Upgrade of the Second Level of the Readout Electronics for the CMS Drift Tubes Subdetector

Álvaro Navarro-Tobar on behalf of the CMS DT group

Abstract–**The readout server (ROS), which constitutes the second level of the CMS drift tubes (DT) subdetector readout architecture, is a complex VME 9U board, currently placed in the CMS cavern, in the racks on one side of the detector wheels. The planned upgrade of DT electronics for 2013-2014 includes the relocation of all the second level of readout and trigger electronics in the CMS counting room, allowing among others, the future redesign of a higher performance ROS board based on commercial devices with no radiation tolerance requirements.**

Simulation studies have been carried out in order to assess the system's performance under the increased luminosity (10^{35} cm^{-2} s^{-1}) planned for the HL-LHC upgrade, ca. 2020. Results show that the ROS board could become a limiting factor due to the event processing time.

The capabilities of currently-available FPGAs allow incorporating most of the ROS functions (input deserialization, input buffer, data processing and multiplexing, slow control interface, test-mode operation) into a single device. In particular, Virtex-6 and Spartan-6 series are being considered. In both families, fully-automatic asynchronous deserialization is carried out by the gigabit transceivers, which are not suitable for our application due to the minimum data rate and reduced availability. Nevertheless, asynchronous data reception can be carried out by making use of the dedicated deserializers present in each of the I/O tiles, plus some additional logic and clocking resources.

The improved performance of these devices (as compared with current ROS technology) allows reducing the event processing time, increasing maximum system's operation frequency.

I. Introduction

The Compact Muon Solenoid (CMS) experiment at CERN [1] aims to expand our knowledge by exploring the physics at the TeV scale, helping to clarify some of the questions still unresolved by the Standard Model. In order to do so, it measures the resulting particles from the proton and heavy ions collisions produced by the Large Hadron Collider (LHC). LHC has been operating for more than two years at center of mass energy of 7 TeV and achieving luminosities up to $2{\cdot}10^{33}$ cm^{-2} s^{-1}. The outermost layer of the CMS detector, the muon system, is in charge of identifying and accurately determining the trajectory described by muons as they are deflected by the magnetic field. In the barrel, the position of the muon is measured by the drift tubes (DT). In the DT cells, the distance to the central anode is calculated from the transit time of the electrons, produced by the ionization of the gas particles and accelerated by the intense electric field. The current pulse is amplified and shaped in the front-end boards (FEB), and its time stamp digitized in the readout boards (ROB) by CERN's ASIC HPTDC (High-Precision Time-to-Digital converter) [2]. The total number of DT cells in CMS is 172 200, and each ROB digitizes the hits from up to 128 of them. Each one of the 1500 ROBs stores the hits' time stamp and sends it to the readout server (ROS) board over a 240 Mbps differential twisted-copper pair when they match within 1.25-µs window determined by the reception of a Level-1 Assert (L1A) trigger signal. The 25 ROB links from each of the 12 sectors of each of the 5 wheels are concentrated into a ROS board [3], situated in racks close to the detector. The ROS is in charge of carrying out a basic processing of the information and multiplexing these data into a single 800 Mbps, fiber optic link. This link is routed through tunnels to the CMS counting room, approximately 60 m away. The ROS links are received by 5 device-dependent units (DDU) that further pack and process this information and hand it out to the data acquisition (DAQ) system for subsequent storage and analysis in the Grid.

II. Motivation

The readout system was designed in the late 90s, and was optimal in view of the expected data volume, the technology availability and cost at the time, and the tight restrictions imposed by CMS on the electronic design (space, tolerance to magnetic fields and radiation, power dissipation, reliability).

The DT readout system manages a volume of information that depends on several factors. The signal of interest is the trace left by the muons going through the detector. Additionally, several other hits are collected, caused by a variety of different particles generated by the LHC collisions, which constitute the background signal and contribute to the occupancy of the readout buffers and the decrease of the maximum processing speed. These two signals are dependent on the luminosity achieved by LHC, which is gradually approaching its nominal value, 10^{34} cm^{-2} s^{-1}. The LHC is planning a gradual upgrade that will lead to a luminosity of 10^{35} cm^{-2} s^{-1}, which is currently known as High-Luminosity LHC (HL-LHC).

During the ROS design phase, various analyses were performed [4] in order to study the expected occupancy of the detector from Monte-Carlo simulations.

The study showed that for LHC's nominal luminosity (10^{34} cm^{-2} s^{-1}), the occupation of the ROB FIFOs and the ROB-ROS link bandwidth is around 10%. Since only the payload, and not the protocol overhead, increases with luminosity, there is a reasonable security margin to operate the ROB under HL-LHC conditions. Similarly, the ROS-DDU link will also able to handle the increased data rate.

To assess the ROS performance, computer simulations of the different FPGAs firmware and interconnection have been

Manuscript received November 11, 2011.

A. Navarro-Tobar is with CIEMAT, Avda. Complutense 40, 28040 Madrid, Spain (tel: (+34)913466024, e-mail: alvaro.navarro@ciemat.es).

978-1-4673-0120-6/11/$26.00 ©2011 IEEE

conducted. These have allowed determining the event processing time as a function of the different possible hit patterns received from the detector. In the ROS board, the input data is processed in five different Spartan-IIe FPGAs, in order to increase the parallelism and reduce the processing time. However, due to the sequential nature of the state machine in each FPGA, the total processing time for a one-hit event varies amongst channels, from 75 to 115 bunch crossings (BX, period of time between proton bunches in LHC, corresponding to ~25 ns). With this information, the maximum sustainable L1A trigger rate (assuming the non existence of readout buffers) was evaluated as a function of the number of hits in the event, and is shown in Fig. 1, for the case of all hits in the same channel (averaged across the different channels' processing time), evenly distributed channels (multiples of 25 hits plus extrapolation) and muon-like distribution (multiples of 44 hits plus extrapolation).

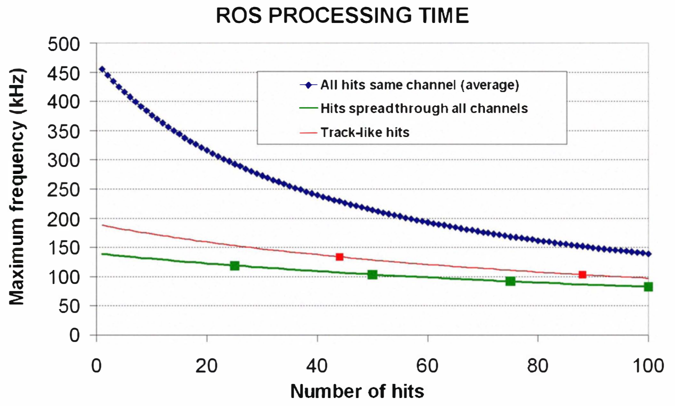

Fig. 1. Maximum sustainable L1A trigger frequency for the ROS as a function fo the total number of hits in the event, for different hit distributions across the input channels. The maximum frequency is calculated as the inverse of the event processing time.

The expected per-event muon and background hit count was inferred from the aforementioned Monte Carlo simulations, and the resulting event sizes seem consistent with extrapolations from experimental data obtained in 2010 at up to $2\cdot10^{33}$ cm^{-2} s^{-1} luminosity. Background signal is expected to be approximately 3 hits per event per sector. The processing time of the background-only events varies greatly with the spatial distribution of these background hits, yielding maximum L1A rates between 200 kHz and 460 kHz. The most probable hit locations are the ones corresponding to the first layer of DTs (MB1). Muon-only events are considerably bigger (a typical muon produces 44 hits), and, consequently, would take longer to process (maximum L1A of 134 kHz). However, with current luminosity levels, the number of muons per event is low, and the ROS board is performing properly, with no foreseeable problems in the short term.

Under the luminosity conditions of HL-LHC, however, both background hits and muon frequency are expected to increase in a factor 10. For 25 evenly-distributed background hits per event, the maximum L1A would decrease to 120 kHz. Adding the delay caused by processing of muons, whose frequency will approach 1 per event per sector, we obtain a processing time estimation that exceeds the threshold marked by the 100 kHz L1A rate.

III. Upgrade to the DT readout system

As a result of the marginal ROS processing time expected during HL-LHC, a new version of the board is being designed. In the new architecture, the ROS boards are to be placed in the counting room, with data being converted from copper links to fiber optic in the current ROS location.

The ROS board is being re-designed accordingly, and the first tests of the improved design are being implemented in Virtex-6 and Spartan-6 FPGA architectures. The increase in performance of these devices over the last years allows integrating most of the current ROS functionality in a single FPGA: input deserialization, FIFO buffers, improved data processing and multiplexing, slow control interface management and test-mode operation. In comparison, these tasks are carried out in the ROS by at least 57 different integrated circuits (ICs): one deserializer and one 4 kB FIFO per channel, 5 Spartan-IIe FPGAs for data processing and 2 Coolrunner-II CPLDs for test mode and slow control interface management. The physical signal reception differs in the new board, because we have switched from electrical to optical. Two high-integration MTP 12-fiber optical receiver modules and a dual LC receiver module are to be used. The output serialization and laser driving tasks are currently carried out by the CERN's GOL (gigabit optical link) IC [5], which plugs to the main ROS board in a mezzanine board that also includes the VCSEL.

In the next two sections we discuss our main conclusions regarding the two most critical issues of this new design: link deserialization and multiplexing performance.

A. Asynchronous deserialization

Although the LHC clock is available both at the cavern and the counting room, and it could be used for sampling input data, it was decided to make the deserialization asynchronous, with automatic clock recovery, as is currently the case in the ROS board. This way the reliability of the link is improved since slight frequency variations are allowed between transmitter and receiver.

Fully-automatic serial data reception with clock recovery is restricted in both Virtex-6 and Spartan-6 to the gigabit transceivers, which are available in reduced number and limited to frequencies above 480 MHz and 600 MHz, respectively. Therefore, we have investigated the implementation of asynchronous deserialization compatible with the present ROB-ROS link, since the transmitter part will remain unchanged. Asynchronous deserialization of a higher number of low-datarate links has been achieved in both platforms, by taking advantage of the dedicated serializer/deserializer (SERDES) module available in each IO pin, plus some additional logic.

In the Virtex-6 architecture, the transmission clock can be easily recovered from the data received from the ROB, which includes both a stop and a start bit in every 12-bit word. The data stream is fed to a clock buffer which is enabled during the stop bit and is disabled during the start bit, producing a lowduty-cycle signal that is introduced to a mixed-mode clock manager (MMCM) in order to generate the auxiliary clocks. Sampling time is adjusted by means of the input delay elements (IDELAY). A one-time calibration is needed in order to adjust the optimum sampling point in the center of the eye. This calibration is performed after reset, and involves cycling the input delay applied in FPGA before the SERDES through one bit period. The delay is only applied to the deserialized stream, not to the signal used in clock recovery, which allows maintaining a constant-phase clock. The location of the stop and start bits is obtained, and the delay value at which they shift position is marked as the edge delay. The optimum eye sampling point is set opposed to the edge delay value and words are aligned with the start bit. This calibration can be repeated whenever it is considered necessary, for example, when the receiver losses lock. It is noteworthy that, once the sampling delay has been established, there is no additional feedback on whether it continues being at the optimal position, and performing another routine calibration would cause data loss. However, this should not be necessary, because the clock's frequency and phase is extracted from the signal, and thus, should follow its variations. A problem could arise, however, if the LHC frequency varies since calibration occurred and there is a big difference in different links' lengths, because the clock recovered from one of them is used to deserialize all (the number of MMCM is not enough to do the clock recovery on a per-channel basis).

Although the problem was successfully solved and a test design experimentally validated in a Xilinx ML605 Virtex-6 test board, the high cost of the FPGAs from this family (around $600 for the simplest of them) led us to try to develop the ROS replacement in the much cheaper Spartan-6 family ($10-$300 price range). Our experimental tests were done using a Xilinx SP605 test board that includes a LX45T FPGA, which is a medium-sized Spartan-6 FPGA, with the highest speed grade (-3).

Porting the Virtex-6 deserialization schema previously explained to the Spartan-6 architecture is not possible. The configuration of the input-output blocks (IOBs) in Spartan-6 does not allow outputting the undelayed input signal to the FPGA fabric when it is being used as an input to the SERDES. Although this problem could be circumvented by introducing one of the links in two different IOBs, we found that the clock recovery was not possible due to limitations on the performance of the Spartan-6 clock buffers and clock management resources.

In the Spartan-6 architecture, each SERDES includes a phase detector that can be used to dynamically adjust the amount of input delay to ensure optimal data sampling, compensating from drifts in the IDELAY element. It works by taking additional "edge" samples of the incoming data stream with a half-bit delay from the "eye" samples. The comparison of these "edge" samples with the preceding and subsequent "eye" sample values allows determining the relative position of the sampling point to the data eye, and adjust it. However, it cannot be used to keep the sampling time in the center of the eye in case the deserialization clock is asynchronous to the data stream clock, because there is no control on the exact time at which the delay is changed, and, consequently, data bits are lost or repeated when the delay value transitions from its maximum to zero or vice versa. For this same reason, it is required that, even in the case of synchronous reception, the data and the deserialization clock maintain a certain phase relationship, only altered by small deviations (such as the ones caused by the drift in the IDELAY elements).

Taking advantage of the idea behind the phase detector, we have developed an asynchronous receiver that locks to the data stream eye by oversampling the incoming stream and analyzing this information to select the appropriate sample and use it as the bit's value. The incoming data stream is sampled in a SERDES at 960 MHz, 4 times its data rate (240 Mbps), with a clock that is asynchronous to the data's source clock. Consecutive samples are compared, and when inequality is found, it marks the presence of a data transition in the interval between those samples. Thus, it is safe to assume that, by taking any of the remaining two samples as the bit's value, the chosen sample is no more than 1/4 of a bit away from the center of the eye. This accuracy can be improved to 1/8 of a bit if we are able to identify which one of the two "eye" samples is closer to the center of the eye. In order to do so, we measure the time elapsed between shifts in the "edge" samples (i.e., the times when the transitions start to be detected between two samples different than the ones it was previously being detected at). This time, together with the time elapsed from the last shift in the "edge" samples, is used to select from the two "eye" sample candidates the one that is actually closer to the center of the eye. A schematic of this process is provided in Fig. 2. Note that in case the two clocks are exactly the same, there should be no changes in the samples in which the transition is detected, and therefore, there is no way of determining which one of the two "eye" samples is closer to the eye center.

The ability of this scheme to lock to the incoming data stream while varying the difference between the remote clock and the local clock was tested. Two different data patterns were used: a pseudorandom (PR) data stream and also the signal generated by a ROB board in the absence of valid data, which has two transitions in every 12-bit word. The lock was maintained for clock frequency differences up to 7.5 ‰ for the case of PR data, 5 ‰ for ROB data. This divergence responds to the fact that the generated ROB stream only has one transition every 6 bits, while the PR data has more. Also, it was observed that the lock performance is lost abruptly for the ROB data at 5 ‰ clock frequency difference, due to it being a deterministic signal. For the PR data, the period between losses of lock decreases gradually as the frequency difference is increased. As a reference, the LHC signal is expected to vary in maximum ± 3.5 kHz while it is ramping up the energy of the beams, which translates to a relative difference smaller than 0.1 ‰ (twice as much if the variation is considered relative to the lowest or highest frequency). Consequently, our system is able to lock to the LHC clock frequency variations

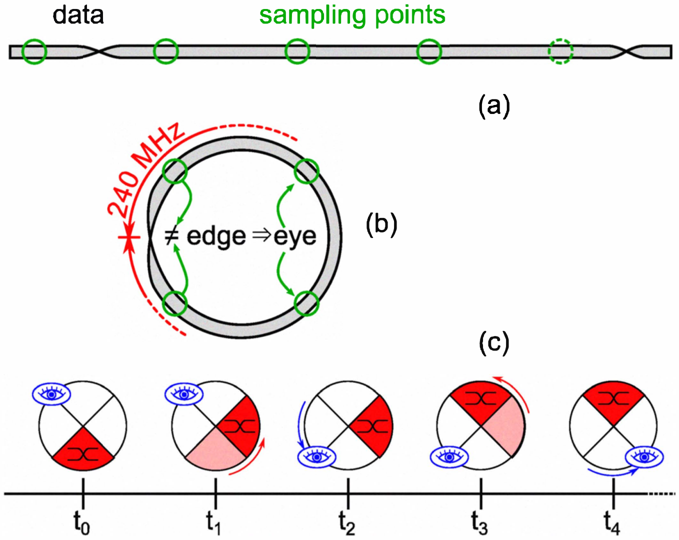

Fig. 2. Schematic view of the deserialization process. In (a), the data stream is shown, along with the sampling points, at a rate 4 times that of the data. In (b) the data stream has been wrapped to ease visualization of the transition detection and consequent selection of the two samples closest to the eye center in each cycle. In (c), the selection of the sample that is within 1/8 bit from the eye center is illustrated: initially (t_0) the transition is being detected between two samples, and at t_1 it moves to a different interval. After $\Delta t = t_3 - t_1$ the transition moves again, and $\Delta t/2$ after this time, at t_4, the optimum eye position is moved to the next sample.

that will happen at different times at the ROS and ROB boards due to its remote location. Furthermore, bit error rate (BER) was measured by means of a 65-hour test of PR data reception, and determined to be under 10^{-13}.

ROB data is sent through the serial link forming 12-bit words. Each of these words includes a stop bit and a start bit, with values 0 and 1. This ensures there are, at least, two transitions in each word. The stop and start bits are identified and used for word alignment. Two additional overhead bits (valid and parity) are used to flag valid data and to check for integrity. The remaining 8 bits are concatenated to form 32-bit words that constitute the ROB's minimum communication unit. These words are written to a FIFO for subsequent multiplexing into the output link.

B. *Multiplexing performance*

Because of the changes in the readout system, not all of the functionality of the ROS board will be re-implemented as-is in the new design, with changes involving mainly the test mode and the slow control interface management. However, the data multiplexing module, which is the most critical part for the elimination of the system bottleneck, has already been implemented.

The state machine that takes data from the 25 channel FIFOs and writes it to the output FIFO runs at 120 MHz. It writes the ROS header and trailer, and cycles through the 25 input FIFOs, with only one "dead" cycle in between channels. The time spent processing an event with one hit in each channel is 34 BX, approximately 10 times better than the current ROS, which spends 330 BX processing the same event. The additional delay it takes to process more than one hit in any channel is 1 BX / 3 hits, 6 times better than the current ROS (2 BX/hit). The approximate processing time for a typical event (1 muon + 25 background hits) would be 1.25 μs, yielding a maximum sustainable L1A rate of 800 kHz.

With this improved processing speed, the size of the FIFOs is of less importance. However, it is worth noting that the amount of buffer storage that is included in the ROS (25 × 4 kB) could be implemented in the RAM blocks of a $60 Spartan-6 FPGA (XC6S25LX-3). This FIFO depth would make the probability of data loss negligible, even under bursts in L1A trigger frequency.

In fact, because each one of the 25 channel FIFOs is written at 5 MHz, the new ROS would almost be able to keep up in the case where the ROB links contain valid data 100 % of the time. However, this would saturate the ROS-DDU link, whose current payload rate is only 640 Mbps.

IV. Conclusion

An upgrade for the CMS DT readout system is under preparation in view of the future plans of increasing LHC luminosity to 10^{35} cm^{-2} s^{-1}. The motivation, based in the limitations of the processing speed of present design, has been discussed in the text.

The implementation of a new ROS board in current FPGA technology has been studied. The advances in technology over the last decade allow implementing most of the ROS functionality into a single IC. Focus has been paid in the implementation of the 240 Mbps deserialization and its reliability. Different implementations have been developed both for Virtex-6 and Spartan-6 FPGAs, both of them showing very satisfactory results.

An asynchronous deserializer has been designed, implemented and tested in a Spartan-6 FPGA. The input data rate variation tolerance is enough to adapt to the changes in the LHC frequency, and BER tests show excellent results.

The multiplexing performance of the new design will be able to manage the workload expected for the operation of HL-LHC, guaranteeing its proper operation for the expected detector occupancy.

Acknowledgment

We would like to thank Cristina Fernández Bedoya and Carlos Willmott for their significant contribution to the work presented in this paper.

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC," *J. Instrum.*, vol. 3, S08004, Aug. 2008.

[2] HPTDC User Manual. http://tdc.web.cern.ch/TDC/hptdc/docs/hptdc_manual_ver2.2.pdf

[3] C. Fernández-Bedoya et al., "CMS drift tube chambers read-out electronics", *Topical Workshop on Electronics for Particle Physics*, Prague, 3-7 September 2007.

[4] C. Fernández-Bedoya, "Diseño, construcción y validación del sistema de adquisición de datos de las cámaras de deriva del experimento CMS", *Ph. D. Thesis*, Univ. Complutense de Madrid, 2010.

[5] P. Moreira, T. Toifl, A. Kluge, G. Cervelli, F. Faccio, A. Marchioro and J. Christiansen, "G-Link and Gigabit Ethernet compliant serializer for LHC data transmission", IEEE *Nuclear Science Symposium*, vol. 2, pp. 96-99, Oct. 2000.