Clock and timing distribution in the LHCb upgraded detector and readout system

F. Alessio, a,1 S. Baron, a M. Barros Marin, a J.P. Cachemiche, b F. Hachon, b R. Jacobsson, a and K. Wyllie a

aCERN, CH-1211, Switzerland
bCPPM, Marseille, France
E-mail: Federico.Alessio@cern.ch

ABSTRACT: The LHCb experiment is upgrading part of its detector and the entire readout system towards a full 40 MHz readout system in order to run between five and ten times its initial design luminosity and increase its trigger efficiency. In this paper, the new timing, trigger and control distribution system for such an upgrade is reviewed with particular attention given to the distribution of the clock and timing information across the entire readout system, up to the FE and the on-detector electronics. Current ideas are here presented in terms of reliability, jitter, complexity and implementation.

KEYWORDS: Digital signal processing (DSP); Control and monitor systems online; Modular electronics

1Corresponding author.
The upgrade of the LHCb experiment

The LHCb experiment [1] is a high-precision experiment at the LHC devoted to the search for New Physics by precisely measuring its effects in CP violation and rare decays. By applying an indirect approach, LHCb is able to probe effects which are strongly suppressed by the Standard Model, such as those mediated by loop diagrams and involving flavor changing neutral currents.

In the proton-proton collision mode, the LHC is to a large extent a heavy flavor factory producing over 100,000 bb-pairs every second at the nominal LHCb design luminosity of $2 \times 10^{32}$ cm$^{-2}$s$^{-1}$. Given that bb-pairs are predominantly produced in the forward or backward direction, the LHCb detector was designed as a forward spectrometer with the detector elements installed along the main LHC beam line, covering a pseudo-rapidity range of $2 < \eta < 5$ well complementing the other LHC detectors ranges.

LHCb proved excellent performance in terms of data taking [2] and detector performance over the period 2010–2012 accumulating $\sim 3$ fb$^{-1}$ of data and it is foreseen to accumulate other $\sim 5$ fb$^{-1}$ over the period 2015–2018. Due to the foreseen improved performance of the LHC accelerator, the prospect to augment the physics yield in the LHCb dataset seems very attractive. However, the LHCb detector is limited by design in terms of data bandwidth — 1 MHz instead of the LHC bunch crossing frequency of 40 MHz — and physics yield for hadronic channels at the hardware trigger. Therefore, a Letter Of Intent [3], a Framework TDR [4] and a Trigger and Online TDR [5] document the plans for an upgraded detector which will enable LHCb to increase its physics yield in the decays with muons by a factor of 10, the yield for hadronic channels by a factor 20 and to collect $\sim 50$ fb$^{-1}$ at a leveled constant luminosity of $1-2 \times 10^{33}$ cm$^{-2}$s$^{-1}$. This corresponds to ten times the current design luminosity and increased complexity (pileup) of a factor 5.
The upgrade of the LHCb readout architecture

In order to remove the main design limitations of the current LHCb detector as described in the previous section, the strategy for the upgrade of the LHCb experiment essentially consists of ultimately removing the first-level hardware trigger (L0 trigger) entirely, hence to run the detector fully trigger-less. By removing the L0 trigger, LHC events are recorded and transmitted from the Front-End electronics (FE) to the readout network at the full LHC bunch crossing rate of 40 MHz, resulting in a $\sim40$ Tb/s DAQ bandwidth. All events will therefore be available at the processing farm where a fully flexible software trigger will perform selection on events, with an overall output of about 20 kHz of events to disk. This will allow maximizing signal efficiencies at high event rates.

The direct consequences of this approach are that some of the LHCb sub-detectors will need to be completely redesigned to cope with an average luminosity of $2 \times 10^{33}$ cm$^{-2}$s$^{-1}$ and the whole LHCb detector will be equipped with completely new trigger-less FE electronics. In addition, the entire readout architecture must be redesigned in order to cope with the upgraded multi-Tb/s bandwidth and a full 40 MHz dataflow [6]. Figure 1 illustrates the upgraded LHCb readout architecture. It should be noted that although the final system will ultimately be fully trigger-less, a first-level trigger similar to the current L0 trigger will be calculated in software. This is commonly referred to as Software LLT and its main purpose is to allow a staged installation of the DAQ network, gradually increasing the readout rate from the current 1 MHz to the full and ultimate 40 MHz. This however will not change the rate of events recorded at the FE, which will run fully trigger-less regardless of the DAQ output rate.

In order to keep synchronicity across the readout system, to control the FE electronics and to distribute clock and synchronous information to the whole readout system, a centralized Timing and Fast Control system (TFC, highlighted in figure 1) has been envisaged, as an upgrade of the current TFC system [7]. The upgraded TFC system will then be interfaced to all elements in the readout architecture by heavily profiting from the bidirectional capability of optical links and FPGA transceivers and a high level of interconnectivity. In particular, the TFC system will utilize the capabilities of the GigaBit Transceiver chipset (GBT) [8] currently in production stage at CERN for its communication to the FE electronics. In addition, the TFC system will also be responsible to transmit slow control (ECS or Experiment Control System) information to the FE, by means of FPGA-based electronics cards interfaced to the global LHCb ECS.

The TFC timing and readout control system

Figure 2 illustrates in detail the logical architecture of the upgraded TFC system. A pool of Readout Supervisors (commonly referred to as S-ODIN) centrally manages the readout of events, by generating synchronous and asynchronous commands, by distributing the LHC clock and by managing the dispatching of events. Each S-ODIN is associated with a sub-detector partition which effectively is a cluster of Readout Boards (TELL40) and Interface Boards (SOL40). While the TELL40s are dedicated to read out fragments of events from the FE and send them to the DAQ for software processing, the SOL40 boards are dedicated to distribute fast and slow control to the FE, by relaying timing information and clock onto the optical link to the FE, and by appending ECS information onto the same data frame. Thanks to the characteristics of the GBT chipset [8], fast
commands, clock and slow control are therefore transmitted on the same bidirectional optical link. This is a major novelty with respect to the current LHCb experiment where fast control and slow control are sent over different networks.

At the FE, the synchronous fast control information are decoded and fanned out by a GBT Master per FE board, also responsible to recover and distribute the clock in a deterministic way. The slow control information is relayed to the GBT-SCA (which stands for Slow Control Adapter) chip via the GBT Master. The GBT-SCA chip is capable of efficiently distributing ECS configuration data to the FE chips by means of a complete set of buses and interfaces, in a generic way [9]. Monitoring data is sent back on the uplink of the same optical link by following the return path, from the GBT-SCA to the Master GBT to the corresponding SOL40.

The hardware backbone of the entire readout architecture is a PCIe Gen3 electronics card hosted in a commercial PC. The same hardware is used for the TELL40, the SOL40 and the S-ODIN boards, only the different firmware changes the flavor of the board. The board will be equipped with up to 48 bidirectional optical link, an Altera Arria 10 FPGA (GX 1150) and a 16x PCIe Gen3 bus interfaced to a multi-core PC. The card is being currently being developed at CPPM in Marseille.

4 Challenges of timing and clock distribution in the LHCb upgrade

Due to the centralized nature of the LHCb Readout Supervisory system (TFC), its implementation within the upgrade of the LHCb experiment poses some challenges in terms of timing distribution, clock recovery, jitter, readout synchronization and ultimately robustness/reliability and control.
The TFC system is in fact a single point of entry for synchronizing the LHCb experiment to the LHC accelerator, by being interfaced to the main LHC 40 MHz clock and its timing information. The clock must be received, recovered, cleaned and fanned out to all elements in the readout architecture, down to the very last FE chip with a deterministic phase and constant latency. In addition, it must be monitored and controlled in a reliable way: once each detector’s partition itself is locally time aligned, the timing of the experiment is adjusted globally to match the LHC beam structure and the fine phase of collisions, ideally with a final tolerance of below 100 ps. As a consequence, every clock source and clock reception device must move accordingly in a deterministic way, in order to have the full system globally aligned.

This is simplified schematically in figure 3 in the context of the LHCb upgrade. Particular care must be given to clock crossing domain paths. In the context of the TFC system, this is particularly important in the SOL40 cards where the clock recovered from the TFC optical stream must also be used to drive the FPGA transceivers which fan out the timing information towards the FE chips. The FE chips ultimately must recover the clock and use it to sample detector data and drive their internal DSPs and analog electronics. At the SOL40, only one optical link is used to connect the cards to the S-ODIN cards to receive timing information to be fanned out, while between 24 to 48 links are used to drive the timing information to the FE depending on the configuration of each partition. Each transceiver must be configured, monitored and controlled so that it maintains the clock quality transmission, its phase and its latency. It is estimated that the whole TFC system will have to drive up to \( \sim 2500 \) destinations at the FE in this way.

Achieving these goals can be very difficult, considering that the current readout is planned to be based on fast optical links and commercial FPGA transceivers, which must be configured and adapted properly to these goals. In addition, the total number of destinations to be reached by the timing information is \( \sim 500 \), including TELL40 and SOL40 cards, while still maintaining the possibility of partitioning and performing local operations.
Choice of clock and timing distribution architecture and technology

In order to fulfill the previous requirements, the architecture of the TFC system has been finalized following currently available technology and the developments from generic R&D projects in our community.

The architecture is illustrated in figure 4, where only a detector’s partition is highlighted: each partition will be controlled via a hierarchy composed of a single S-ODIN and a set of SOL40 boards, enough to cover the entire FE electronics Master GBTs. Each S-ODIN will act as a central Readout Supervisor for that partition, distributing the global clock, fast commands and reset, and other synchronous and asynchronous commands all based on configurable recipes. This will allow each partition to work and run independently from one another, while maintaining scalability.

Synchronicity across all units in the partition is ensured by the usage of Passive Optical Network technology (PON), a technology used in the Fibers-to-The-Home (FTTH) where a passive optical splitter allows to reach many destinations from a single point of start without the need to re-drive the initial stream. Since the clock, timing and readout control signals are centrally generated and they are the same for all destinations in a partition, they can be distributed passively.

A centralized effort at CERN [9] started looking into the possibility of using 10G-PON technology to replace the CERN wide Timing-Trigger and Control (TTC) legacy system at the experiments. The main advantages of having this technology in comparison to the current TTC system are bi-directionality, high number of destinations from a single point of start, high bandwidth, software partitioning and possibilities to use FPGA technology directly to recover the stream of data. This allows the possibility to tune the technology to each experiment’s needs.

Figure 3. Illustration of the clock distribution paths in the upgraded LHCb readout architecture.
Figure 4. Architecture of the TFC system for a sub-detector partition.

In practice, the partition’s timing and readout control information are generated centrally in S-ODIN. These are then synchronized with the LHC main clock which is used to drive the transmitters at the FPGA. The stream of data is split across all destinations in the partitions, where clock and data are recovered with fixed latency and deterministic phase. At the TELL40, the clock and timing information are used to decode the stream of data coming from the detector’s FE. At the SOL40, the clock and timing information — merged together with slow control information — are then relayed onto other FPGA transmitters, to finally reach the FE Master GBTs.

In addition, all SOL40 and TELL40 boards’ FPGAs are configured with a centralized GBT-FPGA core [10], whose aim is to generically be able to drive and receive data from any GBT chip, while maintaining constant phase and minimizing latency. The GBT-FPGA core includes critical features to be able to correctly drive the transmitters at the FE so that the stream of data at the GBT keeps information regarding the clock phase.

Globally, a centralized shift at the LHC clock reception location will propagate correspondingly to the entire partition. This is then followed by a resynchronization mechanism (not described here) whose aim is to get all receivers locked again.

5.1 Critical aspects in the LHCb clock and timing distribution system

The most important aspect of choosing PON technology is the possibility of reaching many destinations without the need of active fan-out and fan-in boards. This reduces drastically the complexity of the architecture as the clock needs to be recovered from an optical serial stream only once: at the SOL40 boards and at the TELL40 boards. At the FE, the GBT chipset has robust clock recovery capabilities as the chip was designed to contain PLLs and CDR blocks for this specific purpose. In the case of the need of an active fan-out/fan-in, the clock would have had to be recovered twice before even being sent out to the FE, augmenting the risk of introducing jitter and noise. On the
Figure 5. A Master S-ODIN is used to centrally manage the global running of the LHCb experiment in the upgraded scenario. The LHC clock is fed to each of the sub-detector partition’s S-ODIN in a sampling mechanism in order to allow for robustness and reliability.

other hand, ad-hoc solutions must be envisaged [9] in order to fulfill the requirements in terms of clock recovery and latency.

Moreover, the possibility of having bi-directionality available across the TTC network allows for a high level of interconnectivity. PON technology uses Time Division Multiple Access (TDMA) for the upstream signals. This implies that each sender is allocated a time slot and that the central receiver (S-ODIN in this particular context) needs to wait for each sender to be done transmitting its information. Again, an advanced study on this was done and the maximum round-trip time was limited to $\sim 5 \mu s$ for up to 128 destinations [9]. This is perfectly compatible with the LHCb upgraded readout architecture as the upstream path is only used to transmit asynchronous busy information (throttle) back to S-ODIN to centrally reject a particular event. Since data are transmitted asynchronously from the FE as well and the TELL40 boards are located in PCs equipped with various GB of memory, the interval of 128 clock cycles in the worst case scenario is highly manageable. In addition, various techniques can be adopted to reduce such delay: not all senders must send at all times, thus reducing the effective round-trip delay or throttling could be centrally extended for more consecutive clock cycles rather than for just one single clock cycle.

5.2 Ensuring synchronicity in global running

Global synchronicity during global running is ensured by a Master S-ODIN who is dedicated to centrally generate and distribute commands, reset and synchronous/asynchronous commands. The concept is illustrated in figure 5.

In this case, the fine and deterministic phase of the clock is maintained as each partition’s S-ODIN is interfaced to the LHC clock thus using it to drive its serializers towards the SOL40 and TELL40 boards. Constant latency is ensured by buffering timing and readout commands at the partition’s S-ODINs in order to compensate for the delay in adding one level in the hierarchy.

6 Hardware requirements

The architecture of the TFC system poses some strict hardware requirement on the common hardware boards which are being developed for the LHCb upgrade.
The need for clock cleaning and resynchronization drives the requirement of having an external PLL on-board. Such a PLL would take the recovered clock from the FPGA, clean it and re-drive the FPGA serializers with deterministic phase.

Another requirement is the presence of a dedicated connector to drive and receive the optical stream through the PON network. It is expected that in the near future 10G-PON compatible SFP+ transceivers will be available.

Lastly, the card needs a specific LHC interface. It is currently foreseen to implement on the card some connectors for single-ended signals to be interfaced to the current LHC-TTC system. In the future, the LHC interface might be an external custom-made module which is interfaced via an optical link to the various S-ODIN boards.

An illustration of such a board is presented in figure 6.

7 Conclusion

Within its upgrade, the LHCb experiment is finalizing the specifications of its sub-system. The timing and readout control system TFC is a crucial system in the upgrade of the LHCb detector as it is responsible to centrally manage the readout of event, the distribution of synchronous and asynchronous commands and the distribution of the global clock, received from the LHC.

In this paper, the current ideas for the architecture of the TFC system have been presented. PON technology seems to be the ideal solution for such a system, together with optical links and FPGAs and high-level of interconnectivity.

As commercial PON components will become available in the near future, an extensive testing campaign within CERN will be performed in order to validate the ideas and deploy the system in its scale.
References

[1] LHCb collaboration, The LHCb detector at the LHC, 2008 JINST 3 S08005.

[2] R. Jacobsson, Performance of the LHCb detector during the LHCb proton runs 2010–2012, IEEE Nucl. Sci. Symp. Med. Imag. Conf. (2012) 1479.

[3] LHCb collaboration, Letter of intent for the LHCb upgrade, CERN-LHCC-2011-001 (2011).

[4] LHCb collaboration, Framework TDR for the LHCb upgrade, CERN-LHCC-2012-007 (2012).

[5] LHCb collaboration, LHCb trigger and online upgrade technical design report, CERN-LHCC-2014-016 (2014).

[6] F. Alessio et al., Trigger-less readout architecture for the upgrade of the LHCb experiment at CERN, 2013 JINST 8 C12019.

[7] F. Alessio and R. Jacobsson, Timing and fast control for the upgraded readout architecture of the LHCb experiment at CERN, IEEE Trans. Nucl. Sci. 60 (2013) 3438.

[8] P. Moreira et al., The GBT Ser-Des ASIC prototype, 2010 JINST 5 C11022.

[9] D. Kolotouros et al., A TTC upgrade proposal bidirectional 10G-PON FTTH technology, in Topical Workshop on Electronics for Particle Physics 2014, 22–26 September 2014, Aix en Provence, France.

[10] M. Barros Marin et al., The GBT-FPGA core: features and challenges, in Topical Workshop on Electronics for Particle Physics 2014, 22–26 September 2014, Aix en Provence, France.