PROCEEDINGS OF SCIENCE

Common Readout System in ALICE

Mitra Jubin[*], **Khan Shuaib Ahmad**

For the ALICE Collaboration [†]
VECC, KOLKATA
E-mail: jubin.mitra@cern.ch

The ALICE experiment at the CERN Large Hadron Collider is going for a major physics upgrade in 2018. This upgrade is necessary for getting high statistics and high precision measurement for probing into rare physics channels needed to understand the dynamics of the condensed phase of QCD. The high interaction rate and the large event size in the upgraded detectors will result in an experimental data flow traffic of about 1 TB/s from the detectors to the on-line computing system. A dedicated Common Readout Unit (CRU) is proposed for data concentration, multiplexing, and trigger distribution. CRU, as common interface unit, handles timing, data and control signals between on-detector systems and online-offline computing system. An overview of the CRU architecture is presented in this manuscript.

7th International Conference on Physics and Astrophysics of Quark Gluon Plasma
1-5 February , 2015
Kolkata, India

[*] Speaker.

[†] A footnote may follow.

© Copyright owned by the author(s) under the terms of the Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

1. Introduction

The LHC (Large Hadron Collider) is the world's largest and most powerful particle collider, operational since the year 2009. It is going for its next major upgrade in 2018, enabling physicists to go beyond the Standard Model: the enigmatic Higgs boson, mysterious dark matter and the world of super-symmetry are just three of the long-awaited mysteries that the LHC is unveiling[1]. The LHC has already attained the maximum energy of 13 TeV centre-of-mass energy in 2015 for proton-proton collisions and 5.5 TeV per nucleon in the case of Pb-Pb collisions. From the year 2020 onwards, HL-LHC (High Luminosity LHC) will be operational whose main objective is to increase the luminosity of the machine by a large factor.

To fully exploit the physics potential provided by the machine, ALICE (A Large Ion Collider Experiment) has decided to go for a major upgrade before the start of the third phase of LHC running (RUN3). Motivated by it successful physics results and past operation experiences the R&D for ALICE upgrade has started. This manuscript presents how the change in physics objective has affected the data rate, that resulted in a new electronic block development called **Common Readout Unit (CRU)** to act as a nodal point for data, control, and trigger distribution. Figure 1 shows how the ALICE major upgrade timeline is aligned with LHC luminosity upgrade road-map.

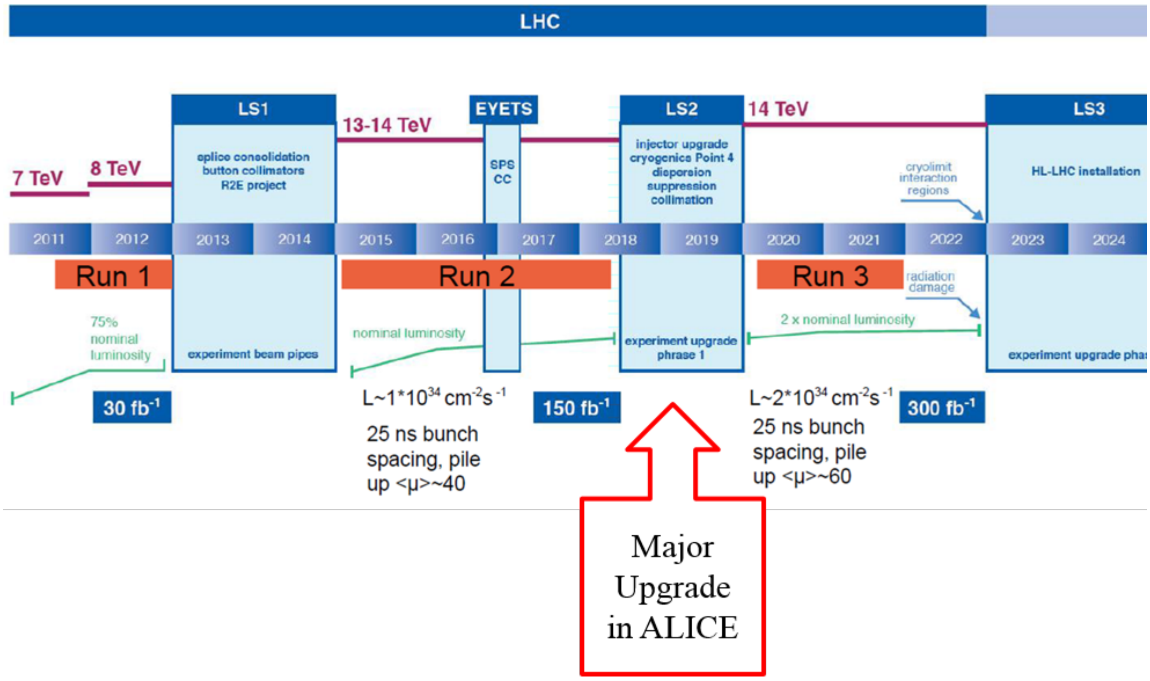


Figure 1: PHASE 1 major upgrade in ALICE to prepare for RUN3 and HL-LHC

For collider experiments, the instantaneous luminosity and integrated luminosity are important parameters to characterize its performance. As the LHC is aiming for higher luminosity, it means more number of events [2] will be generated over the experiment runtime as evident from the above expressions. Precision instrumentation of the ALICE detector is required for proper exploration of this high-intensity physics frontier. Exploration of the rare events require large event statistics. Improved vertexing and tracking with optimum detector resolution. After the planned upgrade the

PoS(ICPAQGP2015)098

readout will be capable of handling anticipated interaction rate of 50 kHz for Pb-Pb events and 200 kHz for pp and p-Pb events, resulting in a peak data flow traffic of about 1TB/s. Figure 2 shows the detectors that are going for the major upgrade as decided by ALICE collaboration.

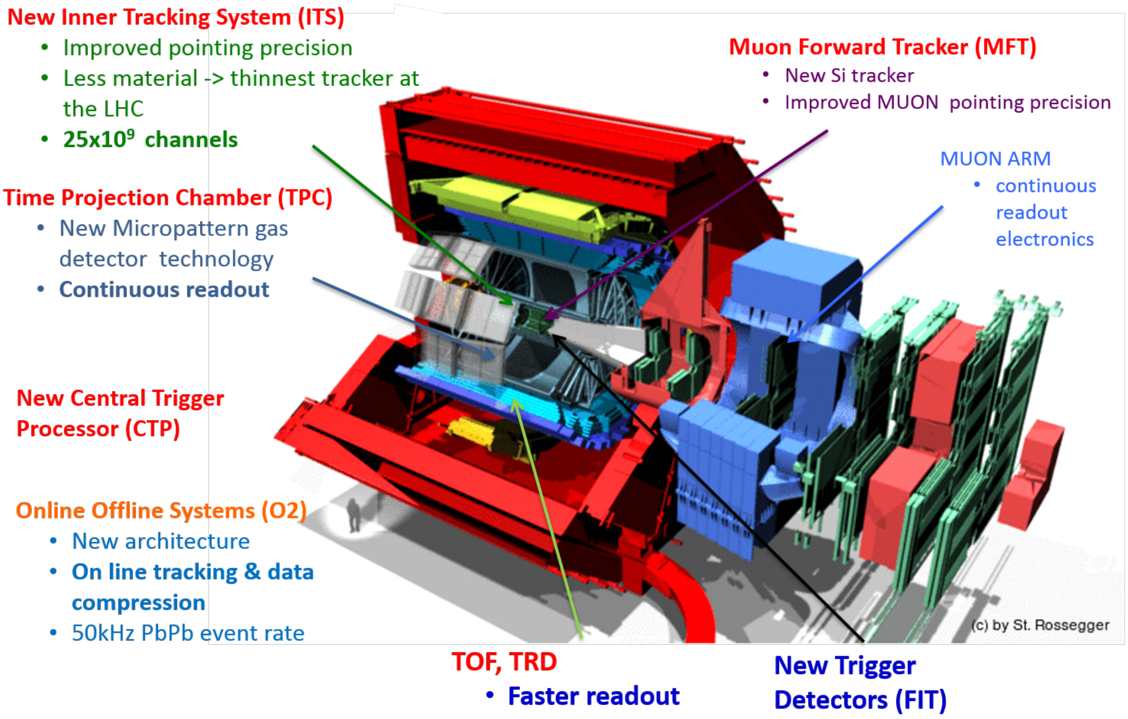

New Inner Tracking System (ITS)
• Improved pointing precision
• Less material -> thinnest tracker at the LHC
• 25x10^9 channels
Time Projection Chamber (TPC)
• New Micropattern gas detector technology
• Continuous readout
New Central Trigger Processor (CTP)
Online Offline Systems (O2)
• New architecture
• On line tracking & data compression
• 50kHz PbPb event rate
Muon Forward Tracker (MFT)
• New Si tracker
• Improved MUON pointing precision
MUON ARM
• continuous readout electronics
(c) by St. Rossegger
TOF, TRD
• Faster readout
New Trigger Detectors (FIT)

Figure 2: ALICE Upgrade from 2021

2. Technical Motivation

A critical component of electronics and computing farm in High Energy experiments is to decide on which data to store and what to discard. In such experiments, the rate at which detector data is sampled is much higher than the rate of physics interactions of primary interest. Here trigger decisions play an important role in the decision on data taking. From the past run-time experience, it is found that detector dead time, busy signal and trigger taking decisions affect data taking rate. In the upgraded architecture, it is decided to acquire data with a marked time-stamp in continuous mode and dump it on computing farms for online processing, where trigger decisions are applied to proper physics event selections. In this manner, we are not losing any significant data samples. However, there are provisions kept in this new design for non-upgraded detectors to use old technical links and trigger architectures.

The paradigm shift in readout strategy calls for ALICE to develop a new design framework for more parallelism, compact layout, and balanced load distribution. This led to the proposal for the use of new data processing block, CRU, to accelerate the system data taking performance. It is dedicated to trigger distribution, data aggregation, and detector control moderation. To keep up with future needs and demands in HEP experiments, there is growing interest in the use of

PoS(ICPAQGP2015)098

reconfigurable hardware like FPGA. With reconfigurability feature we can have faster development time, no upfront non-recurring expenses (NRE) for future upgrades, more predictable project cycle and field re-programmability. This calls for the developers to search for DAQ boards that use FPGA (Field Programmable Gate Array) and also meets with CRU firmware requirement.

3. CRU Location in the ALICE experiment

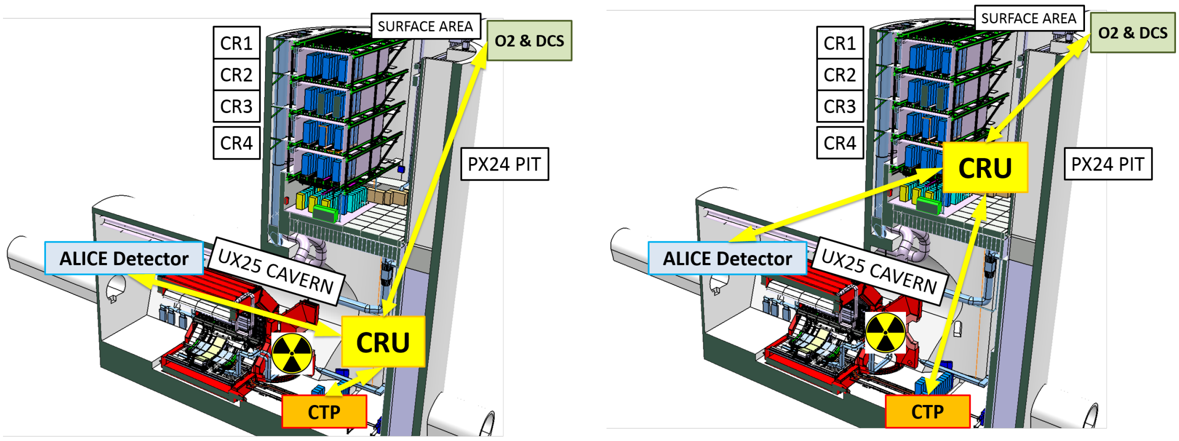
CRU acts as a common interface between ALICE on-detector electronic system, the computing system (O^2 - Online and Offline) and trigger management system (CTP - Central Trigger Processor). Being the central element, CRU has to handle three types of data traffic which include detector data, trigger and timing information and control instructions. There has been an option to keep the CRU either in the cavern or the counting room as shown in Fig. 3a and 3b. Location 1 shows CRU placed in Cavern at critical radiation zone, whereas Location 2 shows CRU placed in Counting room (CR4) at controlled radiation zone. The location choice depends on three parameters: the amount of cabling required, radiation hardness of FPGA boards needed and scope for future maintenance.

Lets us consider 1st Location site for CRU. Here, because of the proximity to radiation zone, the CRU DAQ board need to be radiation hard. It means that we also have to use radiation hard FPGAs. These radiation-hardened FPGA process technology are still many generations behind state-of-the-art commercial IC processes. For example, the rad-hard FPGAs are in the 65-nm or less-dense process nodes, whereas commercial grade FPGAs have gone down to 14-nm FINFET Technology. Now these carries a drawback, that number of logic cells available for programming are much lower than that of commercial grade FPGAs. Besides popularly used digital Single Event Upset (SEU) mitigation technique is Triple Modular Redundancy (TMR) circuits or voting logic, which further lowers the available logic resources. For these reasons the total resource available for user logic development is lower than that of the commercial grade FPGAs. The location-1, however, got some advantages over location-2, like minimum cable length required between Detector - CRU and CTP - CRU.

Now consider the 2nd Location site for CRU. Here in controlled radiation zone, we are free to choose the latest and most advanced FPGA chip available in the market and play with it. It also provides easy hardware access to design engineers even during experiment run. However, this site also has a drawback. The length of cabling required from cavern to the counting room is roughly 75 m. This involves cabling of 8344 links from sub-detectors digitized readout channels. For each optical fibre cable there involves a transmission latency of $\sim$367 ns or 15 (= 367 / 25) LHC clock cycles. So, it clearly means the trigger information pathway between CTP-CRU-Detector are suitable for triggers whose allowed $latency > 2 \times (367\ ns + Asynchronous\ Serial\ Protocol\ Serialization/De-serialization\ latency)$. It is multiplied by factor 2 to account for traversal time of the signal from CTP-CRU and back to CRU-Detector. Hence, to communicate those fast critical triggers they need to be connected directly from CTP to the sub-detectors. Altogether, cable needed is much more than location 1.

From Run3 ALICE experiment will be moving towards continuous readout architecture. In that case trigger and timing information will not be latency critical, and long asynchronous links (like GBT [3], PON [4]) can be used for trigger transmission. However, it would remain critical for

PoS(ICPAQGP2015)098



(a) Location 1

(b) Location2

Figure 3: CRU location in the ALICE experiment

sub-detectors that still depend on trigger based architecture like legacy sub-detectors or upgraded detectors operating in triggered mode for commissioning. The majority of the detector decided to operate in trigger-less architecture, based on the heartbeat software trigger that is used to designate the time frame boundaries for event building at the online computing system. For easy maintenance, the future firmware upgrades and debugging, easy accessibility is required, sometimes even in between experiment run-time data taking. Weighing all the pros and cons for both the location sites, the ALICE collaboration has voted for location 2 as the suitable position for the CRU.

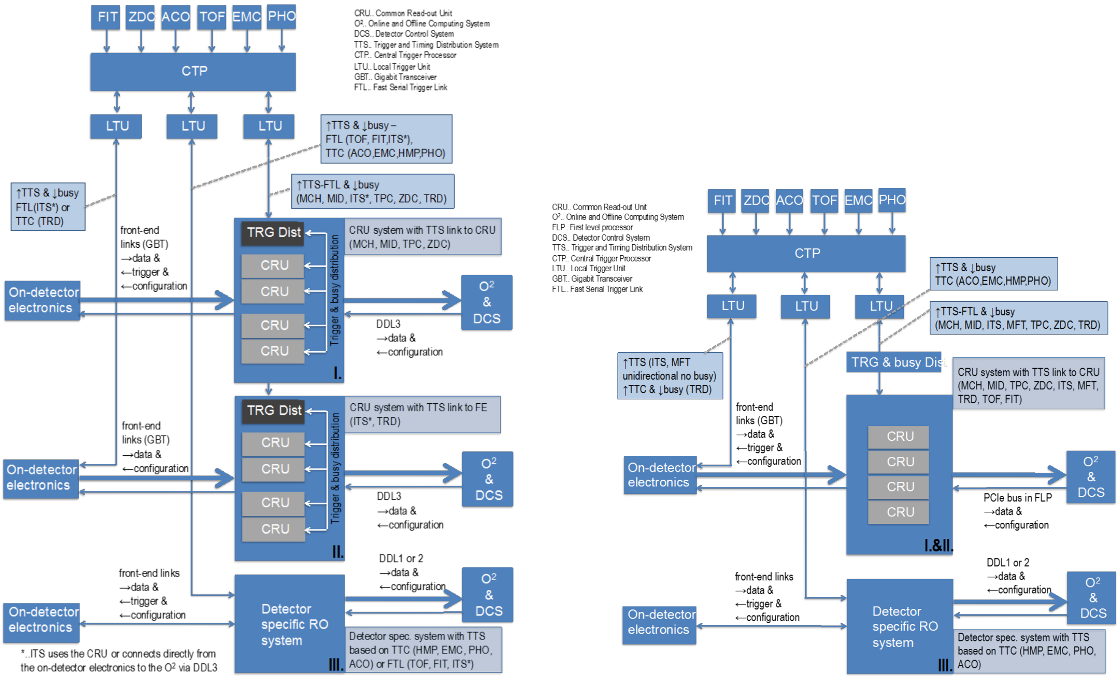
4. CRU Readout configuration

The major task of CRU functionality is to aggregate sub-detector readout channel incoming data over GBT interface links [5], [6] to be aggregated over a limited number of Detector Data Link (DDL) compatible to computing group requirement. This led to a survey of FPGA-based DAQ boards that have a maximum number of incoming optical channels and high bandwidth output channels for pushing the data to the computing systems. We have found two candidate boards suitable to match our CRU system requirement, namely PCIe40 and AMC40. PCIe40 is based on latest 20 nm Altera Arria10 FPGA, having provision for 48 bidirectional GBT links and 16 lanes PCIe channel lanes. AMC40 is based on 28 nm Altera Stratix V FPGA, having provision for 24 bidirectional GBT links and 12 bidirectional 10 Gbps links. As can be seen from table 1, total $\sim$1.1 TB/s of incoming data need to be pushed to the online system. The ALICE collaboration has decided to use separate CRU's for each sub-detectors, and also for proper load distribution again each sub-detector will not use complete CRU hardware resources at its full occupancy. Load distribution among CRU boards is critical, as it controls heat dissipation, system failure due to overload and efficient aggregation of events at the event builder of the online computing system. Therefore, an average CRU will not need more than 24 GBT links per board.

Now both the boards are our suitable candidates. The choice now depends on whether to go for ATCA (Advanced Telecommunications Computing Architecture) or PCIe based architecture. ATCA based architecture provides modularity for design framework and high-speed backplane for

PoS(ICPAQGP2015)098

trigger and control information distribution among CRU boards. While PCIe form factor needs no DDL link as it directly connects to the PCIe bus of the CPU system. However, this creates a risk, as PCs got very fast up-gradation cycle and whether presently selected PCIe Gen 3 slots would be supported in future is unclear. It means new CRU boards need to be designed. However, assurance has been given by PCI-SIG community that PCIe Gen 3 provides legacy support for upcoming next 2 generations of PCIe. So, based on two form-factor of CRUs, there can be two types of readout configuration as shown in figure 4. For details refer to ALICE Electronics Technical Design Report [7].


FIT
ZDC
ACO
TOF
EMC
PHO
CTP
LTU
CRU.. Common Read-out Unit
O².. Online and Offline Computing System
FLP.. First level processor
DCS.. Detector Control System
TTS.. Trigger and Timing Distribution System
CTP.. Central Trigger Processor
LTU.. Local Trigger Unit
GBT.. Gigabit Transceiver
FTL.. Fast Serial Trigger Link
↑TTS & ↓busy – FTL (TOF, FIT,ITS*), TTC (ACO,EMC,HMP,PHO)
↑TTS-FTL & ↓busy (MCH, MID, ITS*, TPC, ZDC, TRD)
↑TTS & ↓busy FTL(ITS*) or TTC (TRD)
TRG Dist
CRU
Trigger & busy distribution
CRU system with TTS link to CRU (MCH, MID, TPC, ZDC)
CRU system with TTS link to FE (ITS*, TRD)
On-detector electronics
front-end links (GBT) →data & ←trigger & ←configuration
front-end links (GBT) →data & ←configuration
front-end links →data & ←trigger & ←configuration
O² & DCS
DDL3 →data & ←configuration
DDL1 or 2 →data & ←configuration
Detector specific RO system
Detector spec. system with TTS based on TTC (HMP, EMC, PHO, ACO) or FTL (TOF, FIT, ITS*)
I.
II.
III.
*..ITS uses the CRU or connects directly from the on-detector electronics to the O² via DDL3
↑TTS & ↓busy TTC (ACO,EMC,HMP,PHO)
↑TTS-FTL & ↓busy (MCH, MID, ITS, MFT, TPC, ZDC, TRD)
↑TTS (ITS, MFT unidirectional no busy) ↑TTC & ↓busy (TRD)
TRG & busy Dist
CRU system with TTS link to CRU (MCH, MID, TPC, ZDC, ITS, MFT, TRD, TOF, FIT)
PCIe bus in FLP →data & ←configuration
I.&II.
Detector spec. system with TTS based on TTC (HMP, EMC, PHO, ACO)

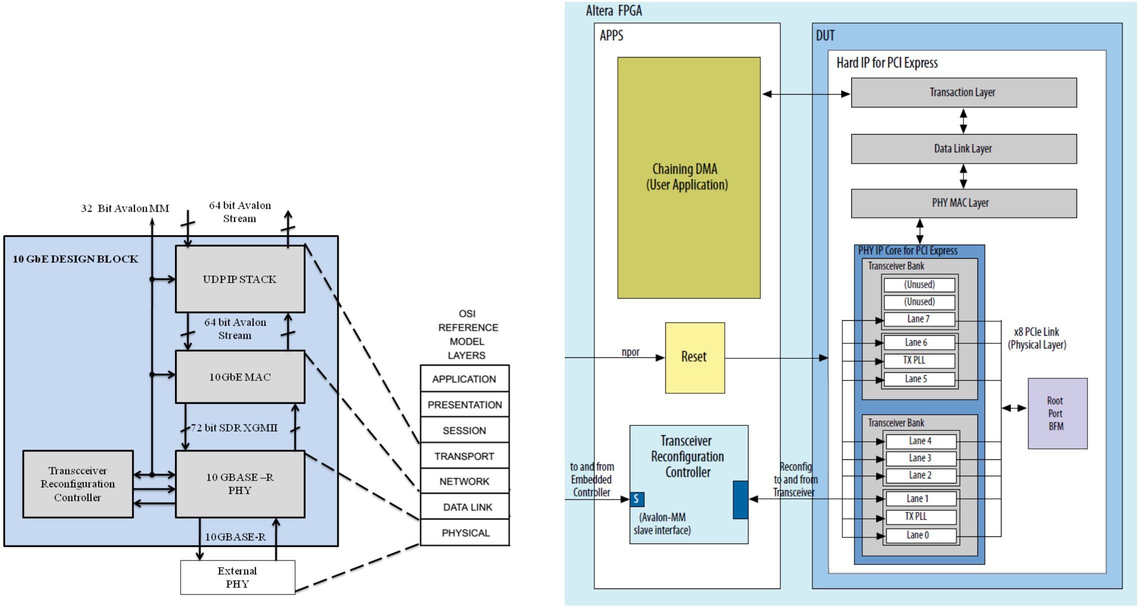
(a) Configuration 1

(b) Configuration 2

Figure 4: CRU Readout Configurations

The major decision parameter was to select FPGA board that has sufficient logic resources for detector data sorting, clustering and compressing. For Arria10 FPGA (in PCIe40) the number of logic resources is roughly double that of Stratix V FPGA (in AMC40). It means now we have to check after implementing the periphery logic, which board is left over with more resources for detector core logic development as shown in figure 5. Since Arria10 has PCIe hard IP whereas in Stratix V there is no hard IP for 10 Gigabit Ethernet IP, more logic building blocks are utilized in case of Stratix V. Clearly Arria10 is the winner and hence ALICE collaboration has opted for PCIe40 in a joint venture with LHCb Experiment group. Altera also provides vertical migration from Arria10, which means when more advanced Stratix 10 FPGA will be available on the market same firmware and hardware board can be used over again, without any recurring development cost.

PoS(ICPAQGP2015)098



(a) UDP IP stack over Stratix V (b) PCIe Interface protocol stack over Arria 10

Figure 5: Showing the implementation of two protocol stack and its interface with user application layer

5. CRU Usage

Detectors that use CRU are listed in table 1. Other detectors are not listed here. The table summarises the link usage for each detector along with the number of CRU boards needed. Moreover, the link count includes CRU-FE links that carry hit data from the on-detector electronics to the CRU and TTS-FE links that carry trigger data from the CRU to the on-detector electronics.

Table 1: Detector Specific CRU usage [8]

User Groups	**FEE /**	**No. of**	**Maximum Readout**	**Data Rate for**	**Readout Mode**	**Link Type**	**No. of Links**		**No. of**
	Readout Boards	**Channels**	**Rate (kHz)**	**Pb-Pb (GB/s)**			Bidir	Unidir	**CRU boards**
CTP (Central Trigger Processor)	FPGA (Kintex 7)	–	200	0.02	Triggered / Continuous	GBT & 10G PON	14 + 1	0	1
FIT (Fast Interaction Trigger)	FPGA (Virtex 6)				Triggered	GBT	22	0	1
ITS (Inner Tracking System)	FPGA (Kintex 7)	25×10^9	100	40	Triggered/ Continuous	GBT	192	384	24
MCH (Muon Chamber)	ASIC (SAMPA)	10^6	100	2.2	Triggered / Continuous	GBT	550	0	25
MFT (Muon Forward Tracker)	FPGA (Kintex 7)	500×10^6	100	10	Triggered/ Continuous	GBT	80	80	10
MID (Muon Identifier)	FPGA (8x Max10, 2x Cyclone V)	21×10^3	100	0.3	Continuous	GBT	32	0	2
TOF (Time Of Flight)	FPGA (IGLOO2)	1.6×10^5	100	2.5	Triggered/ Continuous	GBT	72	0	3
TPC (Time Projection Chamber)	ASIC (SAMPA)	5×10^5	50	1012	Triggered / Continuous	GBT	7200	7200	360
TRD (Transition Radiation Detector)	FPGA	1.2×10^6	200	20	Triggered (8b/10b)	Custom	0	1044	54
ZDC (Zero Degree Calorimeter)	FPGA (Vertex 5,6)	22	100	0.06	Triggered	GBT	1	1	
Total				1087.08			8164	8344	480

PoS(ICPAQGP2015)098

6. Summary

In this paper, we have introduced the reader the motivation for CRU design and also the challenges faced for CRU hardware location, configuration, and board selections. More details can be found in [9].

References

[1] L. Rossi, O Brüning, *et al.*, “High luminosity large hadron collider,” in *European Strategy Preparatory Group-Open Symposium, Krakow*, 2012.

[2] G. L. Kane and A. Pierce, *Perspectives on LHC physics*. World Scientific, 2008.

[3] J. Mitra, S. A. Khan, M. B. Marin, J.-P. Cachemiche, E. David, F. Hachon, F. Rethore, T. Kiss, S. Baron, A. Kluge, *et al.*, “GBT link testing and performance measurement on PCIe40 and AMC40 custom design FPGA boards,” *Journal of Instrumentation*, vol. 11, no. 03, p. C03039, 2016.

[4] D. M. Kolotouros, S Baron, C Soos, and F Vasey, “A TTC upgrade proposal using bidirectional 10G-PON FTTH technology,” *Journal of Instrumentation*, vol. 10, no. 04, p. C04001, 2015. [Online]. Available: `http://iopscience.iop.org/article/10.1088/1748-0221/10/04/C04001/pdf`.

[5] S Baron, J. Cachemiche, F Marin, P Moreira, and C Soos, “Implementing the GBT data transmission protocol in FPGAs,” in *TWEPP-09 Topical Workshop on Electronics for Particle Physics*, 2009, pp. 631–635.

[6] P. Moreira, R Ballabriga, S Baron, *et al.*, “The GBT project,” in *Proceedings of the Topical Workshop on Electronics for Particle Physics*, 2009, pp. 342–346.

[7] ALICE Collaboration, “Technical Design Report for the Upgrade of the ALICE Read-out & Trigger System,” *CERN-LHCC-2013-019 / LHCC-TDR-015*, 2014.

[8] Wigner R.C.P. for ALICE Collaboration, “CRU User Requirements,” *ALICE Internal Document*, no. v0.6 (Draft), 2016.

[9] J. Mitra *et. al.* for ALICE Collaboration, “Common Readout Unit (CRU) - A new readout architecture for the ALICE experiment,” *Journal of Instrumentation*, vol. 11, no. 03, p. C03021, 2016.

PoS(ICPAQGP2015)098