Versatile firmware for the Common Readout Unit (CRU) of the LHC ALICE experiment

O. Bourrion[∗]**, J. Bouvier**

Univ. Grenoble Alpes, CNRS, Grenoble INP[†]*, LPSC-IN2P3, 38000 Grenoble, France*
† *Institute of Engineering Univ. Grenoble Alpes*

F. Costa

CERN Route de Meyrin, Geneva, Switzerland

E. Dávid, J. Imrek, T.M. Nguyen

Wigner Research Centre for Physics, 29-33 Konkoly-Thege Miklós str,
H-1121 Budapest, Hungary

S. Mukherjee

Bose Institute, Kolkata 700 054 WB, India

For the third running period of the CERN LHC, the ALICE experiment will use a Common Readout Unit (CRU) at the heart of the data acquisition system. The CRU, based on the PCIe40 hardware designed for LHCb, is a common interface between the front-end, the computing system, and the trigger and timing system. The 475 CRUs will interface 10 different sub-detectors with 3 sub-systems and reduce the total data throughput from 3.5 TB/s to 635 GB/s. The ALICE common firmware framework is under development. It supports data taking in continuous and triggered mode, clock, trigger and slow control delivery. In this paper, the architecture and results are presented.


[∗]Speaker.

arXiv:1910.08804v1 [physics.ins-det] 19 Oct 2019

1. Introduction

For the third running period of the CERN LHC, most of the detectors from ALICE will be upgraded to be able to operate in continuous readout mode. However, triggered readout will still be used by some detectors, and also for commissioning and calibration runs. The experiment will produce 3.5 TB/s of data, and thus to cope with this new scheme and with this demanding throughput the 10 upgraded detectors will use the Common Readout Unit (CRU). Faithful to the design reuse strategy for the LHC experiments, the PCIe40 electronics designed for the LHCb experiment [1] was used to be the CRU of ALICE. With the hardware being the same, a common firmware was developed for the CRU such that the development efforts between sub-detectors was shared.

The CRU is the interface between the detector front-end electronics, the O^2 facility, the Detector Control System (DCS) and the Trigger and Timing System (TTS), see Fig. 1.
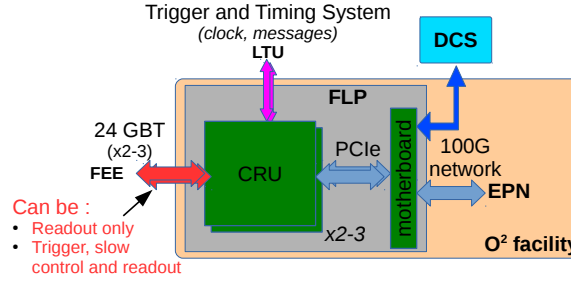

Figure 1: The CRU is the interface between the detector front-end electronics, the O^2 facility, the DCS and the TTS via the Local Trigger Unit (LTU).

The O^2 facility is a computer farm composed of the First Level Processors (FLP) and Event Processing Nodes (EPN). The FLPs gather information from the FEE via the CRU and communicate with the EPN though the 100GbE network. Each FLP, which is in fact a server, hosts two or three CRUs. The CRU communicates to and from the FEE via optical fibers (up to 24). Depending on the sub-detector considered, the optical fiber can be used for readout only, and/or for trigger and timing, and/or for slow control. For most of the sub-detectors, the GBT protocol [2, 3] is used. The communication with the TTS is done via a dedicated bidirectional Passive Optical Network (PON), it is time multiplexed in the upstream direction. It allows the reception of the LHC machine clock and to exchange messages. The CRU is connected to the server's motherboard via PCIe gen 3 x16. The readout and the detector control messages are passed through this interface.

2. Continuous readout

As stated previously, in Run 3 ALICE will operate in continuous readout mode. In practice the sub-detectors will provide a continuous stream of data. For event reconstruction at the EPN level, it was decided to have the continuous stream sliced in Time Frames (TF) of 22 ms. The TF are then divided in 256 Heart Beat Frames (HBF) of one orbit duration (89.4 µs). In this scheme, the task of the CRU is to collect the data continuously and to check the successful HBF reception in each FLP. For each HBF, *an acknowledge* (HBACK) or *not acknowledge* message (HBNACK) is delivered

to the Central Trigger Processor (CTP) which needs to assess the quality of the TF reception. In the case the latter is not satisfactory, the CTP can request that all CRU of the experiment drop the remaining HBF of a TF by sending a Heart Beat Reject (HBr) command, which allows the system to recover for the subsequent Time Frame. An example of two incomplete time frame transmissions is shown in Fig. 2.
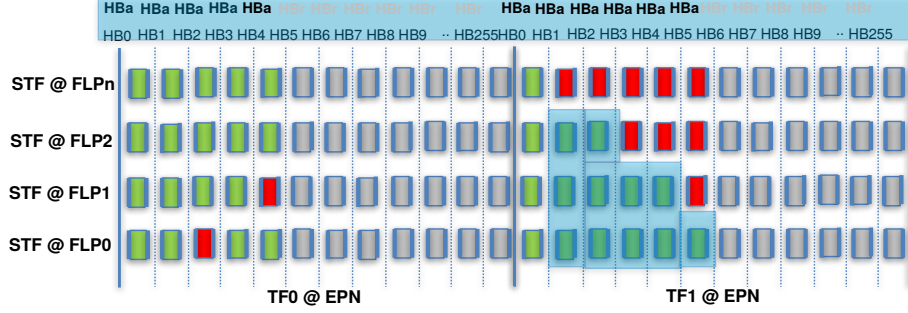
Figure 2: Illustration of the continuous readout (extracted from [4]). Each rectangle represents a Heart Beat Frame (green: successfully received; red: bad reception or missing fragment; grey: deleted HBF). There are 256 HBF per Time Frame (TF). In this example, the remaining HBFs of a TF are rejected after a bad reception is detected. HBa: Heart Beat accept; HBr Heart Beat reject.

3. Requirements

The CRU common firmware was developed to fit the ALICE needs. In particular, it was decided (i) to share the common features and interfaces (PCIe, trigger and timing, and GBT protocol), (ii) to provide the possibility to read-out all detectors in 'raw-mode' (that is with no data processing in the CRU), (iii) to allow reference clock and trigger signals distribution and (iv) to permit FEE configuration. Additionally, some sub-detector specific features called 'user-logic' can be integrated in the common firmware via a specific compilation. This 'user-logic', which is an online data processing, can consist of features such as baseline correction, zero suppression, or others. A requirement of ALICE is to be able to switch between 'raw-mode' and 'user-logic' at any moment. The last important requirement is to include many self-testing capabilities in order to ease commissioning and system maintenance.

The main flavor variations of the common firmware come from the GBT link mode used to communicate with the front-end and the information it is supposed to carry, the integration of a 'user logic' or not, and the type of slow control required to configure the FEE.

Indeed the GBT link can be either operated in 'GBT-mode' (80 bits of payload and forward error correction) or in 'wide-bus' (112 bits of payload and no forward error correction). In 'GBT-mode' the data can be produced in stream (one data transferred at each machine clock cycle) or already in packet, while in 'wide-bus' only stream is possible. The GBT-link may be used to send trigger messages or the reference clock only to the FEE, while upstream it is used for data readout and optionally to acknowledge specific slow control transaction. At this stage of the project, the requirements of the various detectors can be summarized in the table shown in table 1.

2

Detector	Number of CRU	TX/RX per CRU	User Logic	GBT with FEC ("gbt-mode")	TTS through CRU	DCS through CRU
CPV *(Charged Particle Veto)*	1	24/24	N	Y	Y	Y (SCA)
FIT *(Fast Interaction Trigger)*	2	24/24	N	Y	Y	N
ITS *(Inner Tracker System)*	26	12/24	TBD	Y	clock only	Y (single word transfert)
MCH *(Muon Tracking Chamber)*	32	24/24	Y	Y	Y	Y (SCA)
MFT *(Muon Forward Tracker)*	11	12/24	N	Y	clock only	Y (single word transfert)
MID *(Muon Identifier)*	2	24/24	TBD	Y	Y	Y (SCA)
TOF *(Time Of Flight)*	4	24/24	TBD	Y	Y	Y (SCA)
TPC *(Time Projection Chamber)*	360	24/24 (50%) 0/24 (50%)	Y	N	Y	Y (SCA)
TRD *(Transition Radiation Detector)*	37	0/36	Y	not GBT	N	N
ZDC *(Zero Degree Calorimeter)*	1	24/24	N	Y	Y	N

Table 1: Table summarizing the requirements of the various detectors. In white features not required; in blue compatible requirements with an adapted common firmware; in green and orange features requiring a specific bit file generation.

4. Firmware description
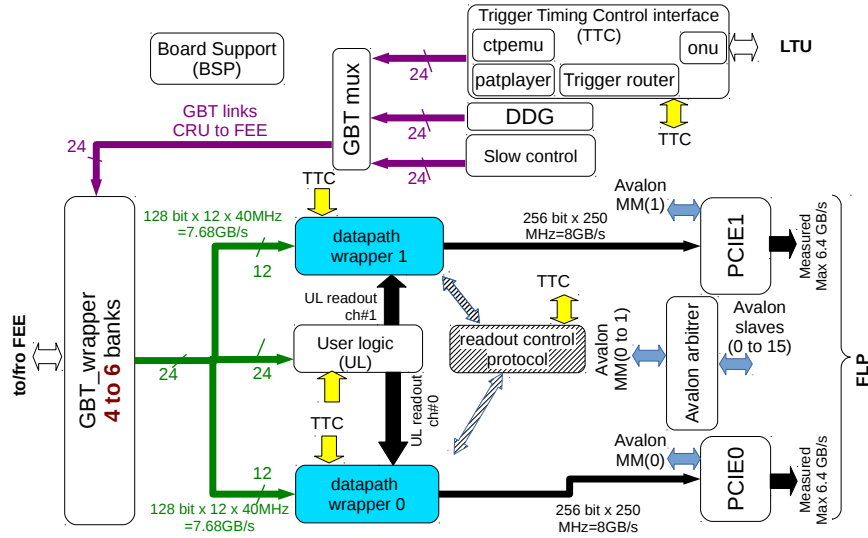
Figure 3: Overview of the common firmware. The main parts are shown: the GBT interface, the Trigger and Timing Control interface, the data path and the PCIe endpoints.

An overview of the firmware is given in Fig. 4. The main parts are shown: the GBT interface, the Trigger and Timing Control interface, the data path and the PCIe endpoints. Starting from the front-end side on the left, the 'GBT_wrapper' interface is shown, it is the interface with the FEE. The 'GBT_wrapper' is a forked version of the GBT-FPGA developed at CERN [2, 3]. The main differences are that (i) it has a user data path operating at 240 MHz (six times the machine clock), (ii) a dynamic switching is possible between 'GBT-mode' and 'wide-bus' to cover more use cases, (iii) the clock domain crossing between the transceiver domain and the user part is achieved thanks to timing constraints (no phase scanning is required at link startup) and (iv) the test data pattern generator is shared between all links to save resources. Moreover, the 'GBT_wrapper'

allows external (with optical fibers) and internal (inside the FPGA transceivers) loop-back modes that allows respectively to validate the CRU-FEE communication and to validate CRU proper data path operation once installed in the system. Indeed, as it will be shown later, the 'data generator' allows the emission of representative data toward the FEE that can be looped back by them in the CRU. This feature is nice to have in order to be able to stress the system in-situ.

On the downstream path (CRU to FEE), depending on the detector requirements or test requirements, several sources can be selected to supply the 'GBT_wrapper'. These are the Trigger and Timing Control (TTC) interface, the Dedicated Data Generator (DDG) or the slow control.

The TTC interface is composed of four components. The first is the Optical Network Unit (ONU) [5, 6] that allows the CRU to recover the machine clock from the PON and to forward it through the GBT to the FEE. The ONU is also used to receive from the central system the 200 bits of the trigger and timing message at each clock cycle (trigger bits, bunch crossing number, Heart Beat ID, ...). The CRU uses the upstream direction to send the HBACK or HBNACK message that is composed of 56 bits. The second component is a trigger emulator ('ctpemu') that is used for tests and system diagnostic purposes. It can produce trigger messages, like the ones provided through the ONU, and simulate readout flow control by producing HBa and HBr commands. The third component is the 'pattern player' that can generate a programmable sequence to be transmitted to the FEE, it is fired by a trigger bit issued either by the 'ONU' or 'ctpemu'. This feature is used to generate a specific control sequence required by the readout ASIC used by the TPC and MCH detectors [7, 8]. The fourth is the trigger router which is used to replicate and reroute the trigger bits to several positions of the downstream GBT board (replicate bits on various e-links). This feature is used by the Muon Identifier (MID).

The DDG can be dynamically configured to produce either streaming or packet type data. The packet data can be produced with fixed or random packet length and inter packet duration. With this powerful tool, the DAQ system and the data flow of the CRU can be stressed in real and large scale conditions.

Two 'datapath_wrapper' are implemented in the firmware, they receive the trigger messages, collect and aggregate the FEE data provided by the GBT link or use directly the 'user logic' input if required to and finally provide some monitoring information to the 'readout control protocol' component (see Fig. 4). The first task of each 'datapath_wrapper' is to receive in parallel the data from up to 12 GBT bus, one user logic link and of the 'readout protocol' component (trigger acknowledge or decision message). The 'gbt_datapathlink' is designed such as to be able to receive the data in stream or packet format. When selecting the stream mode this component construct data packets, i.e. it chops the data stream and inserts the Reduced Data Header (RDH). The RDH is 64 bytes long and contains information that is useful for the readout protocol, such as the Heart Beat ID (HBID), the Link ID, the page counter and the Stop bit. The last two parameters indicate for each Link ID the number of the packet within the HBF and if the packet is the last of the HBF. The 'ul_datapathlink' receive already formed packets. At the output of this first stage only packets having a maximum size of 8 kB and a RDH are available. Then the second stage performs data aggregation, i.e. it scans in a round robin manner the possible data sources and collects data packets. At the output of this stage, the packets from the various links are interleaved. This is followed by the removal of all packets from the data flow, if required by the CTP, with a HBr message (number 4 in Fig. 4). Then, the packets are stored in the 'bigfifo' (a large buffer) to be
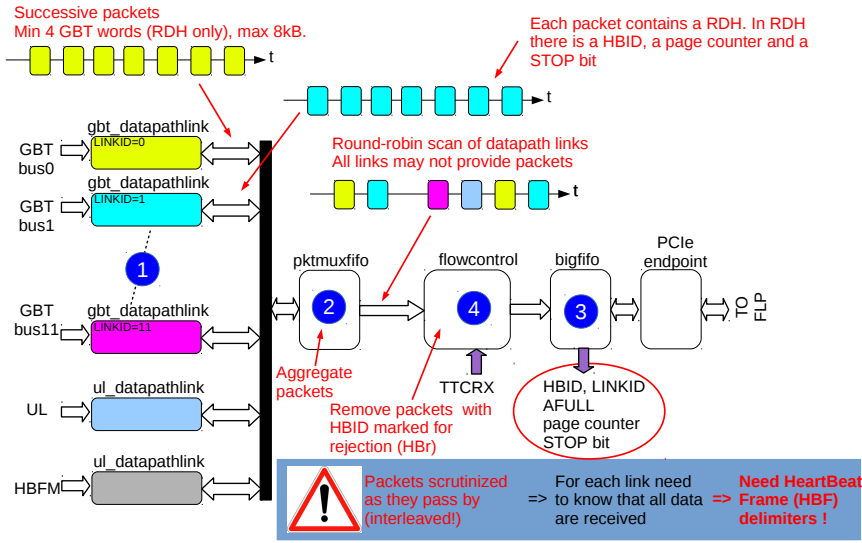
Figure 4: Overview of the 'datapath_wrapper'.

made available to the PCIe endpoint. Note that while being stored, the packets are scrutinized and useful parameters (HBID, LINKID, fifo status, etc) are presented to the 'readout control' protocol component.

The 'readout protocol' uses the information provided by both 'datapath_wrapper' to check the interleaved packets as they are flying-by. The HBF reception is declared successful only if for each LINKID included in the readout, start (page counter of 0 in RDH) and stop packets (stop bit at one in RDH) were received in consecutive occurrences and properly stored in the 'bigfifo' buffers before the timeout for reception elapsed. Then a HBACK or HBNACK is transmitted to the CTP, which in turns, according to the rules of the data-taking, updates the HBa/HBr to be transmitted to the CRU [4, 9].

5. Conclusion

An adaptable common firmware was developed to cover the needs of the upgraded detectors of ALICE [10]. We showed that by carefully designing the firmware many features could be adjusted to be configurable and thus allow the development and validation efforts to be shared for the firmware and for the associated readout software as well. The CRU and its common firmware is already used extensively and successfully by several detectors. The continuous readout mode was already validated by the most demanding sub-detectors.

References

[1] J.P. Cachemiche *et al*, *The PCIe-based readout system for the LHCb experiment*, 2016 JINST 11 P02013

[2] M. Barros Marin *et al*, *The GBT-FPGA core features and challenges*, 2015 JINST 10 C03021

[3] The GBT project website https://espace.cern.ch/GBT-Project/GBT-FPGA/default.aspx

[4] A. Kluge, P. Vande Vyvre, *The detector read-out in ALICE during Run 3 and 4*,
     ALICE-TECH-2016-001

[5] D.M. Kolotouros *et al*, *A TTC upgrade proposal using bidirectional 10G-PON FTTH technology*,
     2015 JINST 10 C04001

[6] E.B.S. Mendes *et al*, *The 10G TTC-PON: challenges, solutions and performance*,
     2017 JINST 12 C02041

[7] S.H.I. Barboza *et al*, *SAMPA chip: a new ASIC for the ALICE TPC and MCH upgrades*,
     2016 JINST 11 C02088

[8] J. Adolfsson *et al*, *SAMPA Chip: the New 32 Channels ASIC for the ALICE TPC and MCH Upgrades*,
     2017 JINST 12 C04088

[9] F. Costa *et al*, *The detector read-out in ALICE during Run 3 and 4*,
     2017 J. Phys.: Conf. Ser. 898 032011

[10] *Projet gitlab* https://gitlab.cern.ch/alice-cru/cru-fw