

A Low-Cost, Low-Power Media Converter Solution for Next-Generation Detector Readout Systems

Alberto Perro, Mitja Vodnik, Paolo Durante



Outline

Current Situation

Future DAQs & Motivation

Proposed Solution

Proof of Concept

Testbench

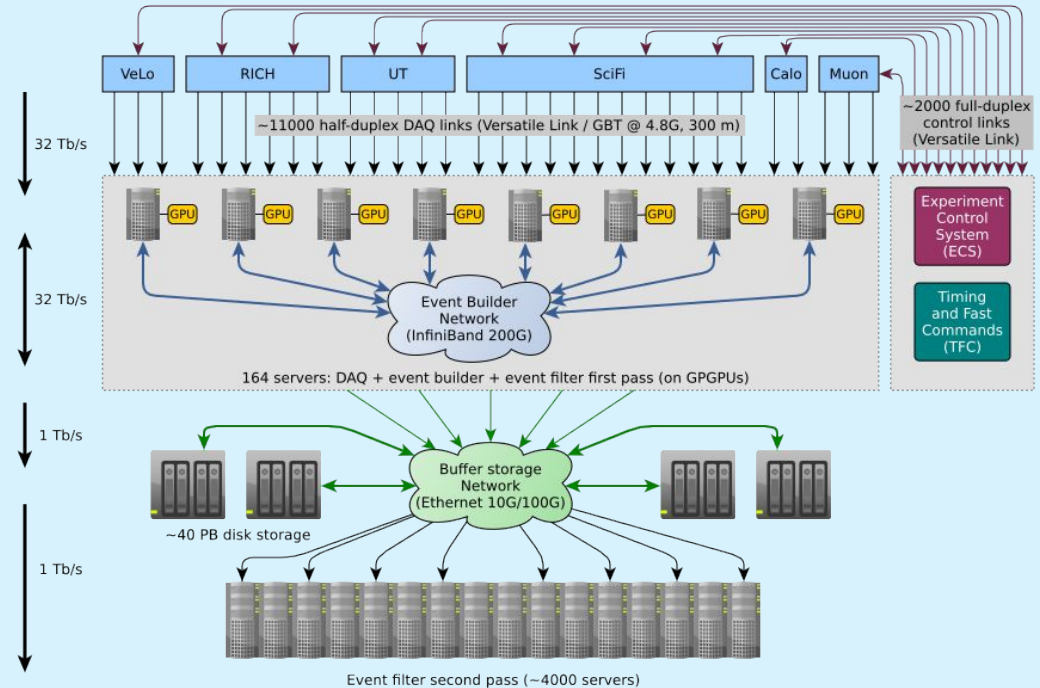
Achievements & Future Directions

LHCb Readout System in Run 3

The LHCb Data Acquisition (DAQ) system reads out the full detector using **~11k GBT optical links @ 4.8 Gbps**, for a total throughput of **32 Tbps**.

These links are read by **custom high-end FPGA** boards which aggregate, process the data, and transfer it to the Event Builder **via PCIe**.

The DAQ architecture will remain mostly the same in Run 4 with the addition of new sub-detectors.



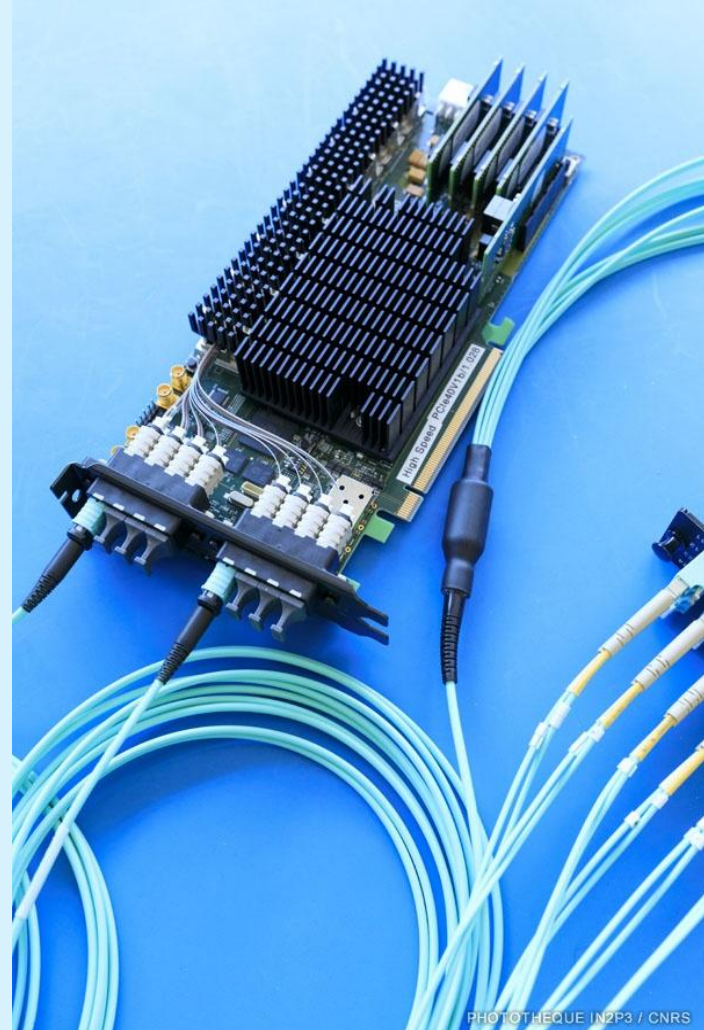
Current Hardware

PCIe40 FPGA Backend Boards:

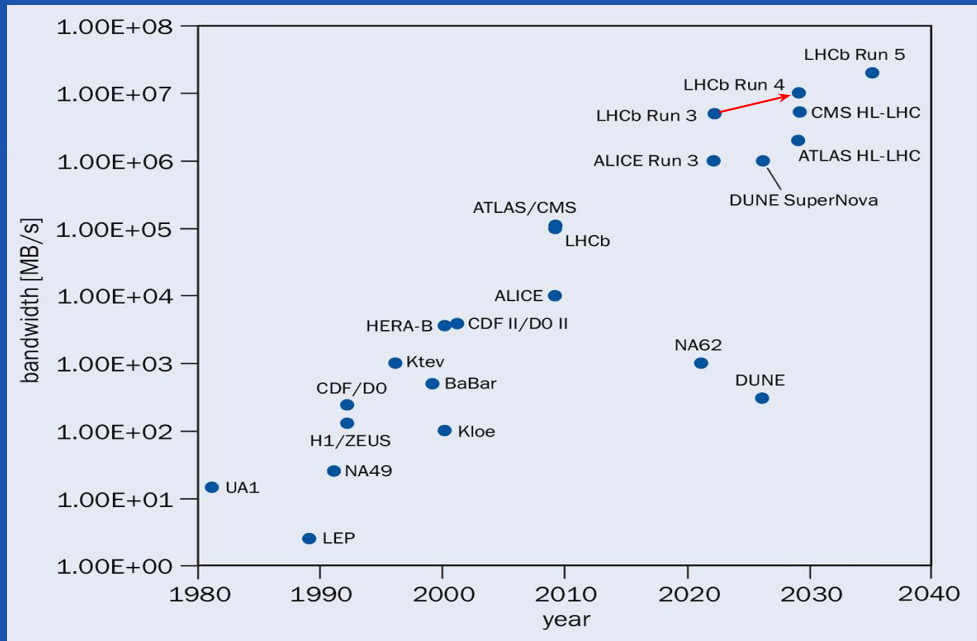
- 1.2M Logic Elements (Altera Arria 10GX)
- 48 GBT Radiation Hard Optical Links @ 4.8Gbps
- Two SFP+ modules
- Two PCIe Gen3 x8 interfaces

520 cards are used in LHCb, of which 445 are dedicated to read out the data. The rest are used for control and timing.

The PCIe40 successor - *PCIe400* - will be used from Run 4 for the new sub-detectors.



Future DAQs in the HL-LHC era



In the HL-LHC era, detector will require major upgrades on the Front End Electronics (FEE) to take advantage of the higher luminosity.

In practice, these upgrades will use a faster link protocol (IpGBT) which runs at 10.24 Gbps and they will need many more links (~30k estimated in LHCb).

DAQ systems will need to be upgraded to manage the higher throughput:

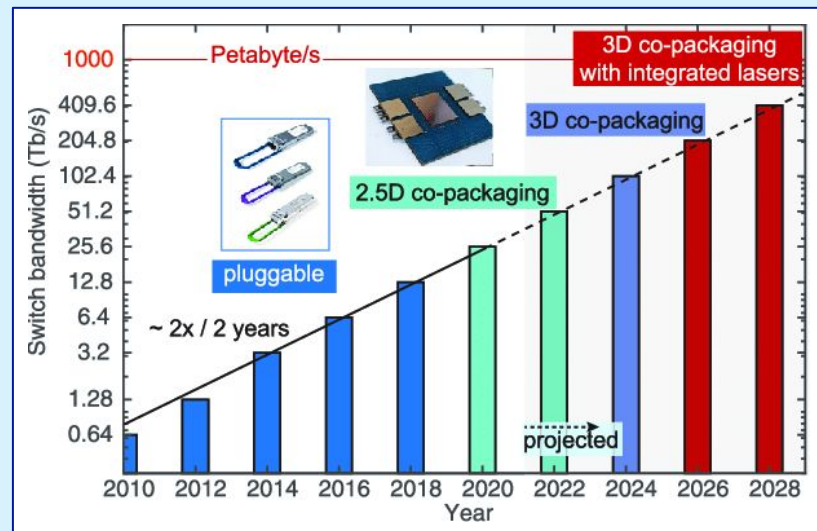
	Run 3	Run 5
Throughput (Tbps)	32	300
Links (k)	10	30

Network Fabric of the Event Builder

The Event Builder in Run 3 uses **InfiniBand HDR @ 200 Gbps**. This design decision, made in 2020, was based on the performance difference between Ethernet and InfiniBand for this specific use case at the time.

Ethernet, however is **cheaper, widely adopted**, and it is evolving at a fast pace.

Modern Ethernet ASICs can switch up to **50 Tbps**, with throughput expected to **double every two years**.



Chao Xiang [CC-BY-4.0](https://creativecommons.org/licenses/by/4.0/)

Motivation

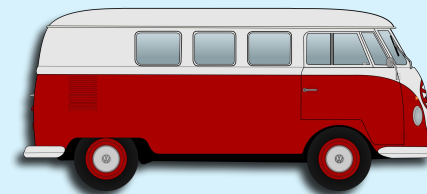
High-end FPGAs are moving towards offering **fewer transceivers** that support **very high data bandwidths** (currently up to 112 Gbps), while FEE links prioritize **radiation hardness** and **power consumption**.

The objectives of this work are:

- Assess the potential of **lower-end FPGAs**, which offer slower transceivers at a fraction of the cost.
- Investigate how much the DAQ architecture can be simplified by leveraging **integrated Ethernet on FPGAs**, instead of using PCIe for data transmission.



FPGAs right now



FPGAs we need

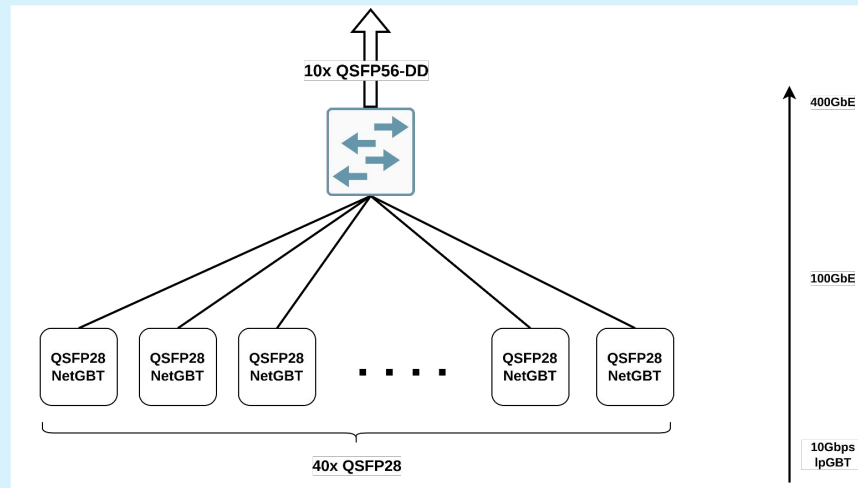
Proposed Solution: NetGBT

Our solution is a **smart media converter** based on a **mid-end FPGA** with many transceivers available.

The media converter is capable of interfacing with a number of **IpGBT links** and, with little processing, convert them into **UDP/IP packets**.

The Ethernet standard makes the design **highly flexible**:

- Supports **direct connection** to a network interface for testbench and small test beam setups
- **Scalable for large-scale deployments**, where multiple converters can be aggregated using consumer off-the-shelf (COTS) Ethernet switches.



A network example using 100GbE-capable NetGBT.

Proof of Concept: Hardware

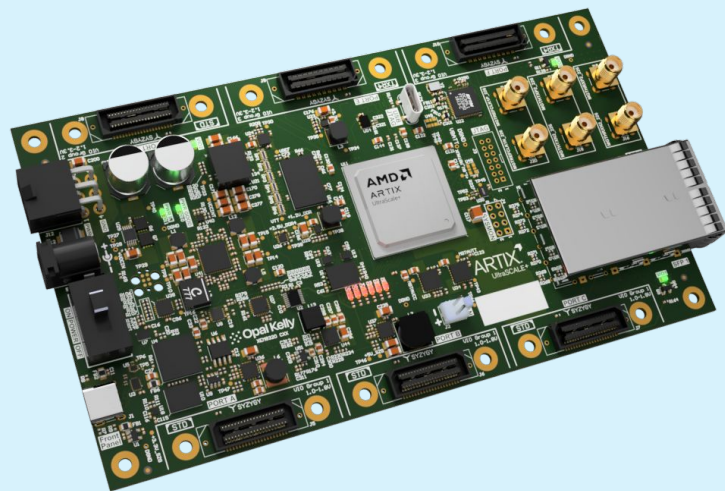
The hardware chosen for the Proof of Concept is an **AMD Artix UltraScale+ AU25P**.

This choice was made due to the availability of an **off-the-shelf development kit** which makes all transceivers available.

This board can implement:

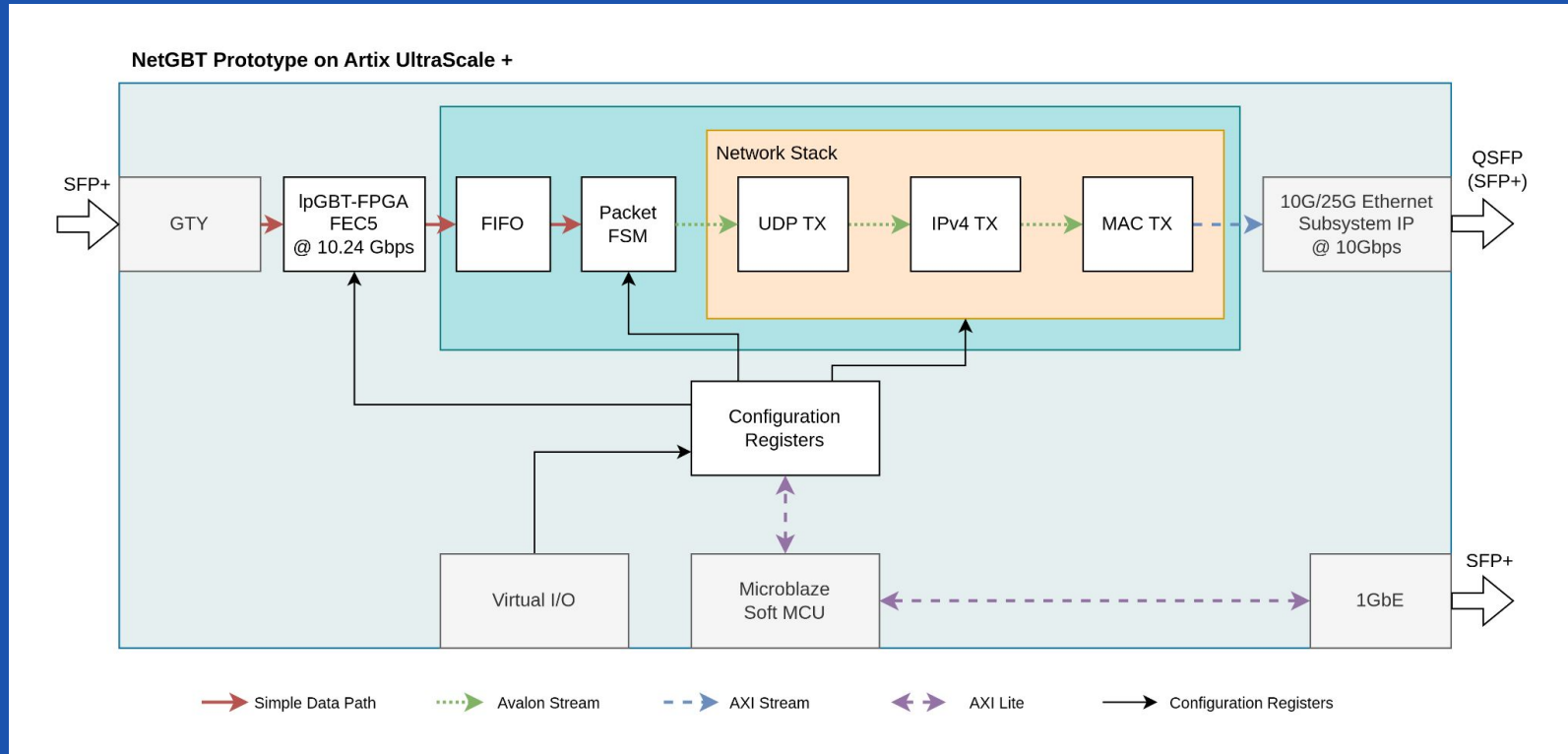
- Up to **4 lpGBT links @ 10.24 Gbps**
- Up to **4 Ethernet links at 10 Gbps**
- A dedicated 1 Gbps Ethernet link for configuration

The board however does not offer 25 Gbps transceivers (GTY), which could be used to implement a 100GbE link.



Opal Kelly XEM8320 Development Kit

Proof of Concept: Gateway



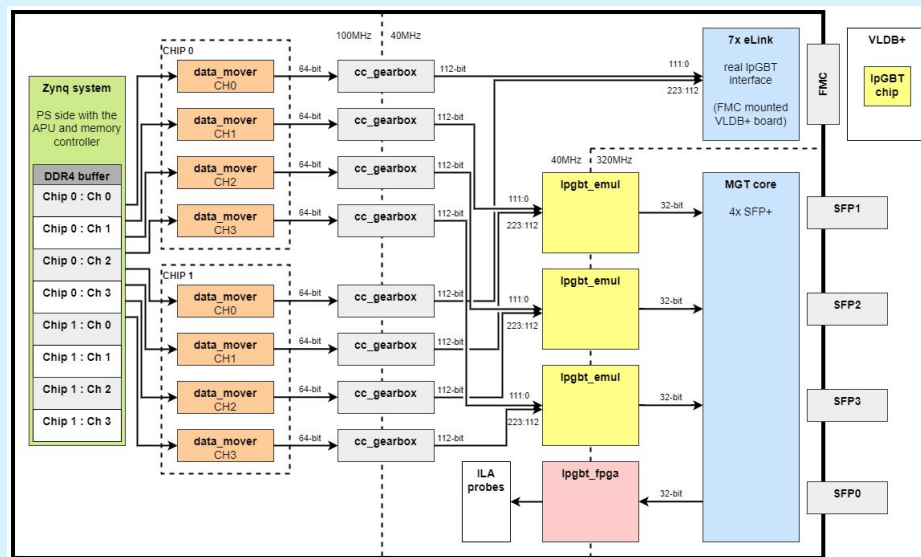
This board (Artix US+) can handle up to 4 IpGBT links, which correspond to a full VTRx+, and a full 4 lane QSF+ (4x10GbE) on the Ethernet side.

Testbench: Stimuli Generator

The prototype has been tested using a **Zynq Ultrascale+ development kit (ZCU102)** together with the **VLDB+** evaluation board for the IpGBT ecosystem.

The Zynq acts as an **FE emulator** sending a predefined sequence of data that is loaded into the DDR4 RAM.

The system is capable of generating up to **5 IpGBT links**: one using the real IpGBT chip on the VLDB+ and the other four emulating the IpGBT chip.

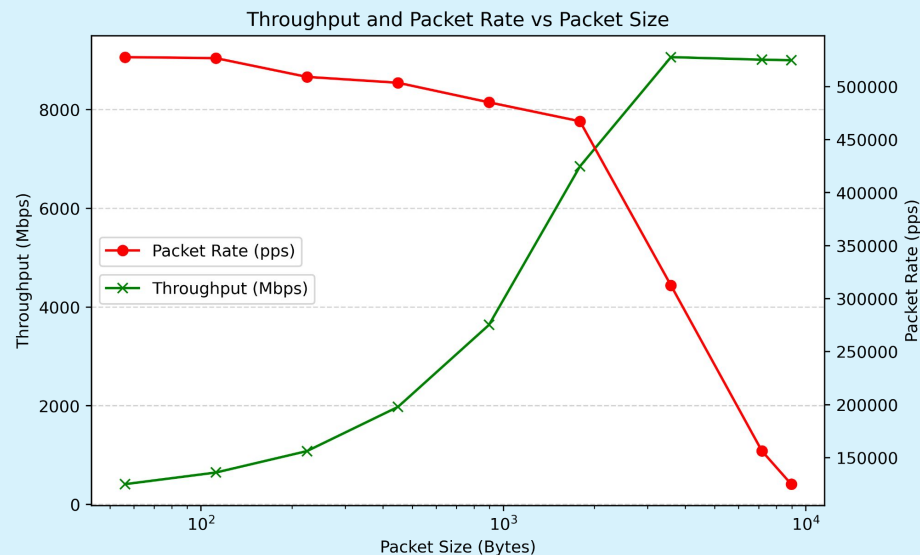


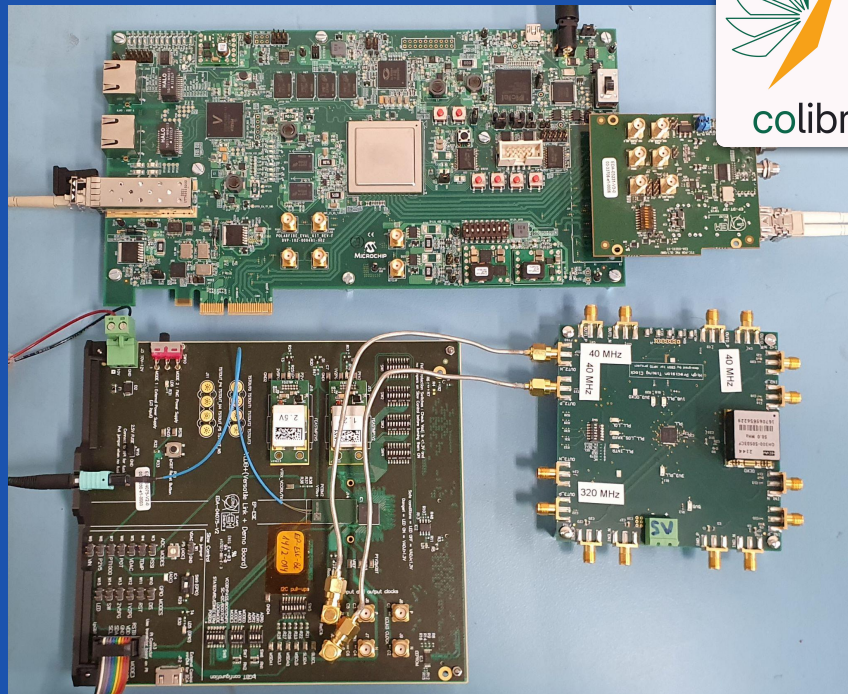
Testbench: Results

Measurements were taken to evaluate which **packet size** is optimal to **reach peak throughput** and to avoid back pressure.

The NetGBT was connected directly to a **10Gbps** Network Interface via a Direct Attach Cable.

Benchmarks show that the point-to-point connection is saturated when **packet sizes are more than 4 kB**.





Vendor Independence

The gateway has been designed to be **vendor independent** using the *colibri* library, enabling testing of different architectures to find which one fits the best.

The vendor independence of the design has been proven by the colleagues of CERN EP-ESE, who ported the NetGBT gateway to a Microchip FPGA Development Kit.

colibri supports:

- AMD Vivado
- Intel Quartus
- Lattice Radiant
- Microsemi Libero
- Efinix Efinity
- Gowin EDA
- Aldec Riviera Pro
- Mentor Modelsim
- NVC
- GHDL

Credits to Valentin Stumpert of EP-ESE, check his [poster!](#)

Check the [talk on colibri](#) at 5:20 PM

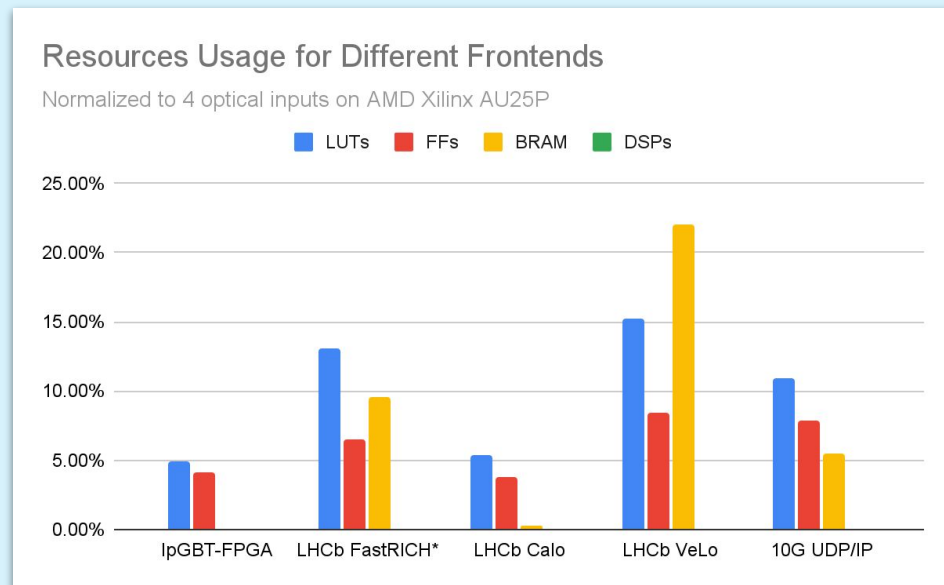
Data Processing

The current gateway does **only packing** of the Front-End data, using a small amount of resources.

We evaluated the footprint of **different data processing** elements on the FPGA to understand how much can be offloaded on the device.

Both simple (CALO) and complex (VeLo) data processing can be fitted on the FPGA.

Key consideration: balance the ratio between resources and transceivers to select the optimal device.



**FastRICH only implements Aurora 64b/66b decoding*

Prototype Development: Phase 2

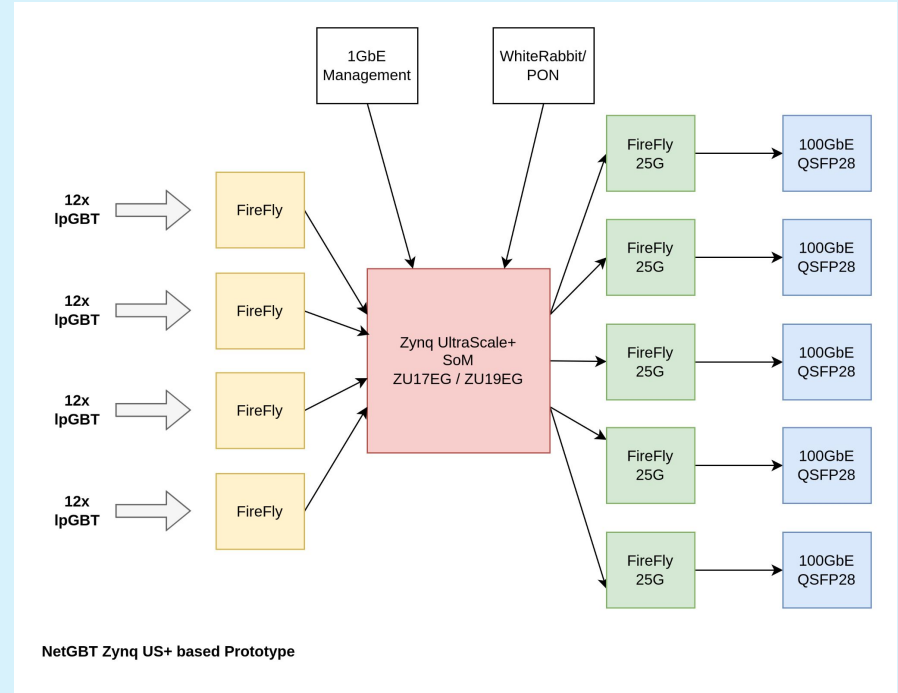
Building on the promising results from the proof of concept, we are advancing to **Phase 2** of the prototype.

This prototype will utilize a **System on Module (SoM)** based on Zynq UltraScale+, featuring **25G capable transceivers**.

The device will support processing of up to **48 IpGBT links** and forwarding via **5 x 100GbE uplinks**.

Multiple SoMs can be housed in a **1U 19"** rack-mounted box.

This solution will further **reduce costs** and provide **high link density** per rack unit.



Conclusions and Future Directions

The Proof of Concept presented showed **promising results**:

- Successful **conversion of IpGBT to UDP/IP** with no back pressure
- **Low resource utilization**, leaving space for data processing offload
- **Cost effectiveness**, using an off-the-shelf mid-end FPGA development kit
- **Flexibility**, thanks to the standard ethernet uplink
- **Modularity**, links can be aggregated using COTS switches

Future developments:

- **Aggregate links** over faster transceivers (Artix UltraScale+ is not capable of 25 Gbps)
- Use an FPGA with **more transceivers** to improve efficiency, cost effectiveness, and to enable more complex data processing

Thanks for the attention

Special thanks to:

- LHCb Online Team
- CERN EP R&D WP 9.3
- ECFA DRD 7.5b