CRI board for CBM experiment - preliminary studies

Wojciech M. Zabołotny[a], Adrian P. Byszuk[a], David Emschermann[b], Marek Gumiński[a], Dirk Hutter[c], Grzegorz H. Kasprowicz[a], Krzysztof T. Poźniak[a], and Ryszard Romaniuk[a]

[a]Institute of Electronic Systems, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warszawa, Poland
[b]GSI-Helmholtzzentrum für Schwerionenforschung GmbH, Planckstraße 1, 64291 Darmstadt, Germany
[c]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Straße 1, 60438 Frankfurt am Main, Germany

ABSTRACT

The Common Readout Interface (CRI) is an important component of the new architecture of readout and DAQ chain for CBM. The paper presents the results of preliminary analysis and experiments performed to assess the possibility to implement the CRI firmware in the selected prototyping hardware platform. It also reviews functionalities provided by the Zynq UltraScale+ platform regarding their usability for the creation of the PCIe-based data concentration system.

Keywords: FAIR, CBM, readout chain, FPGA, DAQ

1. INTRODUCTION

The Compressed Baryonic Matter (CBM) experiment is a fixed-target experiment prepared at the new accelerator complex FAIR in Darmstadt. Its aim is the analysis of nucleus collision at high energies, allowing exploration of the QCD phase diagram in the region of high baryon densities.[1,2]

The products of collisions will be recorded by multiple particle detectors: Micro Vertex Detector (MVD), Silicon Tracking System (STS), Ring Imaging Cherenkov Detector (RICH), Muon Chamber (MUCH), Transition Radiation Detector (TRD), Time of Flight Detector (TOF), Projectile Spectator Detector (PSD).

The CBM readout chain is a distributed system connecting the detectors' Front End Boards (FEBs) with the Data Acquisition System. The electronics controlling the FEBs and receiving the measurement data will be placed in the service building, located in the vicinity of the detector. In the case of STS and MUCH readout, the FEBs will be equipped with STS/MUCH-XYTER2 (SMX2)[3] ASICS, connected via copper links to the Common Readout Boards (CROBs).[4] The CROBs implement a bridge between the SMX2 links and the GBT[5,6] links used to transmit the control signals and the measurement data to the CBM service building. The data received from CROBs should be here concentrated and transmitted via long distance (ca.700 m) links to the First Level Event Selector (FLES) computer cluster located in the central data center. The previous design of the readout chain assumed using two layers of the FPGA-based boards. The first layer, located in the MTCA crates in the service building should be the Data Processing Boards (DPB)[7,8] used to control the FEBs, receive the hit data via the 4.8 Gb/s GBT links, and send the concentrated data the 10 Gb/s long-distance optical links to the second layer. The second layer, located in the entry node computers in the computer center, should be the FLES Interface Boards (FLIB).[9,10] They should receive the data in the form of so-called microslices, and transfer them via the PCIe interface into the memory of the FLES entry nodes. Further processing and building of so-called time slices is performed by the InfiniBand interconnected FLES nodes. The block diagram of the readout chain based on that concept is shown in Fig. 1.

Unfortunately, this approach increases the cost of the readout chain due to a higher number of FPGA chips needed and blocks the possibility to utilize the newest technology of optical links available at the time of assembling of the readout chain. The optical transceivers supporting the long-distance link must be selected at the time of designing of the final
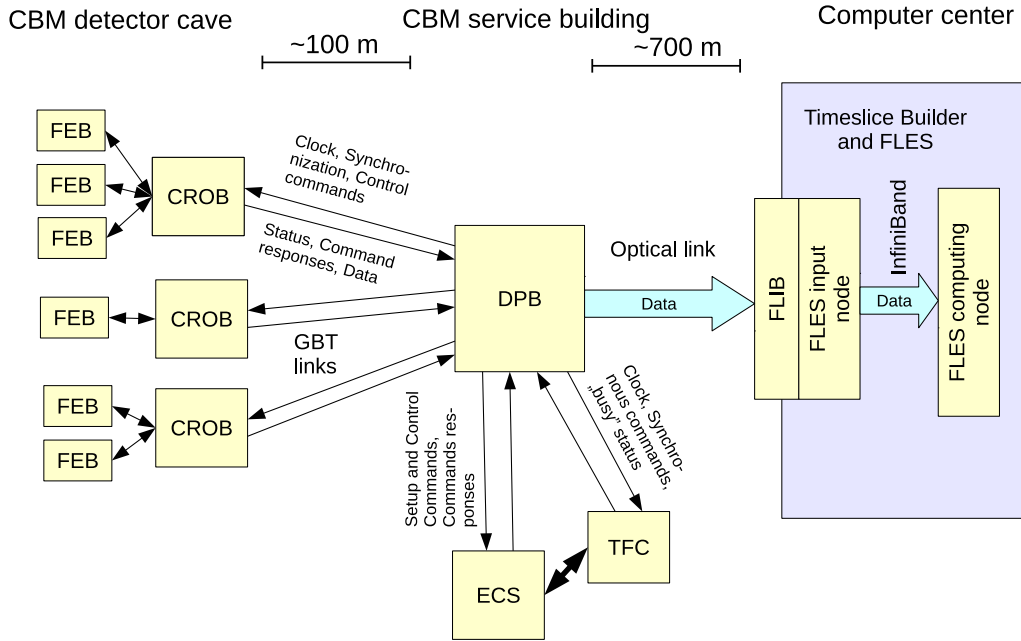

Figure 1: The old proposition of the CBM readout chain with separate DPB and FLIB boards connected via proprietary long-distance optical link [11].

version of DPB and FLIB boards. Those limitations may be worked around by moving the FLES entry nodes to the service building, and joining the DPB and FLIB boards into a single PCIe board - so-called Common Readout Interface (CRI) board. A similar approach is also adopted by other experiments like LHCb[12–14] and ATLAS[15] detectors at LHC. The only specialized optical interface incorporated in the CRI must support the GBT link, and the technology used is already defined by the design of the GBTx ASIC. Therefore, the right transceiver may be selected in advance. The long-distance links in this solution may be implemented using the standard network adapters (Ethernet or InfiniBand), that should minimize the cost and allow using the best technology available at the time when the whole system will be assembled. The block diagram of the modified DAQ chain is shown in Fig. 2.

In the last year, the studies were performed to find the appropriate candidates for the CRI board prototype. The technical requirements have been defined[11, 16] and the appropriate board was selected.

2. REQUIREMENTS FOR THE CRI BOARD

Main requirements for the CRI board candidate are:

- Sufficient (up to 24) number of multi-gigabit transceivers capable of communicating with the GBTX chip at 4.8 Gb/s.

- High-speed PCIe interface - either 16xGen3 or 8xGen4.

- Availability of yet another multi-gigabit transceiver with the possibility to receive the Timing and Fast Control (TFC) signal transmitting the reference clock and synchronization signals (Pulse Per Second - PPS).

- Possibility to implement the jitter cleaner for the reference clock, and availability of the configurable clock network allowing routing the cleaned reference clock to the GBT multi-gigabit transceivers.
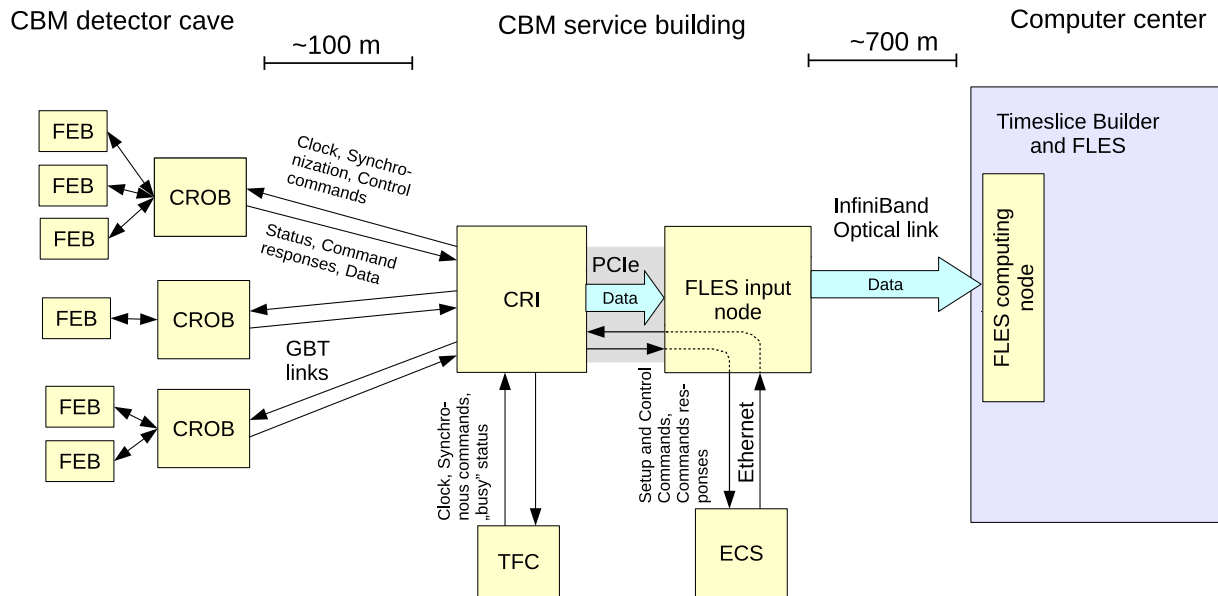
Figure 2: The CBM readout chain architecture with FLES entry nodes located in service building and hosting the CRI boards. The long-distance optical link is implemented with COTS components. [11]

3. SELECTION OF THE PROTOTYPE FOR CRI BOARD

Basing on the above requirements, the analysis performed in [11], and suggestions from the Xilinx[17,18] the HTG-Z920 board was selected. Unfortunately, at the time when this decision was taken, there was no comparable board based on Kintex UltraScale+ FPGA. However, as reported in [11] usage of Zynq UltraScale+ instead of Kintex UltraScale+ FPGA should not have a significant impact on the price of the board. The HTG-Z920 board fulfills the requirements for the CRI supporting up to 8 CROBs, i.e., provides 24 links available for GBT connection (it will be further denoted as CRI-24).

The HTG-Z920 board is equipped with a ZU19EG Zynq UltraScale+ FPGA, offering

- a PCIe x16 Gen3 Interface (using 16x GTH transceivers)

- 16x GTY and 16x GTH transceivers to be used as GBT link inputs.

The number of gigabit transceivers available in HTG-Z920 is sufficient to implement the CRI-24. Unfortunately, the real achievable number of links is limited by connection of those transceivers to the board connectors. Only 16 GTY (up to 30.5 Gb/s) transceivers are connected to the FMC+ connectors, therefore even if we use the Module,[19] offering 6 QSFP28/QSFP+ cages, only 4 quads are connected. Additional limitations result from the fact that one transceiver is needed for the TFC connection. If it is not possible to combine TFC and GBT links in the same quad, then only three quads are available for the GBT links. The simplest solution is to use a single quad for each CROB connection. Therefore implementation of CRI-9 should be easily achievable. If the CROB may be connected to links belonging to different quads (that depends on the results of porting of the DPB firmware to the UltraScale+ platform and on the possibility to conveniently route fibers from different 4-port connectors), then 12 or 15 GBT links are available, and the CRI-12 or CRI-15 board may be implemented. Additional 16 GTH transceivers (up to 16.3 Gb/s) are available via a special Z-RAY[20] connector. However, the currently available Z-RAY modules offer access to the maximum of 12 transceivers.[21,22] Again, if it is not possible to spread CROB GBT links between quads, it offers us only 9 usable GBT links, giving CRI-18 or CRI-21. If the arbitrary allocation of CROB links is possible, we may obtain CRI-24 or even CRI-27 (not planned at the moment).

Possibility of quad sharing	TFC-GBT sharing possible	No	Yes	No	Yes
	CROB GBT links may be spread across quads	No	No	Yes	Yes
# of GBT links (# of CROB boards)	Only FMC+ module	9 (3)	12 (4)	12 (4)	15 (5)
	FMC+ and Z-RAY FireFly modules	18 (6)	21 (7)	24 (8)	27 (9)

Table 1: Number of GBT links and number of CROB boards supported depending on modules used and the possibility to share links between quads.

In the near future, a 6-Port FireFly FMC+ module[23] should be available, that allows connecting the FireFly transceivers via the FMC+ connector. Of course, due to HTG-Z920 limitations, it also allows using only 16 links. However, it eliminates the inconvenient usage of two different standards of optical transceivers.

Possibilities to connect the GBT optical links to the HTG-Z920 board are summarized in Table 1

4. MODIFICATION OF THE HTG-Z920 FOR USAGE AS A FINAL CRI BOARD

It seems that HTG-Z920 may be used as a basis for the final version of the CRI board after moderate modifications.

- The Vita 57.4 FMC+ and Z-RAY connectors should be removed, and instead, 2 FireFly-TX and 2 FireFly-RX modules should be placed on the board, providing 24 GBT links via 12 GTH and 12 GTY transceivers.

- One of unused GTH transceivers should be connected to a standalone SFP cage for TFC connection.

- Unused 3 GTH and 4 GTY transceivers should be connected to QSFP+ cages if there is still free place on the board.

So modified HTG-Z920 board should be usable as a CRI-18 or CRI-24 for the CBM experiment (depending on a possibility to spread GBT links between quads).

5. MOCK-UP DESIGN AND SYNTHESIS RESULTS

To check the viability of HTG-Z920 as the CRI prototype, the mock-up of the firmware was created. Of course the mock-up design was not intended for running in the real hardware. Its purpose was only to verify synthesizability of required functional blocks and assess the resource consumption. The current version of the GBT-oriented DPB firmware was used as a starting point. However, as that design was developed for Kintex 7 based AFCK board, it was necessary to introduce certain changes ensuring that the firmware blocks do compile correctly and are not optimized out, what is possible if the synthezis tools detect that the particular block is misconfigured and therefore disabled.

The especially careful adjustment was needed in GBT-FPGA blocks because the gigabit transceivers available in the UltraScale+ chips differ from those used in Kintex 7.

Other blocks that required similar patching were the TFC receiver and the IPbus subsystem used to implement the slow control. In the final version, the IPbus subsystem will be replaced with registers accessible via the AXI to Avalon Memory Mapped bridge (AXI AMM bridge) IP core, but it should not significantly modify the amount of resources used by the slow control.

To assess the resources consumption, the FLIM module was replaced with the PCIe-based data concentrator built from two Xilinx blocks:

Resource	Available	Used	% used
CLB LUTs	522720	375553	71.85%
CLB Registers	1045440	397984	38.07%
Block RAM Tiles	984	574	58.33%
URAM	128	0	0%
DSP48E2	1968	864	43.90%
Global clock buffers	940	93	9.89%
MMCM	11	2	18.18%
GTHE4_CHANNEL	32	25	78.13%
GTYE4_CHANNEL	16	16	100%

Table 2: Results of synthesis and placement of CRI mock-up firmware in the xczu19egffvc1760-2 device.

- DMA/Bridge Subsystem for PCI Express (PCIe) v4.0[24]

- AXI4-Stream Interconnect[25]

The whole mock-up design implemented the CRI version with 24 GBT links, able to support 8 CROB boards. The mock-up firmware supports two selectable configurations of CROB boards:

- the first one with a single uplink from each SMX2 chip (5 FEB-8-1 boards with 8 SMX2 chips connected to each CROB).

- the second one with 5 uplinks from each SMX2 chip (1 FEB-8-5 board with 8 SMX2 chips connected to each CROB).

The design was successfully synthesized and placed, and post-implementation resource consumption was evaluated. The results are shown in Table 2. Obtained results suggest that it should be possible to implement the CRI-24 functionality in the HTG-Z920 board. However, the LUT consumption is high and near to the value where routing problems may occur.

6. FURTHER WORK

The results obtained with the mock-up design have shown that the HTG-Z920 may be used for CRI-24 prototyping. However significant amount of further work is still needed, that is associated with design decisions and full porting of different subsystems to the new UltraScale+ platform.

6.1 Clock distribution system

One of the important requirements for the CRI board listed in section 2 is the distribution of the reference clock received from TFC to the FEE connected via GBTX. The HTG-Z920 board uses the SiLabs Si5341 chip,[26, 27] and reference clocks for different GTH/GTY quads are connected to its different outputs. The chip also offers internal clock crosspoint, so it is possible to distribute the same clock frequency to different quads.

However, it is necessary to verify that the chip allows implementation of the jitter cleaner with sufficiently low jitter and clock skew between different outputs.

6.2 GBT-FPGA blocks

The GBT-FPGA blocks[28] are essential components for communication with GBTX chips located in CROB boards. The difficulty of porting the GBT-FPGA to UltraScale+ platform is increased by the fact, that this block uses advanced functions of multi gigabit transceivers to ensure the deterministic latency of the communication in the downlink direction, which is required to allow distribution of the reference clock and synchronization commands to the FEE ASICs. Therefore the final implementation will require thorough selection of individual settings and their verification in hardware.

6.3 PCIe transfer of hit data

Transfer of microslice data to the hosts memory previously fell into the regime of the FLIB firmware design and needs to be integrated into the CRI firmware [29]. The interface to the subsystem design can follow the same semantics as the current FLIM interface, however it may need adoption to the new data rates. The host interface should be kept independent from the detector specific part in order to share it between all CRI designs for different subsystems. The design, including data preparation and DMA engine may be ported from the current, Kintex 7 based FLIB firmware to the CRI. However the Ultrascale+ PCIe block differs greatly from the current design rendering this a non trivial task. Currently under investigation is the option to incorporate the Xilinx XDMA core instead of the current DMA engine into the design. However it offers only 4 (instead of 16) virtual channels to the host memory and implies development of a completely new device driver library and application interface.

A central question that needs further investigation is how many logical output streams the CRI should create. Merging all GBTx links into one single stream may need excessive resources for proper stream merging. On the other hand transferring multiple streams to the hosts memory increases resource consumption for the DMA design and leads to significant data handling overhead within the subsequent FLES processing stages.

6.4 Slow control subsystem

Replacement of IPbus with PCIe-based slow control requires checking if the PCIe express interface heavily loaded with the transfer of hit data may provide reliable and efficient transport of slow control commands and responses. Former experiences with the XDMA IP core have shown, that at high DMA traffic, the register access via PCIe may be significantly slowed down - even up to 1 ms per operation.

Fortunately, a viable alternative exists. The Zynq UltraScale+ chip used in the HTG-Z920 provides also a powerfull multicore, 64-bit ARM processor, that may be used to implement slow control and diagnostics.

In that case the access to the registers in programmable logic (PL) may be performed via separate AXI interface, not interfering with the high-speed transfer of measurement data via PCIe. The ARM processor may be accessible via Ethernet network.

7. CONCLUSIONS

The HTG-Z920 board seems to be a reasonable prototype for the CRI board for CBM. Initial analysis of functionalities offered by that board confirm that it fulfills requirements for the CRI. However, depending on the possibilities of sharing the GTH/GTY quads the maximum number of supported GBT links may vary between 18 and 27. Previous experiences with Kintex 7 chips suggest, however, that 24 GBT links should be available.

Results of synthesis of the initial version of the firmware, not working, but with all important functions implemented, show that for 24 GBT links the resource consumption is high but still acceptable.

According to preliminary analysis, a moderate modification is needed to convert HTG-Z920 into the final version of CRI-24. However, further firmware development and testing are needed to confirm that acceptable parameters, especially those related to reference clock recovery and distribution are achievable.

REFERENCES

[1] Ablyazimov, T., Abuhoza, A., Adak, R. P., and etal, "Challenges in QCD matter physics –the scientific programme of the Compressed Baryonic Matter experiment at FAIR," *The European Physical Journal A* **53**(3), 60 (2017).

[2] "CBM - The Compressed Baryonic Matter experiment." http://www.fair-center.eu/for-users/experiments/cbm.html .

[3] Kasinski, K., Szczygiel, R., and Zabolotny, W., "Back-end and interface implementation of the STS-XYTER2 prototype ASIC for the CBM experiment," *Journal of Instrumentation* **11**(11), C11018 (2016).

[4] Lehnert, J., Byszuk, A., Emschermann, D., Kasinski, K., Müller, W., Schmidt, C., Szczygiel, R., and Zabolotny, W., "GBT based readout in the CBM experiment," *Journal of Instrumentation* **12**(02), C02061 (2017).

[5] Moreira, P. et al., "The GBT-SerDes ASIC prototype," *Journal of Instrumentation* **5**(11), C11022 (2010).

[6] Moreira, P. R. S., "The radiation hard GBTX link interface chip," (Nov. 2013). https://indico.cern.ch/event/267408/ .

[7] Zabołotny, W. M. and Kasprowicz, G., "Data processing boards design for CBM experiment," *Proc. SPIE* **9290**, 929023–929023–11 (2014).

[8] Zabolotny, W., Kasprowicz, G., Byszuk, A., Emschermann, D., Gumiński, M., Juszczyk, B., Lehnert, J., Müller, W., Poźniak, K., and Romaniuk, R., "Versatile prototyping platform for Data Processing Boards for CBM experiment," *Journal of Instrumentation* **11**(02), C02031 (2016).

[9] Hutter, D., de Cuveland Jan, and Volker, L., "CBM FLES input interface developments," in [*CBM Progress Report 2015*], Friese, V., Sturm, C., and Toia, A., eds., 112, GSI, Darmstadt (2016). http://repository.gsi.de/record/186952.

[10] Hutter, D., de Cuveland Jan, and Volker, L., "Evaluation of the CBM FLES demonstrator," in [*CBM Progress Report 2016*], Selyuzhenkov, I. and Toia, A., eds., 157, GSI, Darmstadt (2017). http://repository.gsi.de/record/201318/.

[11] Zabołotny, W. M., Kasprowicz, G. H., Byszuk, A. P., Emschermann, D., Gumiński, M., Poźniak, K. T., and Romaniuk, R., "Selection of hardware platform for CBM common readout interface," *Proc. SPIE* **10445**, 1044549–1044549–8 (2017).

[12] Cachemiche, J.-P., Duval, P.-Y., Hachon, F., Gac, R. L., and Marin, F., "Study for the LHCb upgrade read-out board," *Journal of Instrumentation* **5**(12), C12036 (2010).

[13] Bellato, M., Collazuol, G., D'Antone, I., Durante, P., Galli, D., Jost, B., Lax, I., Liu, G., Marconi, U., Neufeld, N., Schwemmer, R., and Vagnoni, V., "A PCIe Gen3 based readout for the LHCb upgrade," *Journal of Physics: Conference Series* **513**(1), 012023 (2014).

[14] Cachemiche, J., Duval, P., Hachon, F., Gac, R. L., and Réthoré, F., "The PCIe-based readout system for the LHCb experiment," *Journal of Instrumentation* **11**(02), P02013 (2016).

[15] Anderson, J., Bauer, K., Borga, A., et al., "FELIX: a PCIe based high-throughput approach for interfacing front-end and trigger electronics in the ATLAS Upgrade framework," *Journal of Instrumentation* **11**(12), C12023 (2016).

[16] Zabołotny, W., Kasprowicz, G., Byszuk, A., Gumiński, M., Poźniak, K., and Ryszard, R., "Towards the common readout interface boards for STS," in [*CBM Progress Report 2016*], Selyuzhenkov, I. and Toia, A., eds., 150–151, GSI, Darmstadt (2017). http://repository.gsi.de/record/201318/.

[17] Xilinx, "Kintex UltraScale+ Boards, Kits, and Modules." https://www.xilinx.com/products/boards-and-kits/device-family/nav-kintex-ultrascale-plus.html .

[18] Xilinx, "Zynq UltraScale+ MPSoC Boards, Kits, and Modules." https://www.xilinx.com/products/boards-and-kits/device-family/nav-zynq-ultrascale-mpsoc.html .

[19] Global, H., "6-Port QSFP28 (6x100G) / QSFP+ (6x40G or 6x56G) FMC+ Module (Vita57.4)." http://www.hitechglobal.com/FMCModules/x6QSFP28.htm .

[20] Global, H., "Z-Ray® Micro Array." http://www.hitechglobal.com/ZRAY/ZRAY-Modules.htm .

[21] Global, H., "CXP (12x10G or 12X12G) Z-RAY Module." https://hitechglobal.us/index.php?route=product/product&path=59_91&product_id=229 .

[22] Global, H., "12Tx / 12Rx FireFly Z-RAY Module." https://hitechglobal.us/index.php?route=product/product&path=59_91&product_id=230 .

[23] Global, H., "6-Port Samtec FireFly (6x100G) FMC+ Module (Vita57.4)." http://www.hitechglobal.com/FMCModules/FMC+FireFly.htm .

[24] Xilinx, "DMA for PCI Express (PCIe) Subsystem." https://www.xilinx.com/products/intellectual-property/pcie-dma.html .

[25] Xilinx, "AXI4-Stream Interconnect." https://www.xilinx.com/products/intellectual-property/axi4-stream_interconnect.html .

[26] Silicon Labs, "Si5341/40 Rev D Data Sheet." https://www.silabs.com/documents/public/data-sheets/Si5341-40-D-DataSheet.pdf .

[27] Silicon Labs, "Si5341, Si5340 Rev D Family Reference Manual." https://www.silabs.com/documents/public/reference-manuals/Si5341-40-D-RM.pdf .

[28] Marin, M. B., Baron, S., Feger, S., Leitao, P., Lupu, E., Soos, C., Vichoudis, P., and Wyllie, K., "The GBT-FPGA core: features and challenges," *Journal of Instrumentation* **10**(03), C03021 (2015).

[29] Hutter, D., de Cuveland, J., and Lindenstruth, V., "CBM First-level Event Selector Input Interface demonstrator," *Journal of Physics: Conference Series* **898**(3), 032047 (2017).