Published by IOP Publishing for Sissa Medialab

Received: *October 31, 2024*
Accepted: *January 23, 2025*
Published: *February 13, 2025*

Topical Workshop on Electronics for Particle Physics
University of Glasgow, Scotland, U.K.
30 September–4 October 2024

A low-cost, low-power media converter solution for next-generation detector readout systems

A. Perro[a,b,*] **M. Vodnik**[a,c] **and P. Durante**[a]

[a] *CERN, Geneva, Switzerland*
[b] *Aix Marseille Université, CNRS/IN2P3, CPPM, Marseille, France*
[c] *Jozef Stefan Institute, Ljubljana, Slovenia*

E-mail: alberto.perro@cern.ch

Abstract: High Energy Physics (HEP) data acquisition systems are often built from high-end FPGAs. As such systems scale in the HL-LHC era, severe under-utilization of FPGA transceivers can occur because front-end links prioritize radiation hardness and power consumption over raw data bandwidth. This work evaluates recently introduced low-power, low-cost FPGA devices as an alternative building block for future readout architectures. This study presents the implementation of a readout back-End on FPGA where the front-end protocol is based on the Low-Power GigaBit Transceiver (lpGBT) and the readout protocol is based on 10 Gigabit Ethernet, using the LHCb Run 4 RICH detector as a practical case study.

Keywords: Data acquisition circuits; Data acquisition concepts; Modular electronics; Online farms and online filtering

ArXiv ePrint: 2410.23173

*Corresponding author.

https://doi.org/10.1088/1748-0221/20/02/C02027

Contents

1 Introduction

LHCb readout system in Run 3. The Data Acquisition (DAQ) system of the LHCb experiment [1] currently employs 11,000 GigaBit Transceiver (GBT) [2] links at 4.8 Gbps. The readout system and the Event Builder can handle a cumulative throughput of 32 Tbps. These links are read by custom high-end FPGA boards which aggregate, process, and transfer the data to the Event Builder via PCI Express (PCIe). Each of these FPGA boards has 1.2M Logic Elements, it is capable of handling up to 48 GBT links, and it offers two PCIe 3.0x8 interfaces to the host machine. Of 520 cards employed in LHCb, 445 of them are used for data readout and the rest for control and clock distribution.

The readout cards are hosted in groups of up to three in the Event Building servers. This implies that the data of each event is fragmented over the entire cluster. Data fragments from all sub-detectors have to be collected so that single events can be assembled in the same place to proceed with the reconstruction and selection process. Servers in the Event Building cluster are interconnected by a fast network in order to send and receive data fragments. This network uses the InfiniBand HDR 200 Gbps technology. The choice of this technology over the Ethernet standard was motivated by the performance difference measured for this specific use case at the time of design [3].

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

Upgraded sub-detectors will require both a faster data rate, provided by Low Power GigaBit Transceiver (lpGBT) technology [4] reaching up to 10.24 Gbps per link, as well as a higher number of optical links. DAQ systems have to be upgraded accordingly to manage the higher throughput. LHCb estimates a total throughput of 300 Tbps and 30,000 readout links in Run 5 [5], 9.4x the throughput and 2.7x the links compared to Run 3.

The DAQ upgrade offers the possibility to design the EB network once again: Ethernet is evolving at a fast pace, with decreasing cost per bandwidth and can be sourced from multiple competing vendors. Unlike InfiniBand, FPGA IP modules are also readily available. Ethernet ASICs, at the time of writing, are capable of switching up to 50 Tbps in a single chip — almost 2x the current LHCb throughput — and throughput is expected to double every two years [6].

FPGA market trends. The FPGAs currently used in the LHCb readout system provided a sufficient number of transceivers, transceivers' speed, and logic resources at the time of design. However, current high-end FPGAs are moving towards offering fewer transceivers that support very high data bandwidths (up to 112 Gbps) [7]. FEEs in HEP experiments prioritize radiation hardness and power consumption at the expense of link rates. This study wants to evaluate the potential of lower-end FPGAs as a cost-effective DAQ platform and to investigate how Ethernet can simplify the DAQ architecture by integrating it on the FPGAs.

2 Proposed solution

Design principles. The proposed design — named *NetGBT* — is guided by the following principles:

1. Conversion from custom radiation-hard protocols, such as lpGBT, to a standard network protocol like Ethernet.
2. Selection of an FPGA featuring an optimal price-to-transceiver ratio, possibly with a high transceiver count.
3. Support for 10 Gigabit Ethernet (10GbE), with a preference for 25 Gigabit Ethernet (25GbE) where feasible.
4. Sufficient FPGA resources to decode and aggregate existing and upcoming front-end data protocols.

The use of the Ethernet standard makes the design highly flexible, supporting both direct connections to a network interface card (NIC) for testbenches and small test beam setups, while also scaling efficiently for large-scale deployments. In the latter case, multiple Ethernet uplinks can be aggregated using consumer off-the-shelf (COTS) Ethernet switches.

On top of Ethernet, the Internet Protocol (IP) and the User Datagram Protocol (UDP) Lite are used. These protocols are standard protocols in modern networks and they are supported by every network device. UDP-Lite is a simple message-oriented protocol based on UDP. It has been designed for use in scenarios where error tolerance is acceptable or built in the encapsulated payload. It finds its use in real-time application such as multimedia streaming. The absence of a full packet checksum simplifies the design of the gateware and streamlines the data processing on the receiver side.

3 Proof of concept

Hardware. The Proof of Concept is based on an off-the-shelf development kit from Opal Kelly [8], which hosts the AMD Artix Ultrascale+ AU25P FPGA. This FPGA features 12 GTY transceivers capable of speeds up to 16.3 Gbps. The baseboard provides access to 4 transceivers via two SYZYGY XCVR connectors [9], 2 transceivers connected to SFP+ cages, and the remaining 2 transceivers routed to SMA connectors. This configuration allows for the handling of up to 4 lpGBT links, which can be converted into 4 10GbE links using two SYZYGY QSFP+ mezzanine cards [10]. Additionally, one of the transceivers in the SFP+ cages is allocated for a 1 GbE link, which is used for configuration and control purposes.

However, the selected device does not include 25 Gbps capable transceivers, making it unsuitable for testing the aggregation of multiple lpGBT links onto one 25 GbE output.

2025 JINST 20 C02027

Gateware. The Proof of Concept gateware is designed to handle each lpGBT link separately. The lpGBT links are received using the GTY transceiver hard IP and then decoded using the lpGBT-FPGA core configured for the highest data rate (FEC5, 10.24 Gbps). The decoded words are 224 bit wide and are available at 40 MHz, resulting in a link goodput of 8.96 Gbps. Once the payload is available, a mixed-width FIFO is used to store the incoming data and do the clock crossing from the 40 MHz lpGBT clock to the 156.25 MHz Ethernet clock. The data in the FIFO is read in chunks to form multi-word packets: large packets are necessary to make efficient use of the network bandwidth. Once the packets are formed, the network stack adds the headers for UDP-Lite, IP, and Ethernet. Packets are then forwarded to the 10G/25G Ethernet Subsystem IP which transmits the data to the back-End.

The transmission data path from the FIFO to the MAC header core has been designed using the vendor-independent open source common core library *colibri* [11]. Thanks to the use of this library, the gateware has also been ported to a Microchip-based development kit with little adjustments.

Configuration of the system is done via configuration registers which are accessible both via Virtual I/O over JTAG and via a Microblaze Soft Microcontroller, which hosts an MQTT (Message Queuing Telemetry Transport) client to access the registers via the 1 GbE management link. MQTT is a lightweight low-latency protocol designed for communication between devices and is widely used for remote monitoring, telemetry, and data collection applications.

4 Front-end emulator

The emulator platform. A font-end emulator has been developed separately to support the development of next generation back-end electronics, by emulating the output of various detector front-end electronics on FPGA gateware. The emulator is based on the AMD Zynq Ultrascale+ System on Chip, hosted on the ZCU102 evaluation board. The board includes 4 SFP+ cages connected to a Quad of GTH transceivers located on the Zynq chip. Each of the 4 SFP+ ports can be used to emulate one full-speed lpGBT interface. The ZCU102 board also includes an FMC connector which is used to mount a VLDB+ evaluation board, hosting a single lpGBT ASIC chip, which serves as an additional front-end interface.

FastRICH emulator. One emulator variant is designed to emulate the data output of 7 instances of the next generation RICH detector readout ASIC: the FastRICH. Two main parameters are available to implement different front-end configurations: number of serial lanes and lane speed. The number of lanes can be chosen from 1 to 4, while lane speed can be set to either 320, 640 or 1280 Mbps. The simulated output data is loaded onto the on-board DDR4 RAM component, and can be continuously streamed via 7 independent channels, each with its Aurora 64B/66B encoder. Depending on the lane configuration, up to 28 serial lanes can be produced, which are distributed over 4 lpGBT optical links. One of the links utilizes the physical lpGBT chip, while the other 3 emulate the lpGBT on FPGA and stream via 3 of the 4 SFP+ ports.

CALO emulator. The CALO emulator variant is designed to provide maximum data throughput with just basic data formatting. It is inspired by the LHCb Run 3 Calorimeter readout. It emulates 2 FEE chip instances, each streaming data via 4 channels. The data is formatted into frames with a constant length of 14-bytes, or half of the lpGBT payload. The two instances share 4 lpGBT links, streaming each of the 4 channels on a separate optical fiber. Like in FastRICH emulator, the data is streamed from the RAM component and passed onto the same 4 optical interfaces.

2025 JINST 20 C02027

5 Evaluation

Data processing. The described gateware uses a small amount of the resources available (table 1). This allows the implementation of some FEE-specific data processing on the FPGA, offloading some tasks from the high-level trigger software. Resource utilization results from current and future sub-detector data processing blocks have been measured and normalized to the selected device (figure 1).

Table 1. Resource utilization for 4 lpGBT links on AMD AU25P.

	LUTs	LUT RAM	FF	Block RAM
Utilization	20056	298	34500	34
Utilization (%)	14.22	0.30	12.23	11.33

The results indicate that simple front-end data processing involving bit manipulation, error management, and packing (like LHCb CALO) could easily fit on the device. More complex data processing tasks, such as the clustering algorithms employed in the LHCb VeLo, could also fit within the available resources. Additionally, the upcoming FastRICH sub-detector link decoder can also be offloaded onto the device; however, this implementation covers only a portion of the complete data processing as the final dataflow is not yet implemented.

Throughput. Measurements were taken to establish the minimum packet size required to reach the required throughput while avoiding back pressure, packet drops, and subsequent data loss. The host system is based on a AMD Threadripper 2990WX 32-Core CPU, 64 GB DDR4 RAM, and a Mellanox ConnectX-6 NIC. The OS used is AlmaLinux 9.2 with the 5.14 kernel. The FPGA was connected directly to the NIC via a Direct Attach Copper (DAC) cable. Some parameters were tuned to ensure best UDP performance in terms of packets per second (PPS):

- Enabling Jumbo Packets on the NIC (MTU set to 9000)
- Increase the network buffer sizes
- Configure `iptables` to not track and filter packets for the destination UDP port

The benchmarks measured the throughput and the packet rate by taking 10 samples of 1 second each to reduce the impact from external factors.

The results show that the point-to-point connection reaches the maximum achievable throughput of (9064 ± 2) Mbps with the packet size of 3584 B, the correspondent packet rate is $(312,490 \pm 54)$ PPS (figure 2). Therefore, 4 kB is used as the lower bound for packet sizes in this application.

6 Achievements and future directions

The Proof of Concept demonstrated several promising results, specifically the successful conversion of lpGBT to UDP/IP without back pressure. The implementation utilizes few resources, which leaves ample space for additional data processing offload. Furthermore, the solution proved to be cost-effective by using an off-the-shelf mid-range FPGA development kit at a fraction of the cost-per-link compared to high-end FPGAs. The adoption of a standard Ethernet uplink contributes to its flexibility and modularity, specifically the link aggregation using consumer off-the-shelf (COTS) switches.

2025 JINST 20 C02027

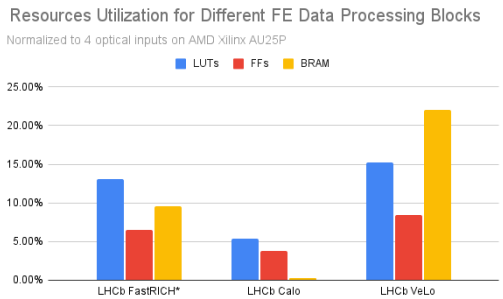

Figure 1. Resource Utilization for different data processing blocks, normalized for 4 lpGBT links on the AMD AU25P FPGA. Note that the FastRICH does not implement the full data processing.

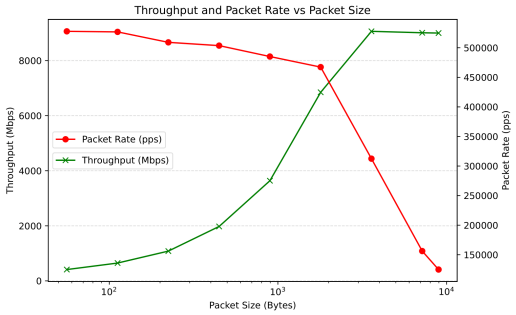

Figure 2. Throughput and packet rate measurements for a single lpGBT link converted to a single 10GbE link.

A second prototype is under development. This next iteration will utilize a System on Module (SoM) based on the AMD Zynq UltraScale+, featuring 25Gbps-capable transceivers. The new device is designed to support the processing of up to 48 lpGBT links and output through 5 100GbE uplinks. Multiple of these SoMs could be housed in a 1U rack-mounted enclosure, optimizing space efficiency.

Cost evaluation. The next iteration of the proposed solution will support 48 lpGBT links. The selected System-on-Module (SoM) is the Enclustra XRU90, priced at $4,123 [12] for a single unit, with potential cost reductions at higher volumes. For optical transceivers, a 12-link Samtec FireFly™ module is listed at $920 from the distributor [13], resulting in a cost of $77 per link. By comparison, a COTS Ethernet transceiver is priced at $40 [14], resulting in $10 per link, suggesting the realistic cost for optical links can be around $43 per link. Using receive-only transceiver modules could further reduce costs to about one-third of this value. Additionally, 100GbE transceivers for the uplink are priced at $43 [15].

The SoM integrates most of the board's cost, and using multiple SoMs on a shared baseboard allows for the distribution of expensive components, such as low-jitter PLLs and voltage regulators. Based on these factors and the pricing of existing SoM baseboards, an additional $500 per SoM is estimated for baseboard costs.

This evaluation highlights the cost-effectiveness of the proposed solution, achieving a total cost of $142 per lpGBT link (table 2).

Table 2. Estimated cost of the NetGBT solution capable of handling 48 lpGBT links. All prices are in $USD.

FPGA SoM Cost	4100
lpGBT Transceivers Cost	2000
100GbE Transceivers Cost	200
PCB Cost	500
Total Board Cost	6800
Cost per lpGBT link	**142**

2025 JINST 20 C02027

Acknowledgments

The authors would like to acknowledge the support from the CERN LHCb Online Team, CERN EP R&D WP 9.3, and ECFA DRD7.5b collaborations.

References

[1] F. Pisani et al., *Design and Commissioning of the First 32-Tbit/s Event-Builder*, *IEEE Trans. Nucl. Sci.* **70** (2023) 906.

[2] P. Moreira et al., *The GBT Project*, in the proceedings of the *Topical Workshop on Electronics for Particle Physics*, Paris, France, 21–25 September 2009, [DOI:10.5170/CERN-2009-006.342].

[3] R.D. Krawczyk et al., *Feasibility tests of RoCE v2 for LHCb event building*, *EPJ Web Conf.* **245** (2020) 01011.

[4] P. Moreira, *The lpGBT: a radiation tolerant ASIC for Data, Timing, Trigger and Control Applications in HL-LHC*, in the proceedings of the *Topical Workshop on Electronics for Particle Physics*, Santiago de Compostela, Spain, 2–6 September 2019 [https://indico.cern.ch/event/799025/contributions/3486153/].

[5] T. Colombo, *Online Infrastructure Design and Performance*, in the proceedings of the *LHCb Upgrade Electronics Workshop*, June 2024, [https://indico.cern.ch/event/1411258/].

[6] N. Margalit et al., *Perspective on the future of silicon photonics and electronics*, *Appl. Phys. Lett.* **118** (2021) 220501.

[7] Intel Corporation, *Agilex™ 7 FPGA and SoC FPGA M-Series*, https://www.intel.com/content/www/us/en/products/details/fpga/agilex/7/m-series.html.

[8] Opal Kelly, *XEM8320*, https://opalkelly.com/products/xem8320/.

[9] Opal Kelly, *SYZYGY*, http://syzygyfpga.io/.

[10] Opal Kelly, *SZG-QSFP Mezzanine*, https://docs.opalkelly.com/syzygy-peripherals/szg-qsfp/.

[11] A. Perro, *COLIBRI: Towards a CERN-wide common cores library*, in the proceedings of the *1st FPGA Developers' Forum (FDF)*, CERN, 11–13 June 2024 [https://indico.cern.ch/event/1381060/contributions/5923223/].

[12] Enclustra GmbH, *Andromeda XZU90 AMD Zynq™ UltraScale+™ System-on-Module (SOM) ZU17EG*, December 2024, https://www.enclustra.com/en/products/system-on-chip-modules/andromeda-xzu90/.

[13] Mouser Electronics, *ECUO-Y12-14-040-0-1-1-2-21 Samtec*, December 2024, https://eu.mouser.com/ProductDetail/Samtec/ECUO-Y12-14-040-0-1-1-2-21?qs=yVhe5xptBcaadGCKNMU3gA%3D%3D.

[14] SFPcables.com, *40GBASE-SR4 QSFP+ Module*, December 2024, https://www.sfpcables.com/40gbase-sr4-qsfp-transceiver-for-mmf-100-150-meters-mpo-mtp.

[15] SFPcables.com, *100GBase-SR4 QSFP28 Transceiver*, December 2024, https://www.sfpcables.com/100gbase-sr4-qsfp28-transceiver-for-mmf-70-100-meters-mpo-mtp-4813.

2025 JINST 20 C02027