A 40 MHz Trigger-free Readout Architecture for the LHCb Experiment

F. Alessio [a], R. Jacobsson [a], Z. Guzik [b]

[a] CERN, 1211 Geneva 23, Switzerland

[b] IPJ, 05-400 Swierk/Otwock, Poland

Abstract

The LHCb experiment is considering an upgrade towards a trigger-free 40 MHz complete event readout in which the event selection will only be performed on a processing farm by a high-level software trigger with access to all detector information. This would allow operating LHCb at ten times the current design luminosity and improving the trigger efficiencies in order to collect more than ten times the statistics foreseen in the first phase.

In this paper we present the new architecture in consideration. In particular, we investigate new technologies and protocols for the distribution of timing and synchronous control commands, and rate control. This so called Timing and Fast Control (TFC) system will also perform a central destination control for the events and manage the load balancing of the readout network and the event filter farm. The TFC system will be centred on a single FPGA-based multi-master allowing concurrent stand-alone operation of any subset of sub-detectors. The TFC distribution network under investigation will consist of a bidirectional optical network based on the high-speed transceivers embedded in the latest generation of FPGAs with special measures to have full control of the phase and latency of the transmitted clock and information. Since data zero-suppression will be performed at the detector front-ends, the readout is effectively asynchronous and will require that the synchronous control information carry event identifiers to allow realignment and synchronization checks.

I. Introduction

The LHCb experiment at the Large Hadron Collider (LHC) at CERN has submitted an Expression of Interest for an LHCb Upgrade [1] which would allow operating LHCb at ten times the current design luminosity and allow improving the trigger efficiencies in order to collect more than ten times the statistics foreseen in the first phase. Improving the trigger efficiencies requires in practice reading out the full detector ultimately at the LHC crossing rate of 40MHz with the consequence that practically all readout electronics have to be replaced.

Fig. 1 shows the upgraded LHCb readout architecture in consideration. The Front-End Electronics will record and transmit data continuously at 40 MHz. The expected non-zero suppressed event size would result in a very large number of links between the Front-End and the new Readout Boards. It has been already shown that almost a factor of ten could be gained by sending zero-suppressed data. The zero-suppression would thus have to be performed in radiation-hard Front-End chips. The consequence is that the data will be transmitted asynchronously to the Readout Boards. Therefore, the data frames must include an event identifier in order to realign the event fragments in the Readout Boards. Fig. 2 shows a logical scheme for the Front-End Electronics which we are investigating together with the new readout control.

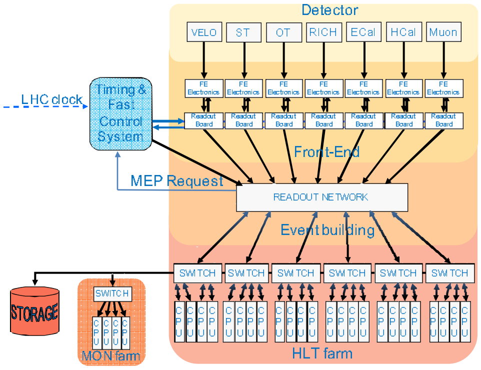

Figure 1: The upgraded LHCb readout architecture

Optical links based on the CERN GigaBit Transceiver (GBT) are being considered for the readout between the Front-End Electronics and a set of about 400 Readout Boards. The Readout Boards will act as interfaces to the event-building 16 Terabit/s network based on IP-Over-InfiniBand. We advocate here that the Readout Boards also act as the FE interface for timing and synchronous control, as well as the bridge for configuration and monitoring. The event filter farm is to be based on COTS multi-cores.

The only exception in the replacement is the current first-level trigger electronics (Level-0 trigger) which already operates at 40 MHz and which may be used to either maintain the readout rate at the current maximum of 1.1 MHz during the time the new readout electronics is being installed or at a rate between 1.1 MHz and 40MHz if the installation of the Data Acquisition (DAQ) network and Event Filter Farm is staged. The use of the current Level-0 trigger system implies that the new Timing and Fast Control (TFC) system will have to support the current distribution system based on the RD12 Timing, Trigger and Control (TTC) development [2].

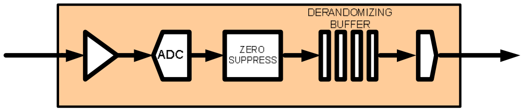

Figure 2: Proposed Front-End architecture

The rate control may also be achieved by implementing local trigger logic in the new Readout Boards (often referred to as "TELL40" as a follower of the current TELL1 [3]) and use the local decisions or rather "recommendations" centrally in the new TFC system in an intelligent trigger throttle mechanism. This type of rate control may also be used to protect the output bandwidth of the new Readout Boards if data truncation is not desired.

The experience with the current Timing and Fast Control system [4] allows a critical examination and inheriting features which are viable in the LHCb upgrade and which have evolved and matured over already eight years. In this paper we propose a new architecture based on entirely new technologies for LHCb together with an outline of the major functions of the system and their implementations. Since the schedule and logistics will probably not allow installing and commissioning the new readout electronics everywhere during only one shutdown, we aim at maintaining support for the old electronics in the new TFC system. This obviously has to be taken into consideration in the DAQ network as well.

II. System and Functional Requirements

Similar to the current system, the new Timing and Fast Control system should control all stages of the data readout between the Front-End Electronics and the online Event Filter Farm by distributing the LHC beam-synchronous clock, synchronous reset and fast control commands, and at least in the intermediate phase a trigger. Below is a list of the global functions which the new TFC system must support. Since the system must be ready before the readout electronics in order to be used in the development of the sub-detector electronics and detector test beams, the ultimate requirements are obviously flexibility and versatility.

A. Bidirectional communication network

The TFC network must allow distributing synchronous information to all parts of the readout electronics and allow collecting buffer status and, at least initially, trigger information to be used for rate control.

B. Clock phase and latency control

The synchronous distribution system must allow transmitting a clock to the readout electronics with a known and stable phase at the level of ~50ps and very low jitter (<10ps). It must also allow controlling fully and maintaining stable the latency of the distributed information. Alignment of the individual TFC links and synchronous reset commands together with event number checks will be required to assure synchronicity of the experiment.

C. Partitioning

The architecture must allow partitioning, that is the possibility of running autonomously one or any ensemble of sub-detectors in a special running mode independently of all the others. In practice this means that the new TFC system should contain a set of independent TFC Masters, each of which may be invoked for local sub-detector activities or used to run the whole of LHCb in a global data taking, and a configurable switch fabric in the TFC communication network.

D. LHC accelerator interface

The system must be able to receive and operate directly with the LHC clock and revolution frequency, and allow full control of the exact phase of the received clock.

E. Rate control

The new system should allow controlling the rate, either relying on a "blind" throttle mechanism based on the buffer occupancies in the Readout Boards or on an "intelligent" throttle mechanism based on local trigger decisions computed in the Readout Boards. The local trigger decisions may then be used as "recommendations" for the TFC system to maintain the rate at a specified level.

At the simplest level, the rate control should be based on the actual LHC filling scheme. The TFC system should therefore have means to predict the bunch structure; possibly even receive information about the bunch intensities as measured with beam pickups.

F. L0 Decision Unit input

As the initial rate control might be based on the old L0 Decision Unit [5], there should be means to interface it with the new TFC system.

G. Support for old TTC-based distribution

In order to replace the current readout electronics and commission the new electronics in steps, and make use of the L0 trigger system which is already operating at 40MHz, the new TFC system must support the old TTC system, at least for a period of time during the upgrade phase.

H. Destination control for the event packets

The system should provide means to synchronously distribute the farm destination to the Readout Boards for each event. This function should also include a request mechanism by which the farm nodes declare themselves as ready to receive the next events for processing. The event transfer from the Readout Boards is thus a push scheme with a passive pull mechanism. The scheme avoids the risk of sending events to non-functional links or nodes, and produces a level of load balancing as well as a rate control in the intermediate upgrade phase with a staged farm. Ultimately this would rather be the only emergency control of the rate when the system has been fully upgraded to a 40 MHz readout.

I. *Sub-detector calibration triggers*

The system must allow generating sub-detector calibration triggers which includes transmitting synchronous calibration commands to the FE electronics.

J. *Non-zero suppressed readout*

Since the proposed Front-End Electronics would perform zero-suppression, a scheme must be envisaged which allows occasionally a non-zero suppressed readout for special purposes. As the bandwidth does not allow this at 40 MHz but there is no requirement for high-rate, the idea is to use the TFC system to synchronize a readout mode in which the readout of a non-zero suppressed event spans over several consecutive crossings.

K. *TFC data bank*

A data bank containing the information about the identity of an event (Run Number, Orbit Number, Event Number, Universal Time) and trigger source information is currently produced by the TFC system and added to each event. A similar block should also be produced in the new TFC system.

L. *Test-bench support*

The system and its components must be built in a way that they can be used stand-alone in small test-benches and test-beams, and they have to be made available at an early stage in the development of the readout electronics.

III. OLD VS A NEW ONLINE SYSTEM ARCHITECTURE

Fig. 3 shows schematically the differences between the current LHCb Readout System architecture and the proposed architecture for the LHCb upgrade as seen from the TFC system point of view.

The current TFC system [4] has a wide timing and fast control network to the Readout Boards (ROB) and to the Front-End Electronics based on the Trigger, Timing and Control (TTC) technology developed by the CERN RD12 team [2]. It also has an independent optical throttle network based on a cheap fibre technology to communicate back-pressure to the trigger rate control logic of the TFC system. In total there are four different types of TFC custom electronics modules (TFC master, partition switch, throttle switch, and throttle fan-in) and two different types of RD12 TTC modules for the distribution backbone (Optical transmitter, optical fan-out). The TFC system receives the first-level trigger decisions from the Level-0 Decision Unit (L0DU) which processes decision data from the Pile-Up System, the Calorimeter and the Muon detectors at 40MHz and is designed to maintain the rate at a maximum of 1.1MHz.

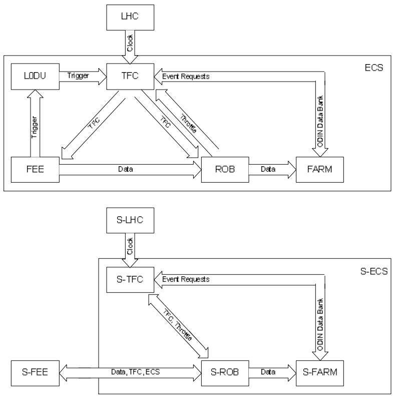

Figure 3: Old vs New Readout System architecture

In the new architecture the many TFC links to the Front-End Electronics are eliminated by profiting from the bidirectional capability of the CERN GigaBit Transceiver (GBT) development [6] and its capability to carry detector data, timing and fast control information, and Experiment Control System (ECS) information such as configuration and monitoring. In this respect the new Readout Boards become the TFC and ECS interface to the Front-End Electronics. The synchronous TFC information would thus be relayed onto a set of GBT links together with the asynchronous ECS information. The number of links from the Readout Boards to Front-End boards (TFC information and ECS configuration data) may be significantly smaller than the number of links from the Front-End boards to the Readout Boards (detector data and ECS monitoring information), possibly by a factor of ten. The TFC and ECS information would then be fanned out locally at the Front-End boards via appropriate bus types. It should be investigated if a common backplane could be envisaged to a large extent (e.g. xTCA).

The separate TFC distribution network and the throttle network between the TFC Master and the Readout Boards in the current implementation would be replaced by high-speed bidirectional optical links based on commercial technology. Unless needed during the staged upgrade to 40 MHz, the Level-0 Decision Unit would be entirely eliminated. The readout electronics would only require a rate control based on the occupancy in the output stage of the Readout Boards.

The Event Packet Request scheme mentioned in the requirements is maintained by implementing the request protocol on the new DAQ network.

IV. NEW TFC ARCHITECTURE

Fig. 4 shows the proposed new TFC architecture to fulfil the requirements of the upgraded LHCb Readout System. In the upgraded scenario, a pool of TFC Masters is instantiated in one single Super Readout Supervisor (S-TFC Master, today called ODIN) based on a single large FPGA for all TFC functions. The S-TFC Master receives the LHC clocks, as well as the LHC

Beam Synchronous Timing information, and distributes them to the instantiations.

The link to the sub-detector readout electronics on the S-TFC Master consists of a set of high-speed transceivers. In order to operate the sub-detectors stand-alone in tests or calibrations, the instantiations are independent from one another, each of which contains the logic described in the requirements. The large FPGA incorporates the configurable switch fabric which allows associating any sets of sub-detectors to the different optional TFC Master Instantiations.

The use of bidirectional links implies a point-to-point connection to each Readout Board. In order to have a manageable set of transceivers on the S-TFC Master, each Readout Board crate has to contain a fan-out/fan-in module. Thus, physically, each S-TFC Master transceiver is connected via a bidirectional optical link to an S-TFC Interface board to the Readout Boards. Hence there are as many S-TFC Interfaces as there are Readout Board crates[1], and consequently as many optical bidirectional TFC links and S-TFC Master transceivers. With 24 TFC links, the system would support up to 480 Readout Boards. If more are required, the S-TFC Interface boards could be cascaded.

The physical connection between the S-TFC Interfaces and the Readout Boards is achieved by high-speed bidirectional copper links of maximum a meter in length. Should it be decided that the Readout Boards require backplane communication, for instance implemented in one of the light-weight xTCA technologies, the TFC communication would be implemented on the backplane. The baseline solution is otherwise using hi-cat copper cables.

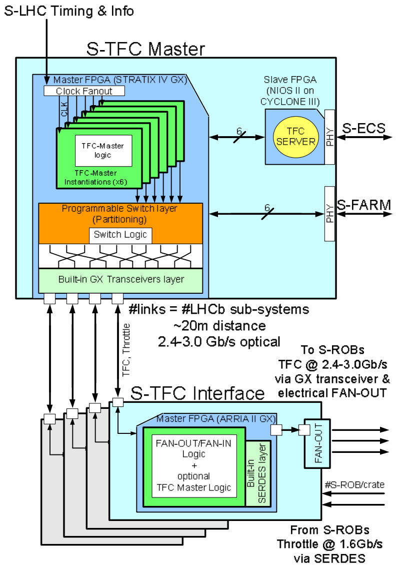

Figure 4: The New TFC architecture

A TFC transceiver block in the Readout Boards performs the clock recovery and decodes the TFC information. It also relays a subset of the information onto the GBT links which goes from the Readout Boards to the Front-End electronics and which is shared with the ECS configuration data.

Therefore the TFC transceiver block should preferably be located in the FPGA with the GBT transceiver block in Readout Boards. The TFC transceiver block also transmits the trigger/throttle information over the TFC link to the S-TFC Interface.

V. R&D Studies and Results from System Simulation

In addition to simulations, the new TFC architecture and the choices of technologies outlined in this paper contain several points requiring feasibility studies on hardware. Below is a summary of issues which need to be addressed:

- Phase and latency control and reproducibility upon power-up with the Altera GX transceivers
- Clock recovery and jitter across the GX transceivers
- Synchronous control command fan-out on the S-TFC Interface and transmission over copper between the S-TFC Interface and the Readout Boards, and effect on jitter
- Clock and synchronous control commands fan-out at the Front-End electronics
- TFC link reset sequence to establish word alignment, and phase and latency calibration across the entire TFC links, including the e-links of the GBTs
- Compounding of the TFC synchronous control information together with the asynchronous ECS information for the GBT links to Front-End electronics
- Implementation to support the old LHCb readout electronics
- Implementation of the control interface based on DIM/TCP/IP in Nios II
- Interface to the DAQ network for the Event Packet Requests and the TFC Data Bank
- Resource usage for S-TFC Master and S-TFC Interface

The use of the GBT-to-FPGA link for data transmission between the Front-End electronics and the Readout Boards is under investigation.

A full simulation framework of the new readout architecture as shown in Figure 1 and 2 has been developed.

It includes a detailed, fully configurable and fully synthesizable clock-level simulation of the new TFC components as described in this paper. It also includes an emulation of the surrounding components such as the GBT links [6], the Front-End electronics and the Readout Boards. The test bench has already allowed defining a preliminary protocol for the new TFC information and has allowed developing the first version of the firmware for the S-TFC Master and the S-TFC Interfaces in their proper environment, estimating the resource usage, studying the latencies of the system, and defining the link reset sequence and timing alignment procedure.

[1] In the case that there are several crates filled with few Readout Boards, the S-TFC Interface would span over more than one crate to keep the number of TFC links low.

Moreover, the development of a common simulation framework allows studying and validating different sub-detector implementations of the Front-End electronics and allows identifying common solutions for the Front-End electronics and Readout Boards, as well as functional inconsistencies.

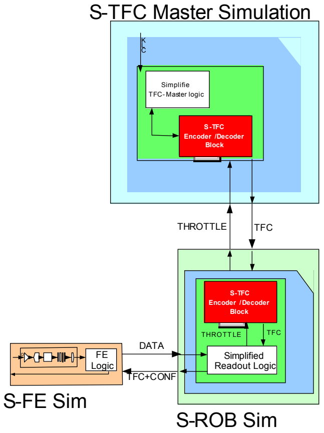

Figure 5: Schematic drawing of the system included in simulation defined as a single slice of the new Readout System.

Here first results from the simulation of a single Readout slice of the proposed architecture are presented. Figure 5 shows the system included in simulation.

A single Readout slice comprises the new Readout Supervisor (S-TFC Master), a Readout Board and one Front-End board, outputting currently one GBT link. The starting point of the S-TFC Master logic is the TFC Readout Supervisor used in the current LHCb experiment, with modifications in the protocol, in the reset sequence and in the links configuration. The implementation of the Readout Board logic concentrates on the relay of the TFC commands onto the GBT link, via a S-TFC Decoder/Encoder block, and emulation of data congestion in the Readout System in order to produce a trigger throttle signal. The Front-End block consists essentially of two parts. A Data Generator emulates the detector response, ADC and zero-suppression by producing data on a set of channels according to a Poisson PDF with a mean occupancy specific to the detector, and the LHC filling scheme. The second part implements the derandomization of the data, the packing of the data onto the GBT link, truncation handling, and emulation of the GBT link.

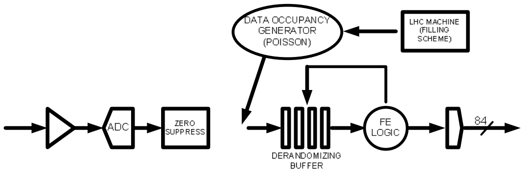

Figure 6: Schematic drawing of a single Front-End channel as implemented in simulation. A VHDL Poisson PDF generator generates ZS data. Data is buffered for processing and then packed onto the GBT link. The nominal LHC machine filling scheme is used in order to exploit the capability of the system during abort gaps and consecutive bunches.

The second part also contains the decoding of the new TFC commands, and applies them to the processing of the events. Figure 6 shows a logical scheme of the Front-End channel.

The system can be customized by changing four main parameters:

- Detector mean occupancy for the data generation
- Channel size in bits
- Number of channels associated to a single FE board, i.e. one GBT link
- Derandomizing buffer depth

The simulation is also prepared in a way that the first part performing the data emulation may be replaced with a different data emulation and data compression to study the requirements of different sub-detectors.

In order to demonstrate the simulation Figure 7 shows the distribution of number of channel with ZS data generated from the Poisson PDF generator for a detector mean occupancy of 30% and 21 channels of 12-bits associated to a single GBT link. The bin of zero occupancy originates from gaps in the LHC filling scheme. Data is buffered in the 15-word deep Derandomizing buffer before being packed and sent over the link. Figure 7 also shows the distribution of the Derandomizing buffer occupancy over almost 3 LHC turns. This particular configuration leads to a peak occupancy of 14 events implying that the truncation mechanism will strongly affect the performance of the system. The simulation shows that in this configuration, 10.5% of incoming events are truncated because of buffer overflow. The simulation also allows demonstrates that the implementation does not lead to any event size bias in the truncation.

With a word size of 80 bits. 80.4 % of the bandwidth of the GBT link is exploited.

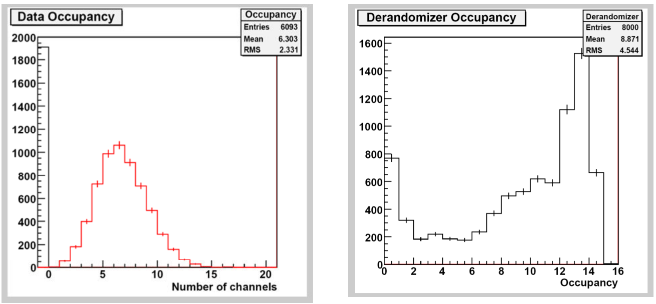

Figure 7: On the left, distribution of channels filled with ZS data in agreement with a Poisson PDF. On the right, distribution of the derandomizing buffer occupancy

The link usage of the GBT link can be improved by optimizing the front-end parameters. In fact, configuring the Derandomizing buffer as 24 words-deep, simulation shows that the system decreases the event loss by a factor 2, resulting in 5.4% of truncated events and a GBT link usage of 83.2%.

Figure 8 shows the trend of the percentage of truncated event as a function of the Derandomizing buffer depth.

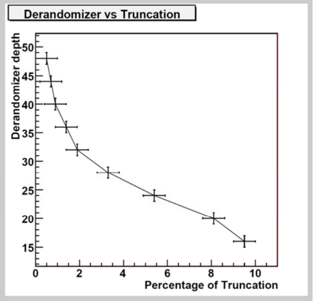

Figure 8: Percentage of events truncated as a function of the Derandomizing buffer depth.

VI. PROTOTYPING PLANS

In order to match the schedule for the Upgrade expressed in the EOI [1] and to have a system ready and robust by the time in which each sub-system will start to test their new readout electronics and validate the conformity with the common specifications, the development of the TFC system must take a lead as was done for the current TFC system. This emphasizes the importance that the system is designed with maximum flexibility and versatility in order to adapt and add functionality as the requirements of the readout system emerge.

A first prototype board is being specified. It is aimed at carrying out the feasibility studies described in Section V. It will be a hybrid S-TFC Master/Interface board with a small set of all the functionalities and I/Os of the two boards, including loopback for all links in order to perform link tests, and latency and jitter studies.

VII. CONCLUSION

In this paper we have outlined a ‘top-down approach’ to the design of a new Timing and Fast Control system for the LHCb upgrade. The new architecture relies heavily on new FPGA and link technologies which allow reducing the number of optical links and boards to provide timing and synchronous control to the entire readout chain of LHCb while adding flexibility and robustness.

A full simulation framework for the TFC components including a readout slice of Front-End electronics and Readout Boards has been implemented. It allows developing the TFC functionality and protocols, and testing the readout control in the proper environment at clock level. It also allows studying and validating different Front-End models, and optimizing latency and buffer requirements.

The choices call for several feasibility studies which will be done based on a first hybrid prototype. The R&D plan and the architecture takes into account the fact that the developments of the new readout electronics will need the new TFC system and that stand-alone operation in test-benches outside the pit must be possible.

REFERENCES

[1] LHCb Collaboration, “Expression of Interest for an LHCb Upgrade”, CERN/LHCC/2008-007, April 22, 2008

[2] S. Baron et al., TTC website: http://ttc.web.cern.ch/TTC/

[3] G. Haefeli et al., “TELL1 - Specification for a common read out board for LHCb”, LHCb 2003-007, October 10, 2003

[4] Z. Guzik, R. Jacobsson, B. Jost, “Driving the LHCb Front-End Readout”, IEEE Trans. Nuclear Science, vol. 51, pp 508-512, 2004

[5] R. Cornat, J. Lecoq, P. Perret, “Level-0 decision unit for LHCb”, LHCb 2003-065, August 22, 2003

[6] P. Moreira, A. Marchioro, K. Kloukinas, “The GBT : A proposed architecture for multi-Gb/s data transmission in high energy physics”, Topical Workshop on Electronics for Particle Physics, Prague, Czech Republic, 03 - 07 Sep 2007, pp 332-336