Journal of Physics: Conference Series **513** (2014) 012010 doi:10.1088/1742-6596/513/1/012010

FPGA-based 10-Gbit Ethernet Data Acquisition Interface for the Upgraded Electronics of the ATLAS Liquid Argon Calorimeters

J. Philipp Grohs on behalf of the ATLAS Liquid Argon calorimeter group

IKTP, TU Dresden, Zellescher Weg 19, D-01069 Dresden, Germany

E-mail: philipp.grohs@tu-dresden.de

Abstract. A stepwise upgrade of the LHC is foreseen starting now until the year 2023 to increase the instantaneous luminosity up to five times of its design value. It implies a challenge for the ATLAS experiment coping with the expected event pile-up, especially for the Level-1 calorimeter trigger system. In order to keep the trigger rates within the limited bandwidth new algorithms have to be applied which in turn requires an upgrade of the ATLAS Liquid Argon calorimeter trigger readout electronics. Towards this upgrade, the ATLAS Liquid Argon calorimeter group develops a high-speed data acquisition interface in ATCA standard using commercial hardware instead of complex and expensive in-house developments where possible. This paper gives an overview of the general concepts of the DAQ interface, the engaged technologies and the current status of the development efforts for an FPGA based fast data link with a standard 10 Gbps Ethernet protocol which may also be useful for DAQ systems of other high energy physics experiments.

1. Introduction

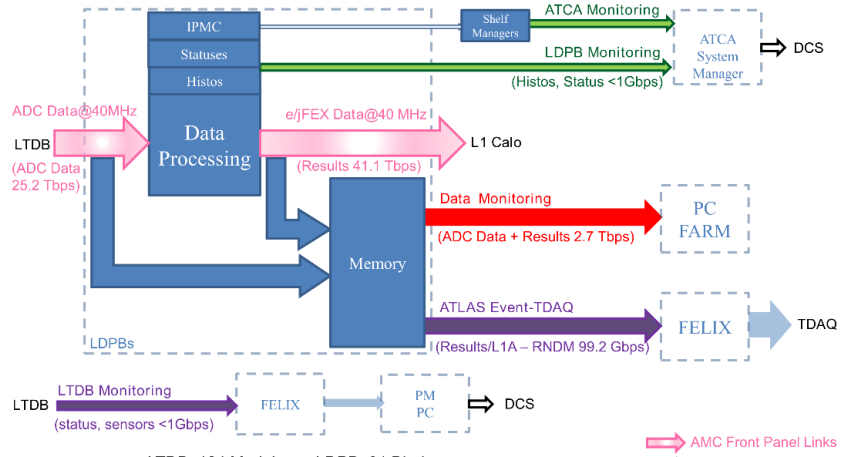
In order to exploit the full discovery potential of the Large Hadron Collider (LHC) [1] and the installed experiments there are two long shutdowns scheduled in the years 2018 (LS2), respectively 2022 to 2023 (LS3). It is thereby foreseen to increase the instantaneous luminosity up to at least five times of its current design value of 1×10^{34} $cm^{-2}s^{-1}$ [2]. The resultant expected event pile-up faces the ATLAS experiment inter alia with the challenge to keep the trigger rates of the hardware based first level calorimeter trigger (L1Calo) within the limited bandwidth of the data readout system while keeping the signal acceptance as high as possible. Therefore, an intensive upgrade and reconstruction program has been initialized during which at first the trigger readout electronics of the ATLAS Liquid Argon (LAr) calorimeter and finally the whole readout electronics of the LAr calorimeter will be replaced. The finer granularity which will be provided digitally to the L1Calo system after the first upgrade step during the LS2 increases the data rate which has to be handled by subsequent PC based trigger stages as well as by the data acquisition system storing trigger information. Therefore, reliable high-speed data links between the FPGA based trigger readout electronics of the LAr calorimeter and the subsequent PC farms are required.

In the following section an overview of the foreseen LAr calorimeter trigger readout electronics upgrade for LS2 with respect to the demand of the high-speed data link to the PC network is given, followed by the illustration of the general concept of the data link itself. Finally the technical details and performance results of the ongoing implementation within a test setup are described.

2. ATLAS Liquid Argon Calorimeter Electronics Upgrade

The upgrade program for the LAr calorimeter for the LS2 is described in detail in [3, 4]. With the envisaged improvements of the data granularity provided by the LAr calorimeter more sophisticated trigger algorithms can be applied by the L1Calo system, to exploit the significant differences of the shower profiles of the objects to be triggered and of their background.

The data flow model of the foreseen FPGA based off-detector electronics is depicted in Figure 1. Data of events triggered by the L1Calo have to be sent to the DAQ PC farm. Moreover, there are monitoring data to be transmitted from the LAr calorimeter electronics to a PC. As these links require less bandwidth (99.2 Gbps respectively 2.7 Tbps in total) and have less strict latency requirements than the continuous data transfer to the L1Calo (41.1 Tbps in total) it is possible to send the data directly to a PC using commercial hardware on the receiving side. Presenting the current state of the development of such links between FPGA and PC is the objective of this paper.



Data Flow Rates	ADC Data @40MHz	e/jFEX Data @40MHz	LDPB Monitoring	Data Monitoring	ATLAS Event-TDAQ	LTDB Monitoring	DCS
LTDB	204 Gbps	-	-	-	-	<<1 Gbps	<<1 Gbps
LDPB	814 Gbps	1.3 Tbps	<<1 Gbps	82 Gbps	3.2 Gbps	-	<<1 Gbps
GLOBAL	25.2 Tbps	41.1 Tbps	<1 Gbps	2.7 Tbps	99.2 Gbps	<1Gbps	<<1 Gbps

Figure 1. Data flow model of the LAr calorimeter off-detector electronics. The data paths are drawn as arrows. The Liquid Argon Trigger Digitization Board (LTDB) is located on-detector and sends the digitized trigger readout data to the off-detector Liquid Argon Processing Boards (LDPB) where FPGAs are performing the energy and time reconstruction (Data Processing). Raw and processed data, are stored in a memory to be sent to PC farms upon the trigger accept signal [4].

3. Concept of the Data Link between the Readout Electronics and the PC network

The implementation of the LAr calorimeter off-detector electronics is planned to be in ATCA standard [5]. This standard defines 10 Gbps XAUI connections between the boards inside one ATCA shelf through the backplane, which are suitable for the above mentioned data transmission between the off-detector electronics and the PC farm for DAQ and monitoring purposes.

The current data readout electronics of the LAr calorimeter necessitates so-called “ROBIN” PCI boards [6, 7] to transfer the data from the detector electronics into the PC environment with the required performance. Contrarily, today’s computer hardware technology provides sufficient processing power to make expensive custom hardware developments, as the ROBIN boards, obsolete. A high-speed link between the FPGA as part of the detector electronics and the subsequent data acquisition PC farm is therefore explored. Although there is commercial PC hardware utilized, some development efforts are required to exploit the full capacity of the 10 GbE connections:

On the one hand the FPGA firmware for the 10 GbE interface has to send data in a format the PC can cope with, at best at minimal consumption of FPGA resources. These constraints lead to the decision to use UDP/IP [8] as it is a standardized, message-based, connectionless network protocol with low overhead. Thus, its implementation in VHDL requires very few FPGA resources. The drawback of the UDP/IP protocol in contrast to TCP/IP [9] is its lack of reliability in terms of packet delivery. To overcome this issue, it is foreseen to utilize the UDP/IP encapsulated IPbus [10] protocol, which is a lightweight protocol for controlling hardware devices with the possibility included to resend lost packages. Although IPbus introduces a further protocol layer the total complexity of the resultant FPGA firmware is expected to be lower than of a possible TCP/IP implementation.

On the other hand, the receiving PC has to have sufficient computing power to receive the data and to transfer them to the existing ATLAS TDAQ software [11] which has to be adapted accordingly. Therefore, load tests with multiple 10 GbE connections in parallel with a single PC are required to identify the CPU time consumption due to the handling of the UDP/IP traffic and due to the ATLAS TDAQ software itself.

4. DAQ Test Setup

4.1. ATCA Crate and Components

The test setup with all its components is depicted in Figure 2 while a schematic can be seen in Figure 3. The ATCA boards are housed in a RadiSys Promentum ATCA Sys-6010 crate with Dual Star backplane [12]. Inside the crate one data processing board populated with a Xilinx Virtex 5 XC5VFX70T FPGA [13] simulates the LAr calorimeter off-detector electronics. The high-speed data link to the PC is realized through the communication over the backplane with a 10 GbE RadiSys ATCA 2210 switch blade [14] inside the same ATCA crate which forwards the data through optical fibers to a PC equipped with two Intel Xeon QuadCore X5550 CPUs with 2.67 GHz [15], 24 GB RAM and four dual-port 10 GbE Myricom 8B2-2S network interface cards [16].

Figure 2. Test setup with a RadiSys Promentum ATCA Sys-6010 crate with Dual Star backplane instrumented with a FPGA test board (right most board), a 10 GbE RadiSys ATCA-2210 switch blade (centered board) and a dual Intel Xeon QuadCore based PC with four dual-port 10 GbE Myricom 8B2-2S (bottom) network interface cards.

4.2. DAQ PC Performance Tests

Although currently only one 10 GbE connection between the FPGA and the PC is establishable due to the properties of the deployed ATCA backplane, multiple 10 GbE NICs have been installed in the PC to enable performance tests in a loopback mode. To ensure that the traffic is transferred physically through the optical fibers and is not redirected by the network driver directly to its destination, two virtual machines have been set up: The first one for sending random data and the second one for receiving them and copying them into a memory buffer. With optimized NIC drivers a total raw data rate of 79 Gbps at 30% CPU load has been observed with this configuration. Therefore, the selected PC hardware is expected to provide sufficient computing power to cope with multiple 10 Gbps connections in parallel.

4.3. FPGA Firmware and Switch Configuration

A schematic of the current test firmware for the FPGA on the test board together with the successive parts of the test setup are depicted in Figure 3. The 10 GbE link over XAUI is realized by the Xilinx XAUI IP core [17]. This IP core has been interconnected via the XGMI interface to the Xilinx IP core XGMAC [18], which handles the media access control. In the preceding user logic module currently a simple UDP/IP receiver and sender has been realized which listens for some basic configuration commands and sends data encapsulated in UDP/IP packages. To comply with the current ATLAS data readout system, on the sending path, the UDP payload is composed of a Readout Driver (ROD) header containing an ID number and the actual ROD data fragment. To check data integrity during the tests, the ID is based on a counter increased for each packed sent. The actual ROD data fragment will contain reconstructed hit energies, but is instead populated with pseudorandom data during the performance tests. One of the configuration commands allows to adjust the UDP package size. This permits to test the performance at different package lengths as well as adjusting the package length to the variable ROD fragment length.

A self-implied obstacle for the performance test of the test board is its lack of the IPMC interface. This have to be overcome by manually configuring the ATCA switch to open the port to the test board. Furthermore, some monitoring features of the switch which occasionally send Ethernet packages have to be disabled as they interrupt the data transfer between the FPGA and the PC and hence causes a considerable amount of package losses at highest transfer rates.

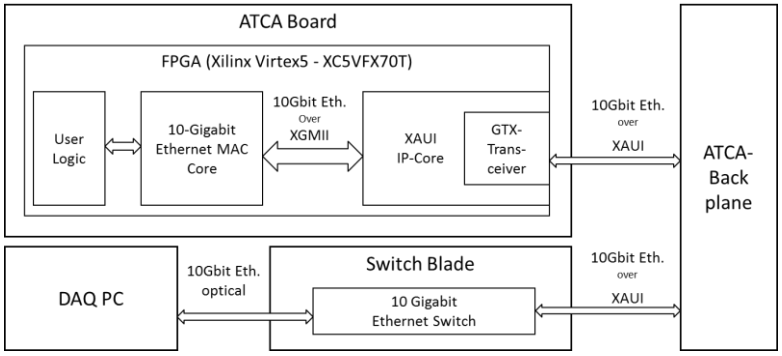


Figure 3. Schematic of the test setup with the utilized IP cores within the FPGA firmware.

4.4. DAQ PC Software and Results

Two types of DAQ PC software have been developed to test the UDP/IP link to the FPGA.

The first one is a standalone application which simply receives the ROD fragments and stores them in a ring buffer of a configurable size. It analyses if any ROD fragment was lost by means of the ID in the ROD header. Furthermore, the bit rate of the data transmission is measured. As expected the best performance is achieved with UDP jumbo frames of 9000 bytes length, which is the maximum size allowed. In this case a raw data rate of 9.98 Gbps without any ROD fragment loss within four hours has been measured. This data rate decreases with smaller UDP frame sizes as the overhead due to the UDP/IP header increases in this case.

The CPU affinity of our DAQ application and the NIC process were set to a specific configuration in order to measure the performance loads caused by the data transmission at 9.98 Gbps. A single core CPU load of 20% was observed for the DAQ application whereas the NIC driver caused additional 60% CPU load.

Beside the measurement of the speed and reliability of the UDP/IP connection the integration of the readout link into the ATLAS TDAQ software has been done. The current ATLAS TDAQ LAr calorimeter readout modules are based on TCP/IP standard. As the FPGA uses the UDP/IP standard the corresponding readout modules of the ATLAS TDAQ software has been adopted accordingly. First tests in conjunction with the FPGA show that this operation mode is functioning and transparent to the TDAQ system. Further performance tests are ongoing.

5. Conclusion and Outlook

The upgrade of the LHC luminosity and the resulting upgrade of the ATLAS detector require the availability of reliable high-speed connections between FPGAs of the detector readout electronics and subsequent DAQ PC farms. We have demonstrated that with today's commercially available hardware it is possible to send data from FPGAs through the backplane of an ATCA crate over a switch blade to a PC and to achieve almost the design transmission rate of 10 Gbps. Intensive performance studies with multiple connections between the FPGA and the PC running the fully adopted ATLAS TDAQ software are planned to be performed in the near future. Furthermore, the ongoing implementation of the IPbus protocol in the FPGA firmware will enhance the reliability of the link significantly.

Acknowledgement

This work was supported by the German BMBF within the research network FSP-101 "Physics on the TeV Scale with ATLAS at the LHC" and by the German Helmholtz Alliance "Physics at the Terascale".

References

[1] L. Evans, P. Bryant, LHC Machine, *Journal of Instrumentation*, vol. **3**, no. S08001, 2008.
[2] L. Rossi and O. Bruning, High Luminosity Large Hadron Collider: A description for the European Strategy Preparatory Group, 2012, CERN-ATS-2012-236
[3] ATLAS Collaboration, Letter of Intent for the Phase-I Upgrade of the ATLAS Experiment, 2011, CERN-LHCC-2011-012
[4] ATLAS Collaboration, ATLAS Liquid Argon Calorimeter Technical Design Report: Trigger Electronics for the LHC Phase-I Upgrade, 2013, to be published
[5] PCI Industrial Computers Manufacturers Group, AdvancedTCA: PICMG 3.0 Short Form Specification, January 2003, http://www.picmg.org/pdf/picmg_3_0_shortform.pdf
[6] B. Green et al., ATLAS Trigger/DAQ RobIn Prototype, *IEEE Trans. Nucl. Sci*., Vol. **51**, 2004, pp. 465-469
[7] B. Green et al., ATLAS TDAQ/DCS ROS ROBIN Production Readiness Review: Design Document, 2005, https://edms.cern.ch/document/555100/1
[8] J. Postel, RFC 768: User Datagram Protocol, 1980, http://tools.ietf.org/html/rfc768
[9] V.Cerf et al, Specification of Internet Transmission Control Protocol, December 1974, http://tools.ietf.org/html/rfc675
[10] R. Frazier, The IPbus Protocol: An IP-based control protocol for ATCA/µTCA, IPbus version 2.0, Draft 9, 19 December 2013, https://svnweb.cern.ch/trac/cactus
[11] ATLAS collaboration, ATLAS high-level trigger, data-acquisition and controls: Technical Design Report, CERN-LHCC-2003-022, ATLAS-TDR-016, 2003
[12] RadiSys, SYS-6010-10G: 0G ATCA system; 12U, 14 slots, http://www.radisys.com/products/atca/chassis-hub-systems/sys-6010-10g/
[13] XILINX, Virtex-5 Family Overview: Product Specification, February 6, 2009 http://www.xilinx.com/support/documentation/data_sheets/ds100.pdf
[14] RadiSys, ATCA-2210: Broadcom Base / Fabric 10G ATCA Switch & Control, http://www.radisys.com/products/atca/switch-control/atca-2210/
[15] Intel, Intel Xeon Processor X5550, http://ark.intel.com/products/37106/
[16] Myricom, 10G-PCIE2-8B2-2S: Two-Port 10-Gigabit Ethernet Network Adapters, https://www.myricom.com/images/stories/10G-PCIE2-8B2-2S.pdf
[17] XILINX, LogiCORE IP XAUI v10.3: Product Specification, April 24, 2012, http://www.xilinx.com/support/documentation/ip_documentation/xaui/v10_3/xaui_ds266.pdf
[18] XILINX, LogiCORE IP 10-Gigabit Ethernet MAC v11.3: Product Specification, April 24, 2012, http://www.xilinx.com/support/documentation/ip_documentation/ten_gig_eth_mac/v11_3/ds813_ten_gig_eth_mac.pdf