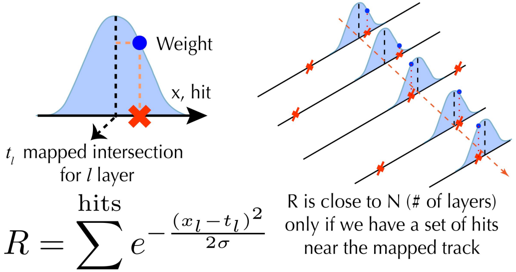


The “Artificial Retina” architecture

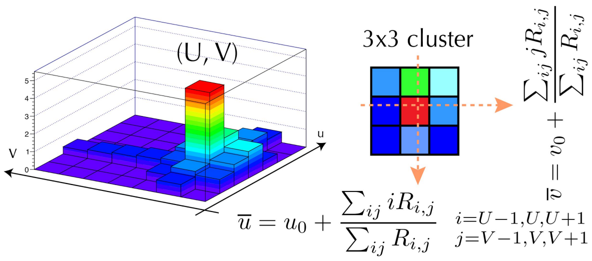
- Track parameter space represented in a matrix of cells
- Each cell computes a weighted sum of hits near the reference track
- Real track parameters are obtained interpolating responses of nearby cells
- To reach high-throughput and low-latencies, cells works in a fully parallel way
- To overcome FPGA size limitations without increasing latency, cells are spread over several chips
- Each FPGA processes a *portion* of every event (vs farm of CPU, each processing a *subset* of events)



Step 2: Accumulating weights (each cell)

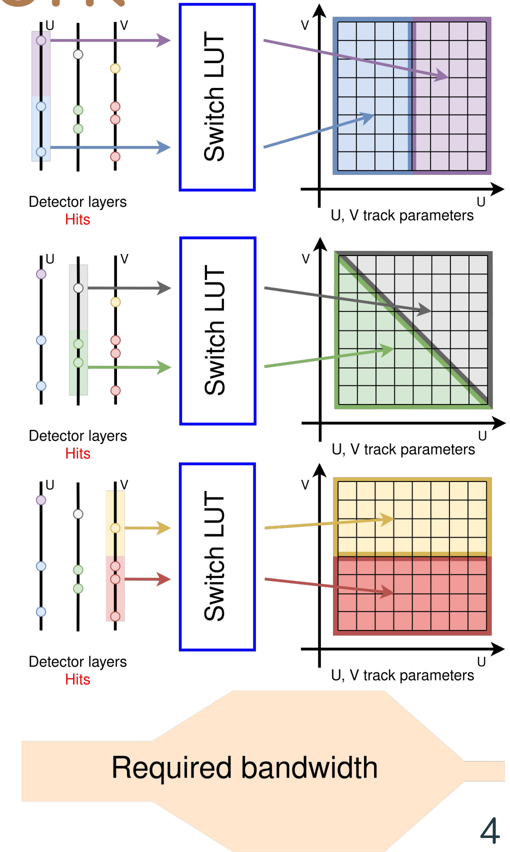


Step 3: Find the local maxima and compute centroid



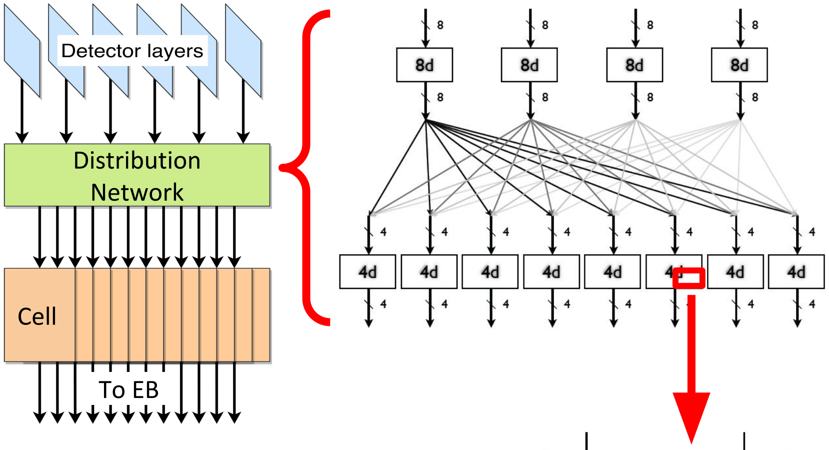
A dedicated distribution network

- Hits are provided to different tracking boards arranged by subdetector readout board
- A custom distribution network rearranges the hits by track parameters coordinates (similar to a “change of reference system”)
- Using Lookup Tables (LUTs), the Distribution Network delivers to each cells only hits close to the parametrized track, enabling large system throughput
- Bandwidth profile increases in the distribution network, and shrinks back to a value lower than input after tracks are found.
  This requires use of fast links

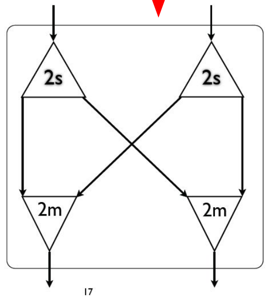


A dedicated distribution network

- We designed a *modular* Distribution Network
- This is implementable within the same array of FPGAs performing the tracking, in separate and independent locations of the chip
- FPGAs have numerous high-bandwidth transceivers (XCVRs)

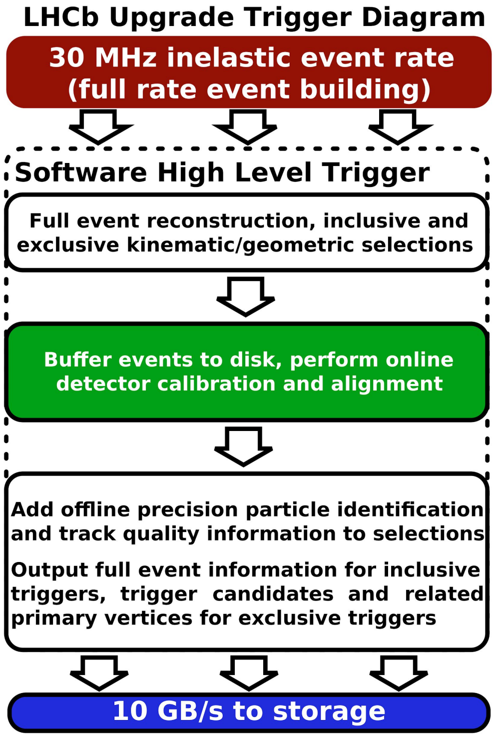


- Optical serial links allow to exchange hits between several boards with great flexibility (e.g. connect distant boards, increase the number of links on a busy path, implement only useful paths)
- The Distribution Network is the crucial backbone of our system - and the main focus of this talk



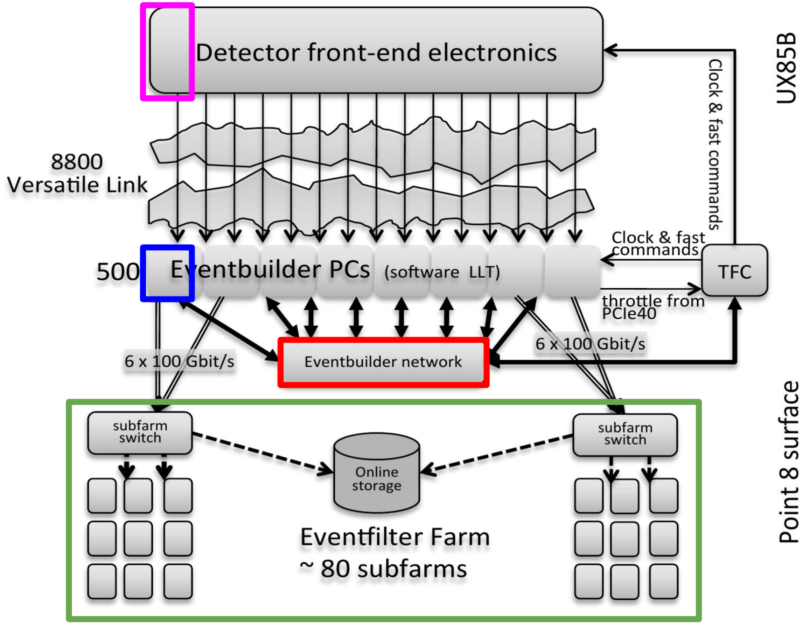
Application to LHCb

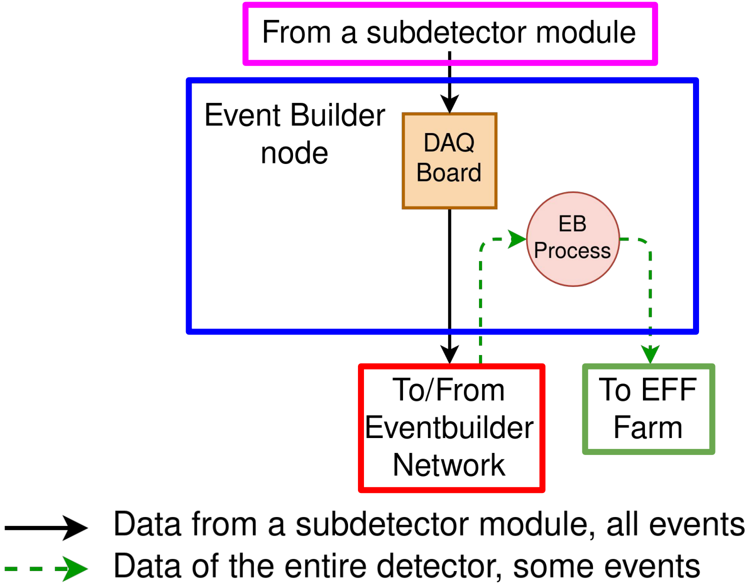
- LHCb is a forward spectrometer designed to study b and c physics
  - is being upgraded and in 2021 will start to acquire data with an almost completely new detector and trigger system [2]
- Instantaneous luminosity will increase by a factor 5
- To take full advantage of increased statistics, LHCb upgrade will have a **trigger-less readout system**
- The full inelastic collision rate of **30 MHz** will be processed by the **full software trigger** (HLT)
- The Event Builder will handle the sizable **bandwidth of 5 TBytes/s**



System integration in LHCb

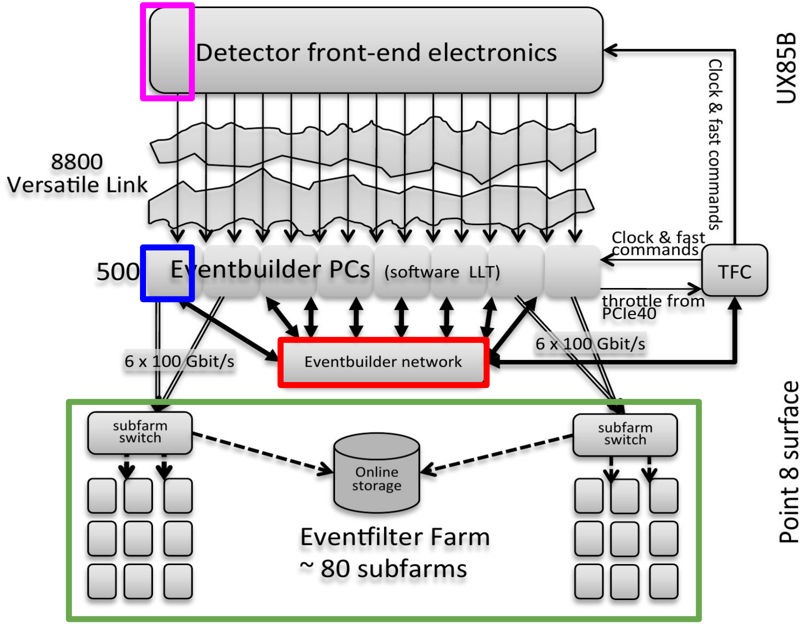
- In LHCb data flows from detector to Eventfilter Farm (EFF) through the Event Builder node and its network

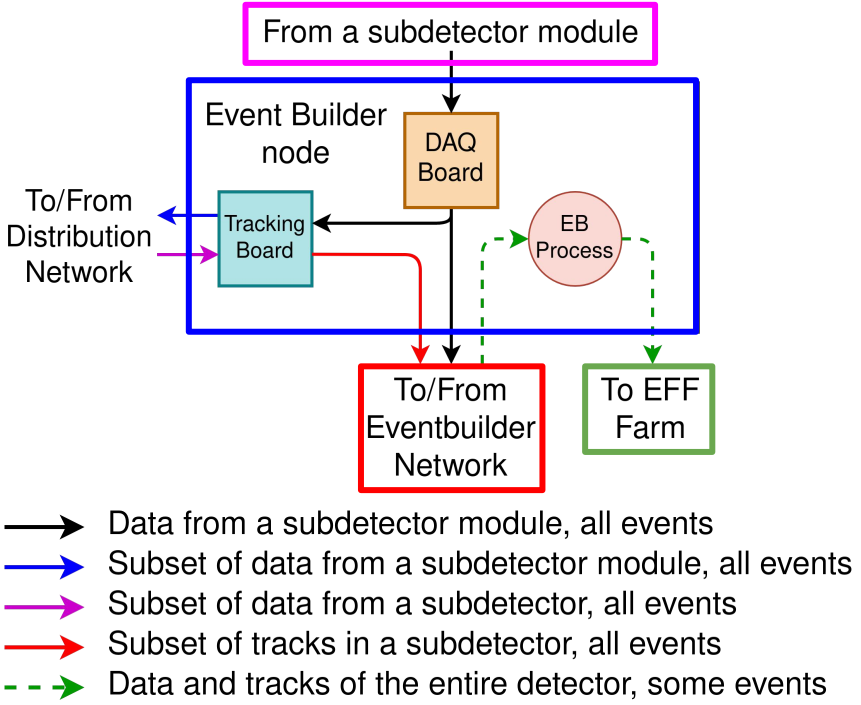




System integration in LHCb

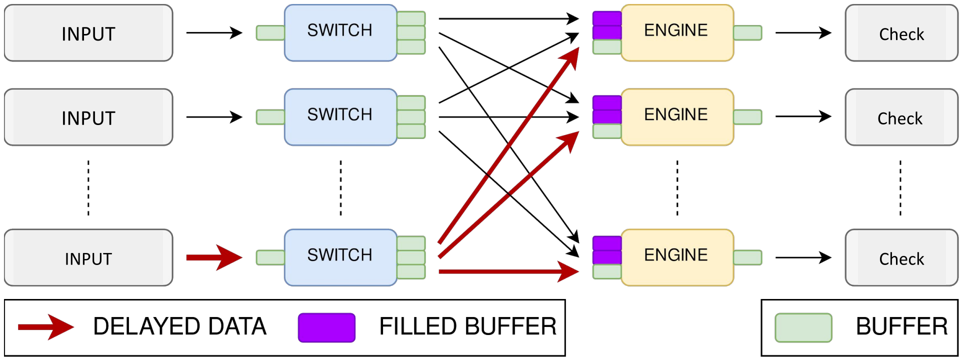
- The “Artificial Retina” could find a place in the Event Builder nodes
- The Event Builder collects the tracks and performs the building, treating the “Artificial Retina” like a virtual subdetector





Testing throughput/latency performance

- We built a prototype based on Intel Stratix-V FPGA, tracking in a generic 6 axial-layers detector
  - “Artificial Retina” architecture includes buffers to decouple the components
- Input from circular buffers injecting 'realistic' hit data at 30 MHz, as in LHC running.
- Important to simulate realistic time skews between inputs, coming from different readout units.
  - In LHCb the maximum expected delay is ~10 µs
- We delayed one input by this value to stress the buffers and verify tolerance to input latencies

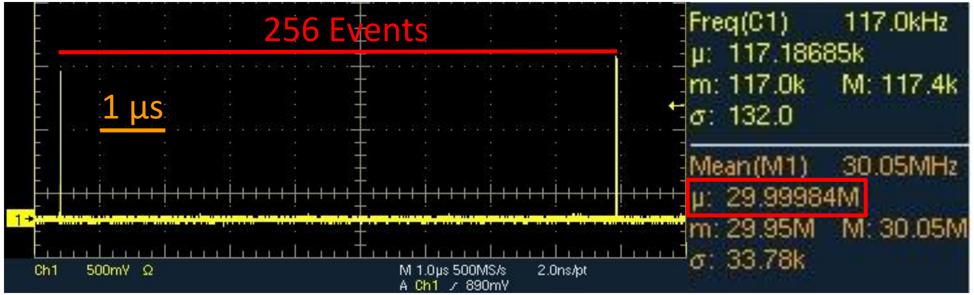


Testing throughput/latency performance

- We built a prototype based on Intel Stratix-V FPGA, tracking in a generic 6 axial-layers detector
    - “Artificial Retina” architecture, includes buffers to decouple the components
- Input from circular buffers injecting 'realistic' hit data at 30 MHz, as in LHC running
- Important to simulate realistic time skews between inputs, coming from different readout units.
    - In LHCb the maximum expected delay is ~10 µs
- We delayed one input by this value to stress the buffers and verify tolerance to input latencies

The test was fully successful:

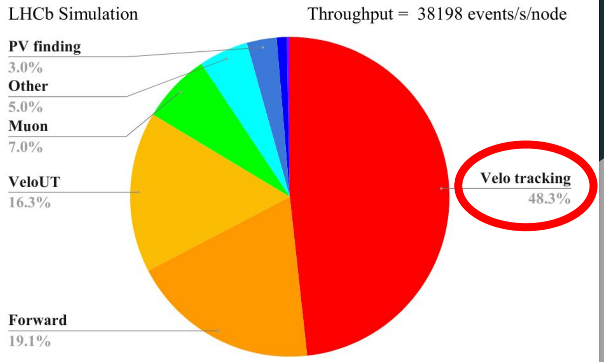
- The buffer size remained under control
- Output rate was equal to the input rate (30 MHz)
- Every track was correctly reconstructed
- Without input skews, system latency is ~0.4 µs

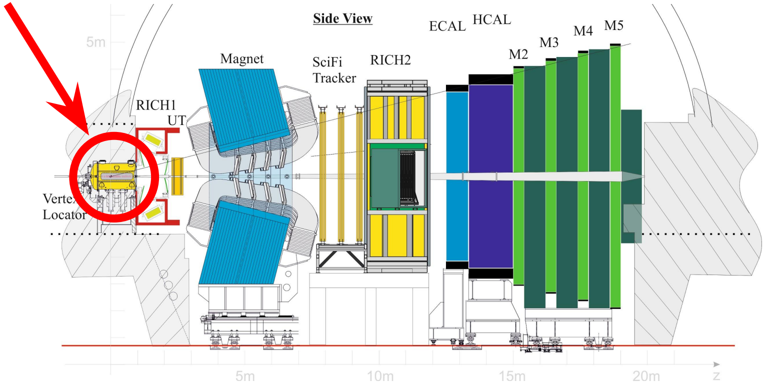


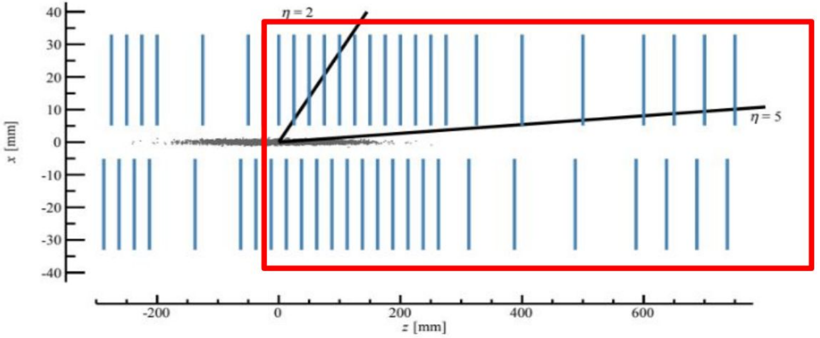
Screenshot of the oscilloscope showing the system throughput

LHCb Vertex Locator

- Reconstructing tracks in the LHCb Vertex Locator (VELO) is the most time consuming task, so it is interesting to study the implementation of the “Artificial Retina” to this subdetector [3]
- VELO is composed of 52 silicon pixel modules, but only 38 are place in the forward region.
  Other layers used mainly to optimize primary vertex precision
- A single DAQ board acquires data from each VELO module → at least 38 Tracking Boards required



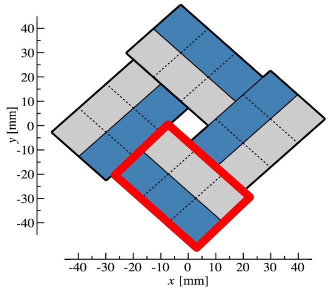


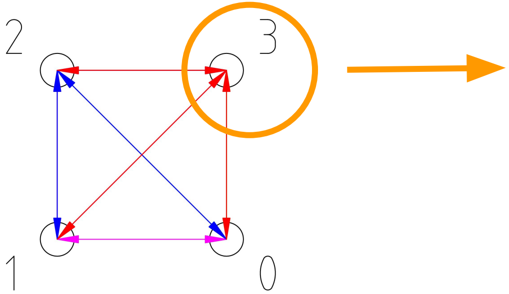


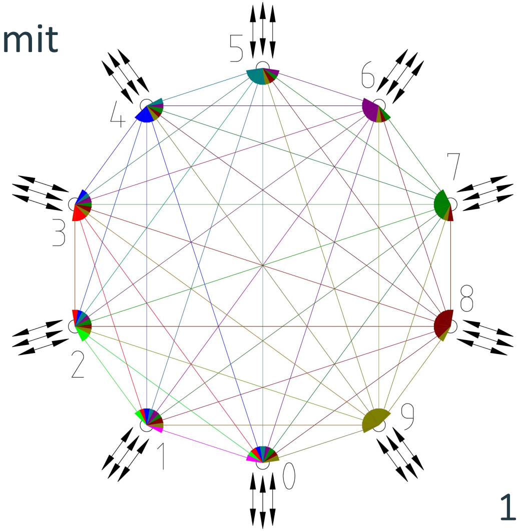
Distribution network topology

- Dividing the VELO track parameters space in 4 quadrants, we could implement 4 full-mesh networks of 10 FPGAs each
- We could connect the i-th FPGA of a network to the i-th FPGA of the other networks creating a full-mesh of full-mesh
- Hit exchange between not directly connected FPGAs require a hop end the communication latency increase, however the total latency remains below the µs limit

- This configuration require 12 links for FPGA
- Several off-the-shelf FPGA boards have 16 XCVRs

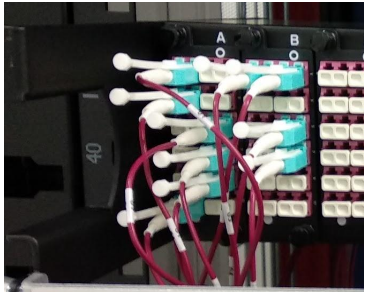


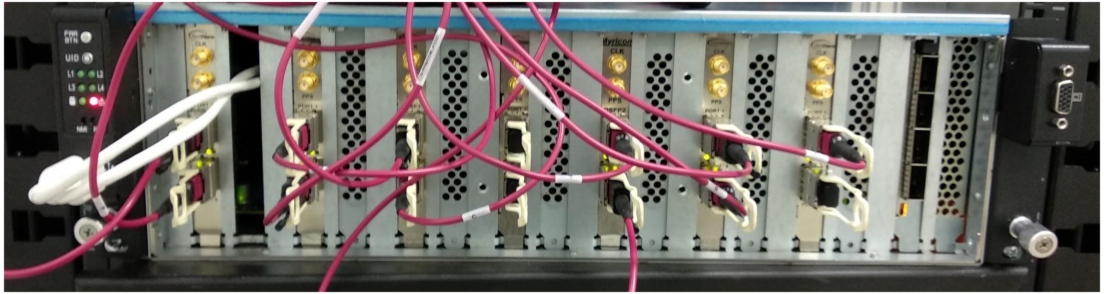




Building a life-size network prototype

- We are building a prototype of one full-mesh network of 10 nodes (half-size FPGA cards) in the LHCb Data Center
- Plan is to start with simulated data, gradually progressing to parasitic operation on real VELO data during Run-3 data taking
- Initial test already in operation with 5 cards sending pseudorandom sequences (**PRBS31**) to the other 4 FPGAs through 10 Gbps optical links
- Continuously tested for 23 days at max speed, measured no transmission error on all but one link **(BER < $2 \cdot 10^{-16}$ CL: 95%)** (single link showing **BER ~ 10^{-13}**, errors disappear lowering speed)





Summary

- In future HEP experiments event building and tracking will be even more challenging
- The “Artificial Retina” is a highly-parallel tracking system that can provide efficient real-time processing of data already at pre-build level
- A fast dedicated distribution network is a crucial element of the system, allowing to collect data from several DAQ nodes, overcome FPGA size limits, reach high-throughput and low-latencies
- We designed a network capable of a real application in LHCb, and we now have a life-size prototype in advanced state of realization, passing all preliminary tests
- Our system will be part of a testbed that LHCb is deploying for the specific purpose of testing new computing accelerators during data taking in the upcoming LHC Run

Backup

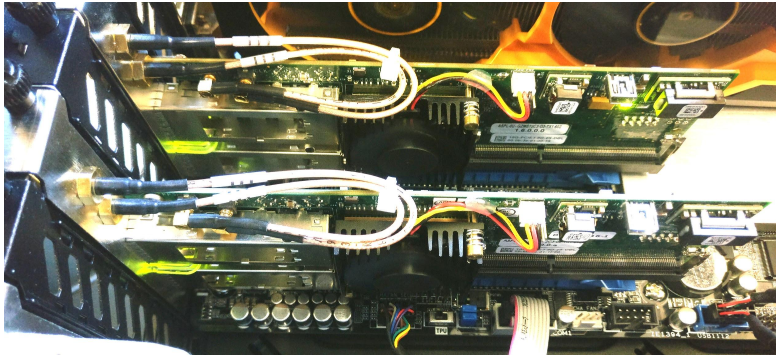
Bibliography

1. G. Punzi et al. on behalf of the LHCb Real-Time Analysis project, ***Real-time reconstruction of pixel vertex detectors with FPGAs***, The 28th International Workshop on Vertex Detectors (Vertex2019) - Tracking and vertexing
2. LHCb Collaboration, ***LHCb Trigger and Online Upgrade Technical Design Report***, CERN-LHCC-2014-016
3. G. Tuci, ***Reconstruction of track candidates at the LHC crossing rate using FPGAs***, CHEP 2019
4. R. Cenci, F. Lazzari, P. Marino, M.J. Morello, G. Punzi, L.F. Ristori, F. Spinella, S. Stracka, J. Walsh, ***Performance of a high-throughput tracking processor implemented on Stratix-V FPGA***, Nuclear Inst. and Methods in Physics Research, A 936 (2019) 344–345
5. R. Cenci, F. Lazzari, P. Marino, M.J. Morello, G. Punzi, L.F. Ristori, F. Spinella, S. Stracka, J. Walsh, ***Development of a High-Throughput Tracking Processor on FPGA Boards***, PoS(TWEPP-17) 136

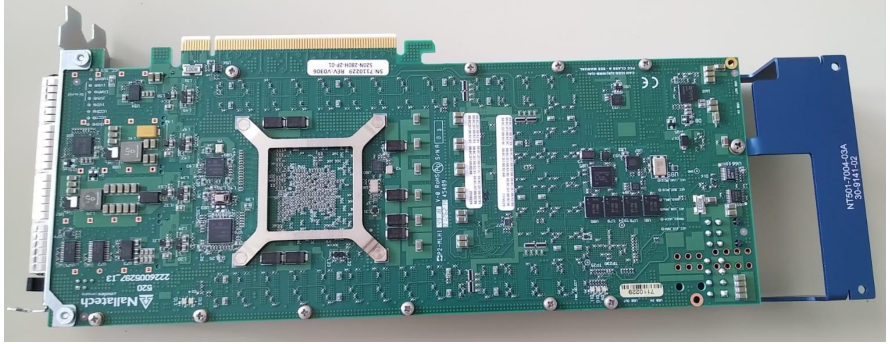
- Riccardo Cenci, Andrea Di Luca, Federico Lazzari, M.J. Morello, G. Punzi, ***Real-time reconstruction of long-lived particles at LHCb using FPGAs***, ACAT 2019

Hardware

- Prototyping board,
  2 Intel Stratix V FPGAs,
  96 optical links



- PCIe 8x board, 1 Intel Arria V GX FPGA, 8 optical links



- PCIe 16x board, 1 Intel Stratix 10 FPGA, 16 optical links

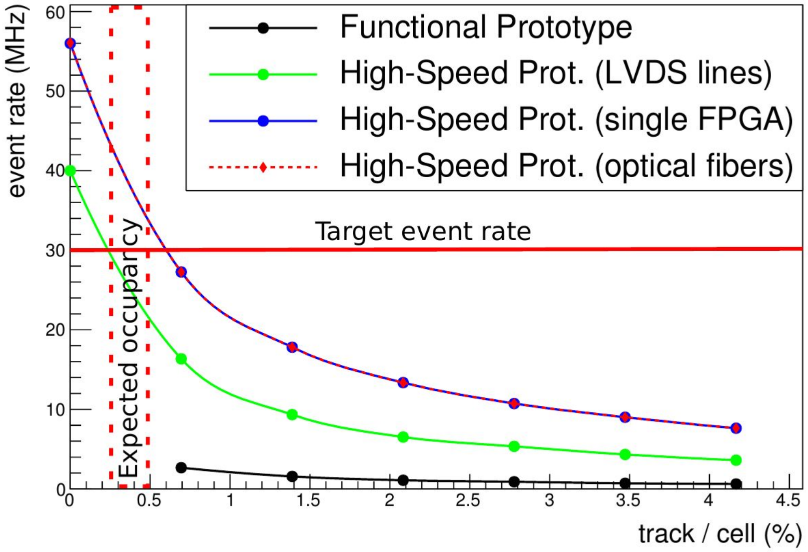
“Artificial Retina” prototype

- We implemented a prototype of a full system for a generic 6 axial-layers detector with 3 input sources and 4 cells matrices [4]

We tested the system with three configuration:

1. Input sources and cells matrices in the same FPGA
2. Input sources and cells matrices in two FPGA connected with on board LVDS lines
3. Input sources and cells matrices in two FPGA connected with optical serial lines

- Due to the limited number of LVDS lines, configuration 2 can not reach the full throughput
- **The throughput** of configuration 3 (optical fibers) **is exactly the same** of configuration 1 (single FPGA) [5]
- The system **latency is ~0.4 μs**



High-speed prototype is implemented on a board with 2 Intel Stratix V FPGAs

VELO tracking with "Artificial Retina"

- Use $B_s \rightarrow \Phi\Phi$ sample from official LHCb full simulation
- Run through simulation of tracking algorithm
- Inject result into LHCb's track performance benchmark
- Comparison with standard CPU algorithm shows very close efficiency performance on fiducial tracks (-200mm < z < 200mm) [3]
- Work still needed in merging back tracks split over multiple retina matrices and finalizing firmware

Track type	ε CPU pat-reco (%)	ε FPGA pat-reco (%) all z	ε FPGA pat-reco (%) fiducial z-region
Long tracks with $p > 5$ GeV/c and hits in VELO> 5	99.84 ± 0.02	99.27 ± 0.06	99.45 ± 0.05
Long tracks from b with $p > 5$ GeV/c and hits in VELO> 5	99.61 ± 0.13	99.24 ± 0.21	99.41 ± 0.18
Long tracks from c with $p > 5$ GeV/c and hits in VELO> 5	99.89 ± 0.12	98.50 ± 0.53	98.62 ± 0.53