

TELL1

A common data acquisition board for LHCb

G. Haefeli, A. Bay, F. Legger, L. Locatelli

LPHE, EPFL, 1015 Lausanne, Switzerland
Guido.Haefeli@cern.ch

Abstract

An off-detector electronics acquisition readout board for LHCb called TELL1 has been developed to read out optical or analogue data from the front-end electronics. The FPGA based board is used for event synchronization, buffering during the latency of the 2nd level trigger, pre-processing including common mode correction and zero suppression. For the data acquisition the board is interfaced to standard Gigabit Ethernet network equipment providing up to four Gigabit Ethernet links. TELL1 accepts 24 optical links running at 1.6GHz and provides for the analogue option 64 10-bit ADC channels sampling at 40MHz.

I. INTRODUCTION TO THE LHCb DAQ AND TRIGGER SCHEME

LHCb is an experiment dedicated to B physics. An overview of the LHCb trigger electronics is given in Figure 1. The multilevel trigger scheme comprises a Level-0 based on high-pt with a rate of 1.11MHz. For the Level-1 trigger the tracks in the Vertex detector (VeLo) are reconstructed and a selection based on tracks with high p_t and significant impact parameter to the primary vertex is used to reduce the event rate to 40KHz. In a final selection called High Level Trigger (HLT) the complete event information is made available and event selection is done with better resolution and selection cuts dedicated to specific final states. A description of the requirements to the front-end and off-detector electronics can be found in [1] and [2].

In the technical proposal [3], the topological Level-1 trigger used only information from the VeLo. During the last year the LHCb detector has undergone a re-optimization [4] with a major impact on the Level-1 trigger system. The base line Level-1 trigger has now access to the data of the VeLo and the first two tracker stations (TT) located in front of the magnet. The trigger optimization study has shown that information of the tracker stations located after the magnet and Muon stations [5] can lead to an improved trigger performance. The necessity of having a Level-1 trigger data path for essentially all detectors in the experiment and the general requirements as buffering during Level-1 latency and the common data path to DAQ, strongly suggests the use of a unique readout board for the experiment. This allows for an upgrade of the Level-1 trigger by adding additional bandwidth

to the readout network without changing the custom electronics.

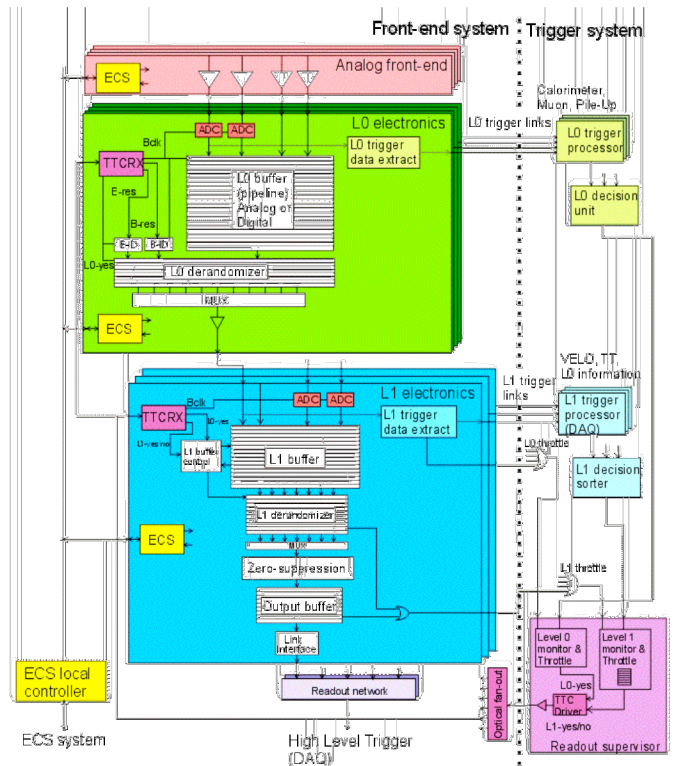


Figure 1: LHCb trigger scheme overview.

The LHCb Level-1 trigger algorithm is processed by general purpose CPUs in the combined Level-1 and HLT farm (see Figure 2). Event data is buffered inside the TELL1 board until the Level-1 decision is taken. To allow a maximum processing time in the CPUs the time to transport and build the event in the readout network has to be kept small. Moreover, flexibility requires the buffer to be made the maximum affordable size. Since the devices used in the readout network for routing the event fragments to the Sub Farm Controllers (SFCs) is standard Ethernet equipment the latency to build an event is expected to be of the order of 10ms and cannot be changed. Even if the average algorithm processing time is less than 0.5ms a small fraction of events has an expected processing time of 10-50ms. As a compromise between Level-1 buffer cost and trigger performance a buffer size of 58254 events corresponding to 52.4ms latency has been chosen for LHCb. The LHC TTC

system [6] is used to distribute the fast timing control information as Resets and Level-0 and Level-1 decisions whereas for monitoring and control a credit card sized PC module is used [7].

II. EVENT BUILDING NETWORK

With a Level-0 trigger rate of 1.11MHz the packet rate in a system where each event fragment is transmitted one by one is too high for a standard Ethernet network. To reduce the packet rate and the overhead caused by the transmission protocol a solution has been found by packing several events into a Multi Event Packet (MEP). The number of event fragments put in one MEP, the packing factor, is an adjustable parameter to be chosen in order to optimize the performance of the network. The packing factor will be set such that one Ethernet frame per MEP is created in average. The maximum value has been (generously) fixed to be 32 for the Level-1 and 16 for the HLT trigger.

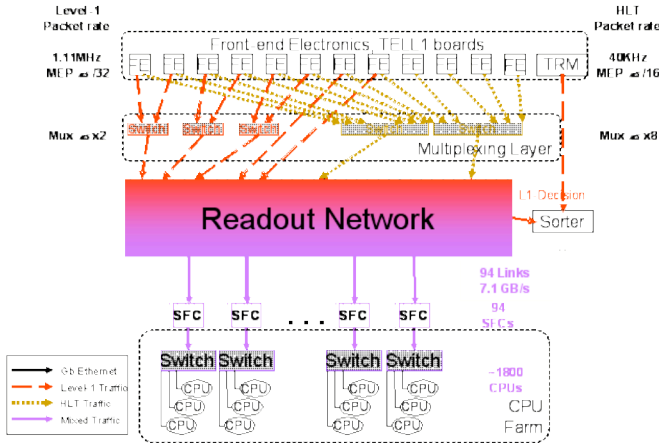


Figure 2: From the TELL1 Level-1 trigger and HLT data is sent on two different links to the data concentrator switches. The readout network routes the data to the SFCs where the MEP packets are disentangled and complete events are distributed to the CPUs.

To minimize the number of ports used on the network switch a first multiplexing stage is inserted after the TELL1 boards. For the readout network a big monolithic switch or several switches can be used. The “Event building” from single MEPs will be done by the SFC PCs from where single complete events are distributed to the sub-farm processing CPUs. Since the entire network is built from standard network equipment the packets provided by the TELL1 board need to comply with the IP protocol which contains among other information the destination in the network. The destination is assigned by the “Readout Supervisor” [8] and distributed via TTC system to the TELL1 boards for each MEP, allowing a dynamic load balancing in the farm.

III. DATAFLOW ON THE TELL1 BOARD

Figure 3 shows an overview of the dataflow on the TELL1 readout board [9]. The motherboard has 4 connectors to accept up to 4 input receiver mezzanine cards. The two types of daughter cards are called A-RxCARD for the analogue

version which performs the digitization of the input signals and O-RxCARD for the optical link receiver cards in charge of de-serializing the optical signals. The data from the receivers is transferred on a parallel bus to one of the four PP-FPGAs which are the main processing units on the board.

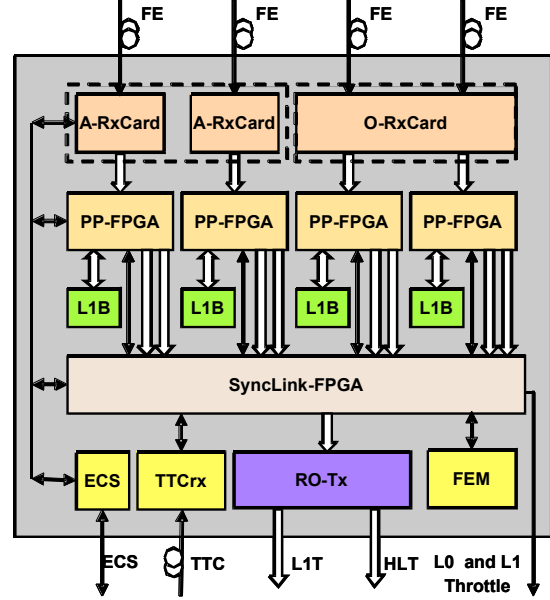


Figure 3: Dataflow overview of the TELL1 board. The blocks indicate the partitioning in different daughter cards, FPGAs and external interfaces.

Level-1 trigger data path For the Level-1 trigger data path, the raw data undergoes the pedestal subtraction, channel masking and common mode suppression before zero correction is applied. Algorithms for the so called Level-1 trigger pre-processing have been studied and implemented in hardware [10]. The data path on the PP-FPGA is shown in a simplified block diagram in Figure 4. After zero suppression the data from different channels are linked together and sent over a fast parallel point-to-point link to the SyncLink-FPGA where a final linking is performed (see Figure 5). Special care has to be taken to avoid buffer overflows at the linking stages due to the restricted bandwidth of the readout network. A large buffer of 64KByte is inserted on the output stage of the Level-1 trigger link of each PP-FPGA. This buffer is used as a de-randomizer and allows to process events with high occupancy without data loss. In case the readout de-randomizer buffer fill state exceeds a certain level, a throttle signal (Level-0 throttle) is asserted which indicates to the “Readout Supervisor” to stop accepting events at Level-0. This mechanism can only prevent from buffer overflows, with a link buffer large enough to store the remaining already accepted events which is done in the 64KByte link de-randomizer buffer.

HLT data path During the processing of the Level-1 trigger decision the data is stored in the Level-1 buffer, implemented as a DDR SDRAM bank for each PP-FPGA. After a Level-1 accept, the data is read out of the memory, linked to one event fragment on the PP-FPGA and sent over a dedicated point-to-point link to the SyncLink-FPGA. Once written into the input

FIFOs, the data is linked again, zero suppressed and encapsulated into Ethernet and IP protocol. Even the necessary bandwidth for this data path is moderate; all buffers need to be prevented from overflow which is done with the Level-1 throttle mechanism. In an identical way as for the Level-0 throttle, it is the “Readout Supervisor” which is in charge of controlling the Level-1 Accept distribution.

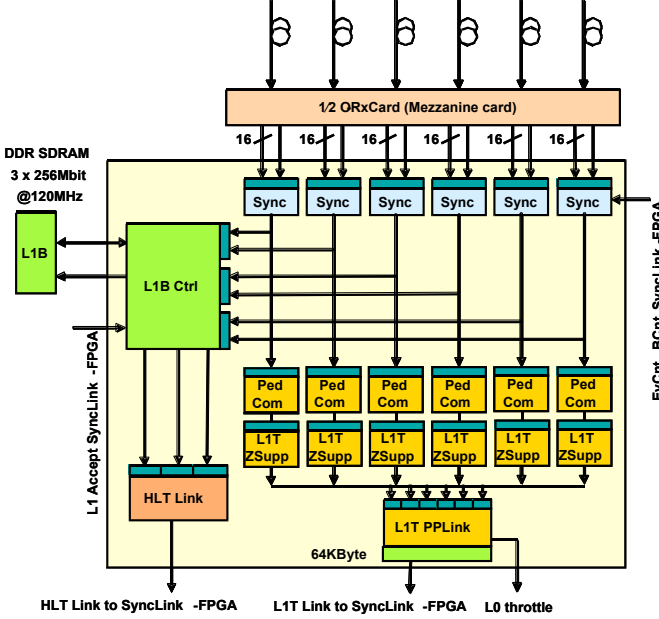


Figure 4: Data flow block diagram from the optical receiver to the PP-FPGA and access to the Level-1 buffer. The blocks marked “Ped Com” represent the pedestal and common mode noise correction.

IV. LEVEL-1 BUFFER IMPLEMENTATION

Several implementations for the Level-1 buffer (L1B) have been studied. During the early design phase the size of the buffer was low enough to allow for static memory e.g. FIFO or single port SRAM. With the latest changes in the readout and trigger system the need for more computing time has increased the Level-1 latency by a factor 64 (!) and is now 52.4ms. This made the change to dynamic RAM (SDRAM) necessary. At the present we use the biggest DDR SDRAM chip available which contains 256Mbit. For the final board we plan to use the next generation memory of 512Mbit since no hardware changes are necessary for this upgrade. The cost for the 256Mbit chip is as low as 10 CHF and the cost for the 512Mbit in one or two years is expected to be comparable. In order to obtain the desired bandwidth for reading, writing and refreshing the memory, the chosen operation frequency is 120MHz with a data transfer rate of 240MHz. In total three 16-bit wide memory chips per PP-FPGA provide a bandwidth of 11.5Gbit/s and a buffer size of 96MByte. With a bandwidth 1.5 times the required write bandwidth the read and refresh operations can be performed without the need of large de-randomizer buffer for the write data. To understand this statement it has to be stressed that on the write side of the memory the Level-0 accepted data is received from the links and no flow control has been foreseen. This data need to be

stored in a write de-randomizer buffer during read and refresh operation of the L1B. It is clear that for a fixed maximum latency Level-1 buffer, the buffer management is simple, the address in the buffer will be set equal to the lower part of the Level-0 event counter.

V. SYNCHRONISATION, DATA LINKING AND NETWORK INTERFACE

A block diagram for this functionality mostly implemented on a FPGA is given in Figure 5. The main tasks are clock distribution, event synchronization, Level-1 data linking, HLT data zero-suppression, HLT data linking, data framing and interfacing to the readout network.

The clock distribution is implemented with the on chip PLLs of the SyncLink-FPGA. The LHC clock recovered by the TTCrx is connected to one of the PLL on the FPGA and is distributed to other PLLs where all other necessary clocks are generated. Clocks with frequencies of 10MHz for the ECS, 80MHz for the pre-processing, 100MHz for the RO-TxCrd interface and 120MHz for the L1B are generated. Signal trace length compensation for the four PP-FPGAs is implemented using phase shifted clocks generated by a PLL.

To add the necessary event identification needed for the processing on the TELL1, the trigger farm and DAQ, the timing information from a local TTC receiver interfacing the SyncLink-FPGA is used. In a first step the TTCrx information is used to generate a local replication of the LHC bunch counter and Level-0 event counter (using the Level-0 Accept and Reset signals). These counters are stored in the Level-0 de-randomizer at each Level-0 accept. Two types of detectors have to be distinguished concerning the event identification information transmitted along the link.

All detectors not using the Beetle readout chip [11] provide themselves the data valid information through the link flow control and include in addition either part of the bunch counter or Level-0 event counter in the header.

Detectors using the Beetle front-end chip (VeLo, Pileup Veto, ST) provide as the only event identification in the header the Pipeline Column Number (PCN) corresponding to the pipeline location that has been used to store the data during Level-0 latency. To synchronize with the start of the event data sent over the links, a local emulation of the front-end chip state machine is required. As already successfully implemented in other high energy particle experiments, a local reference front-end chip is implemented on a small mezzanine card called FEM. The “Reference Beetle” provides among the PCN also a data valid signal which has a fixed time shift to the incoming data on the link. This allows masking the valid data, comparing the PCN and adding the additional event identification as bunch counter and Level-0 event counter to the incoming data.

L1 accepts are distributed either with the TTC “Short Broadcast” or “Long Broadcast” and are interpreted on the SyncLink-FPGA. To read out one event from the L1B, the only information necessary for the L1B controller is the

Level-0 event counter which corresponds to the physical location of the event in the buffer.

The IP destination, corresponding to the SFC IP address, for each MEP assembled on the SyncLink-FPGA is assigned using the TTC “Long Broadcast”. The destination is stored in a small FIFO from where it is read as soon the MEP is ready for Ethernet framing and transmission to the RO-TxCard.

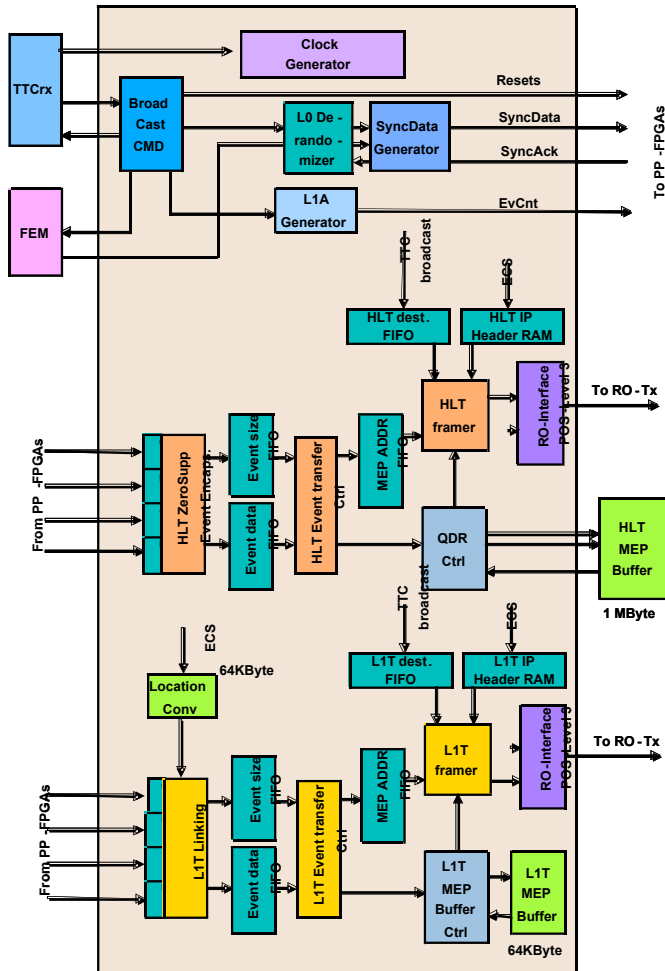


Figure 5: SyncLink-FPGA dataflow overview.

The Level-1 trigger data available on the PP-FPGAs are read into a first FIFO buffer where a final linking leads to the total event fragment on the board. Since not single but multiple events are sent in a packet to the readout, the data is stored on an internal memory (MEP output Buffer). The Ethernet framing can only start after a complete MEP has been assembled, since the total length and the number of frames per MEP have to be added in the header of the IP protocol.

The data arriving on the HLT links have to be zero suppressed at the first stage on the SyncLink-FPGA. The subsequent processing is similar to the Level-1. The only difference is the MEP Buffer being external to the FPGA implemented as dual port memory type QDR.

The interface to the readout network is implemented on a custom mezzanine card. For simplicity reasons the interface chosen is FIFO-like, based on the interface standard called

POS PHY Level-3. This is a commonly used interface for point-to-point data transmission between framer and Media Access Controller (MAC) for Gigabit Ethernet. This interface and the implementation of a dual Gigabit Ethernet card is described in [12] and also in these proceedings.

VI. HARDWARE IMPLEMENTATION

The board layout can be seen in Figure 6. The board is designed as a single slot 9U board compliant with the IEEE 1101.1 mechanical specification. A custom power backplane is used in the J1 position and the transition module space is used for cabling the TTC optical fibre, throttle signals, LAN connection for the ECS and the RO-TxCard.

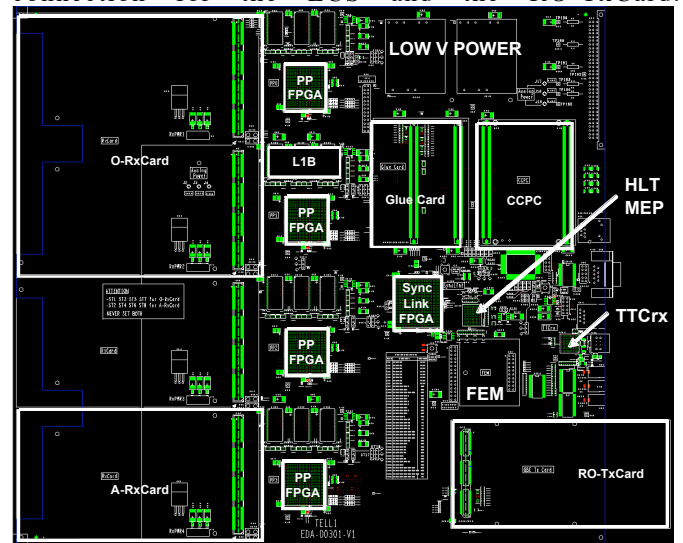


Figure 6: TELL1 board layout.

The fibre ribbon cables for the optical receiver and the copper cables for the analogue receiver are connected on the front panel. Altera Stratix FPGAs have been chosen where the PP-FPGAs are implemented in a 1S20F780-7 and the SyncLink-FPGA uses a 1S25F1020-7. Pin compatible devices with more logic gates do exist for both devices and can be soldered on the board without any changes on the PCB (see Altera Stratix migration path). These FPGAs provide a large number of logic gates, large, medium and small block memories of 64KByte, 4Kbit and 512bit size, PLLs, DSP blocks and all necessary interfaces to interface the external DDR and QDR memories. Special care has been taken to avoid signal integrity problems with the fast interfaces: Terminator technology provided by the FPGAs or discrete termination is used for all signals on the board. Impedance matched connectors with solid power plates are used to avoid signal degradation over the connectors. The layout for the DDR SDRAM and the QDR SRAM has been made with tight constraints on the trace length to obtain a maximum valid data window. All active components providing the boundary scan are connected to a JTAG chain for electrical in-system production testing. To control the various programmable devices on the board I2C, JTAG and a local parallel bus are

used. All interfaces are provided with the LHCb standard CCPC - Glue Card ECS system.

VII. SUMMARY

A readout board for 24 optical links running at 1.6GHz or 64 analogue channels sampled at 40MHz has been developed for the common needs of all detectors in LHCb. The board provides a data path for the topological trigger (Level-1) as well as for the high level trigger data. Sufficient buffer is provided to store the data during the Level-1 latency of 52.4ms and a memory upgrade path is foreseen. A large amount of programmable logic is provided to perform pedestal and common mode correction, zero suppression, linking, framing and setting up the required transport data format (IP) for the two output data streams. A dual or quad Gigabit Ethernet mezzanine is used to interface with the readout network.

VIII. REFERENCE

- [1] Requirements to the L0 front-end electronics, J. Christiansen, LHCb 2001-014.
- [2] Requirements to the L1 front-end electronics, J. Christiansen, LHCb 2003-078.
- [3] Technical Proposal, LHCb, CERN/LHCC 98-4.
- [4] LHCb Reoptimized Detector Design and Performance Technical Design Report, LHCb, CERN LHCC 2003-030.
- [5] LHCb Trigger System Technical Design Report, LHCb, CERN LHCC 2003-031.
- [6] Timing, Trigger and Control (TTC) Systems for the LHC, <http://ttc.web.cern.ch/TTC/intro.html>.
- [7] Controlling front-end electronics boards using commercial solutions, R. Beneyton, C. Gaspar, B. Jost, S. Schmelling, IEEE Trans. Nucl. Sci.: 49(2000) no. 2 pt. 1, pp.474-7.
- [8] Readout supervisor design specifications, R. Jacobson, B. Jost and Z. Guzik, LHCb 2001-012.
- [9] TELL1 Specification for a common readout board for LHCb, G. Haefeli, A. Bay, F. Legger, L. Locatelli, J. Christiansen, D. Wiedner, LHCb 2003-007.
- [10] LHCb VeLo Off Detector Electronics Preprocessor and Interface to the Level 1 Trigger, A. Bay, G. Haefeli and P. Koppenburg, LHCb 2001-043.
- [11] The Beetle reference manual, N. van Bakel et Al., LHCb 2001-046.
- [12] GiGabit Ethernet mezzanines for DAQ and Trigger links of LHCb, H. Muller et al., LHCb 2003-021.