Massachusetts Institute of Technology

The FEROL40, a microTCA card interfacing custom point-to-point links and standard TCP/IP

Dominique Gigi[1]

CERN

23 Geneva, Switzerland

E-mail: `dominique.gigi@cern.ch`

Jean-Marc Andre (5), Ulf Behrens (1), James Branson (4), Olivier Chaze (2), Sergio Cittolin (4), Cristian Contescu (5), Diego Da Silva Gomes (2), Georgiana-Lavinia Darlea (6), Christian Deldicque (2), Zeynep Demiragli (6), Marc Dobson (2), Nicolas Doualot (5), Samim Erhan (3), Jonathan Richard Fulcher (2), Dominique Gigi (2), Maciej Gladki (2), Frank Glege (2), Guillelmo Gomez-Ceballos (6), Jeroen Hegeman (2), Andre Holzner (4), Mindaugas Janulis[a] (2), Michael Lettrich (2), Frans Meijers (2), Emilio Meschi (2), Remigius K. Mommsen (5), Srecko Morovic (2), Vivian O'Dell (5), Samuel Johan Orn (2), Luciano Orsini (2), Ioannis Papakrivopoulos (7), Christoph Paus (6), Petia Petrova (2), Andrea Petrucci (8), Marco Pieri (4), Dinyar Rabady (2), Attila Racz (2), Thomas Reis (2), Hannes Sakulin (2), Christoph Schwick (2), Dainius Simelevicius[a] (2), Cristina Vazquez Velez (2), Michail Vougioukas (2), Petr Zejdl[b] (5)

1. DESY, Hamburg, Germany
2. CERN, Geneva, Switzerland
3. University of California, Los Angeles, Los Angeles, California, USA
4. University of California, San Diego, San Diego, California, USA
5. FNAL, Chicago, Illinois, USA
6. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
7. Technical University of Athens, Athens, Greece
8. Rice University, Houston, Texas, USA

[a]) Also at Vilnius University, Vilnius, Lithuania
[b]) Also at CERN, Geneva, Switzerland

In order to accommodate new back-end electronics of upgraded CMS sub-detectors, a new FEROL40 card in the microTCA standard has been developed. The main function of the FEROL40 is to acquire event data over multiple point-to-point serial optical links, provide buffering, perform protocol conversion, and transmit multiple TCP/IP streams (4x10Gbps) to the Ethernet network of the aggregation layer of the CMS DAQ (data acquisition) event builder. This contribution discusses the design of the FEROL40 and experience from operation.

Topical Workshop on Electronics for Particle Physics
11 - 14 September 2017
Santa Cruz, California

---

[1]Speaker

1.   Introduction

The CMS experiment installed a new Pixel detector during the 2016-17 end-of-year technical stop including new readout electronics. In order to read out the 112 links (10 Gb/s) of the new Pixel back-end electronics, a new custom FEROL40 board has been developed. This microTCA based board is an evolution of the compact-PCI based FEROL board [1] developed in 2013. The FEROL interfaces the back-end electronics of the sub-detectors to Ethernet switches and computing nodes. The input links connecting to the sub-detector back-end electronics are point-to-point serial links using the CMS specific SlinkExpress protocol. The output transmits the data to the DAQ Ethernet network via an on-chip reduced TCP/IP implementation [2].

The new FEROL40 board is based on the microTCA standard with an architecture that maximizes the number of links per board. It has been designed to aggregate four links of 10 Gb/s and uses four 10 Gb/s Ethernet output links. The memory used for the TCP socket buffer is based on DDR3 modules with a bandwidth of 90 Gb/s. The FEROL40 can also receive the TCDS [3] (Trigger Control and Distribution System) information (triggers, commands) enabling the checking of synchronisation of the event fragments received from each link and emulation of event fragments to test the complete DAQ infrastructure. The software for configuration, control and monitoring accesses the FEROL40 on-board resources via a commercial memory mapped PCIe link connecting the microTCA MCH and a control PC.

2.   The FROL40 microTCA board

The FEROL40 is a microTCA board organized around an Altera Arria V GZ FPGA (as shown in Figure 1). Four inputs receive the fragments from FEDs (back-end electronics boards) of the new Pixel detector at 10Gb/s using the CMS DAQ SlinkExpress protocol (see Section 4). The output implements four 10Gb/s TCP/IP connections allowing the data to be sent to a PC with a standard NIC in the CMS DAQ system. Shared between input and output, two blocks of DDR3 memories operating at 350 MHz allow the storage of 4 TCP streams (45 Gb/s bandwidth per block). A connection to the TCDS was added in order to enable checking of event fragment
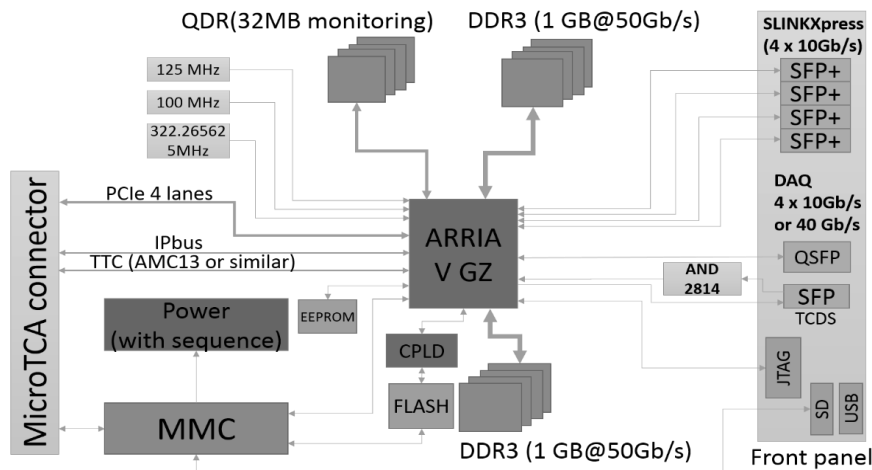

Figure 1: FEROL40 Block Diagram

alignment. This feature also enables FED emulation, generating the data inside the FEROL40 board on receipt of an external trigger. The MMC system [5] is implemented on the board to comply with the microTCA standard. All resources on the board are controlled and monitored using the PCIe bus of the microTCA crate. The components on the board are powered by 8 different voltages, each of which is generated by a DC/DC converter (12 V is used as main power) controlled by a power sequencer to follow the requirements of the FPGA and DDR3 powering sequence. The MMC is able to monitor the voltage and the current of each DC/DC converter.

3. PCB Design

The board has 16 layers and uses the NELCO 4000 13 EP-SI material, which has a constant dielectric value over a large signal frequency range. The dissipation factor is low in order to minimize the signal loss on traces. A special routing technique is used for the 16 pairs of signals operating at 10 Gb/s. It uses 5 layers spaced to increase the section of the lines and decreasing the physical impedance. Blind VIAs are used as well to avoid reflection that a stub would create. (see Figure 2). The DDR3 address, data and control buses have been routed taking care of the length (in order to minimize the signal length and to ensure the same length for all signals in the bus) and the impedance to be able to operate at the highest frequency allowed by the DDR3 technology.

Another parameter of the board is the thickness that should be controlled across the production because the board uses an edge connector to be inserted in the microTCA backplane.

Figure 2: PCB Stackup and 10Gb/s eyes

4. SlinkExpress

The link implemented between the FED and the FEROL40 uses a point-to-point lossless protocol on a duplex 10 Gb/s optical connection (Figure 3). The link is based on the CMS DAQ SlinkExpress protocol. From the FED side, the interface is seen as a FIFO. An IP-core implements

Figure 3:SlinkExpress interface Block Diagram

a common interface across all FEDs. Data are sent in blocks of 4 kB and retransmitted in cases of transmission error or back-pressure. This gives low latency as soon as back-pressure is de-asserted. The DAQ system can monitor registers and control a data fragment generator instantiated inside the SlinkExpress IP-core, implemented on the FED side.

5. TCP/IP output

The output uses a QSFP+ 4x10 Gb/s transceiver to send the data to the DAQ system (via Ethernet switches) using the TCP/IP protocol. This protocol is implemented in logic gates (no CPU) as it was for its predecessor, the FEROL board [2].

The FEROL40 implements a simplified (uni-direction) version of the TCP/IP protocol - data flows from the FEROL40 to a receiving PC. Only control (acknowledgement) packets are allowed to go back. The FEROL40 doesn't accept any connection open request and, hence, works as a TCP client only. The amount of DDR3 memory installed on the FEROL40 for the TCP socket buffer, enables to avoid any back-pressure in the TCP stream up to 800 ms at 10 Gb/s.

All basic protocols like ICMP, probe ARP, PING are also implemented in order to be able to debug connection problems.

6. Memory (TCP stream)

Two blocks of DDR3 are used to buffer the four TCP streams between the input and the output. The DDR3 operates at 350 MHz and the bandwidth of a block is sufficient to handle the TCP stream of two links and two SlinkExpress links (45 Gb/s throughput). The memory was tested successfully up to 550 MHz (70 Gb/s throughput). The DDR3 interface is implemented with an IP from Altera. A "glue" layer is added on top of the IP to manage the arbitration between read and write cycles. Additional buffers are implemented in FPGA block RAM between the DDR3 IP-core and the 4 links (2 In , 2 Out) in order to smooth the data flow (due to latency requests on DRR3 IP interface, DDR3 refresh and arbitration,..). Altera tools have been used to validate the PCB production and measure the DATA timing margin for read and write accesses. The results of the margin are shown in the Figure 4; the green part shows that there is sufficient margin.
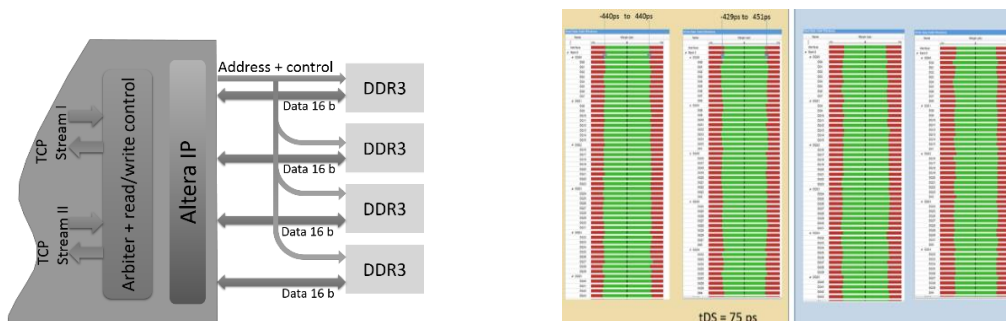

Figure 4: DDR3 implementation

7. Additional functions

The board includes additional functions. The MMC implemented with an Atmel 32-bit processor includes a user part. This allows the reprogramming of the FPGA flash using the SD

card. The main FPGA firmware is loaded using a CPLD. This component reads a NAND flash memory which can be reprogrammed also using the MMC user part.

A TCDS [3] connection via an SFP cage is implemented to receive the TTC (Trigger, Timing and Control) and send the TTS (Trigger Throttling System) signals. This function is useful to use the board in a simple system. For a larger system with multiple boards, the FEROL40 has connections on the backplane to communicate with the AMC13 hub board [4] to receive the TTC and send the TTS signals. In latter case, the AMC13 is connected to the TCDS .

The microTCA Ethernet backplane lines are connected to the FPGA. This gives the possibility to control the board via the IP-bus [6] or via an Ethernet connection using an embedded processor in the FPGA (like the NIOS).

8.   Conclusions

The FEROL40 has been developed and produced in less than 1.5 years. The production has been done in a single cycle due to the tight schedule. A number of minor issues emerged which could be addressed without requiring a new production cycle.   An add-on board (power card) has been developed and adapted on the FEROL40 to replace one malfunctioning DC/DC converter. Remote update is not possible due to an FPGA IP-core constraint not noticed before the production. The tri-state data bus cannot be shared between the Altera IP-core and the user logic needed to reprogram the Flash via the PCIe bus. The NAND flash has to be replaced by NOR flash for the future production to avoid the complexity of the Error Correction Code used by NAND to correct single bit errors on read operations.

A total of 50 boards have been produced out of which 49 passed the acceptance tests. Thirty-two boards are installed in three microTCA crates to readout the data from the 112 Pixel FEDs. All FEROL40s are operating according to all functional and performance requirements for the new Pixel detector in CMS and data was acquired successfully in the 2017 LHC run.

References

[1] H.Sakulin et al.,"The new CMS DAQ system for run-2 of the LHC", IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 62, NO. 3, JUNE 2015

[2] P. Zejdl et al., "10 Gbps TCP/IP Streams from the FPGA for the CMS DAQ Eventbuilder Network", *Journal of Instrumentation* **8** (2013), no. 12, C12039

[3] J.Hegeman et al, "The CMS Timing and Control Distribution System", Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015 IEEE; DOI 10.1109/NSSMIC.2015.7581984

[4] E. Hazen et al., "The AMC13xg: a new generation clock/timing/daq module for CMS microtca", Journal of Instrumentation (2013), no. 12, C12036

[5] MMC(Module Management Control) board developed by CERN EP-ESE group.

[6] C. Ghabrous Larreabet al, "IPbus: a flexible Ethernet-based control system for xTCA hardware", DOI 10.1088/1748-0221/10/02/C02019