Design of a high throughput electronics module for high energy physics experiments*

Chun-Jie Wang(王春杰)[1;2] Zhen-An Liu(刘振安)[1;1)] Jing-Zhou Zhao(赵京周)[1] Zhao Liu(刘兆)[1;2]

[1] State Key Laboratory of Particle Detection and Electronics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: High-energy physics experiments enable us to explore and understand particle properties and interactions. An increase in luminosity in the accelerator, which allows us to study particles in higher energy ranges, demands faster data transmission and processing. Aimed at this, a high throughput uTCA-compliant electronics module, based on the latest FPGAs, has been designed. It contains 48 10.0 Gb/s optical fiber input channels and 24 10.0 Gb/s optical fiber output channels, supporting up to 480 Gb/s input bandwidth and 240 Gb/s output bandwidth. It complies with the uTCA standards, providing high speed data exchange capability and functioning as a compact and key module in a trigger and DAQ system for a large experiment. A reliable 10.0 Gb/s data transmission among two boards has been verified and one functionality that merges 6 1.6 Gb/s data channels into one single 10.0 Gb/s channel has been achieved. The hardware, firmware and software together with a performance evaluation are given in this paper.

Keywords: high throughput, concentration, trigger, DAQ, uTCA

PACS: 29.85.Ca, 84.30.-r, 07.05.Hd **DOI:** 10.1088/1674-1137/40/6/066102

1 Introduction

Since the announcement of the Higgs-like boson discovery [1, 2] at the LHC (Large Hadron Collider) experiment, CERN, Geneva, Higgs research has become a hot subject. In order to search for and study such high energy particles more deeply, the luminosity of the collider needs to be upgraded to produce more events per second, which then requires improved performance of the trigger and DAQ systems to handle high-speed and huge-volume data. Under these circumstances, the capability of throughout in the trigger and DAQ system becomes a key challenge.

The LHC at CERN has recently been upgraded to provide proton collisions at a center-of-mass energy of at least 13 TeV and eventually close to 14 TeV, with a luminosity greater than $2\times10^{34}\cdot\mathrm{cm}^{-2}\cdot\mathrm{s}^{-1}$. For the detectors, the substantial increase in luminosity comes with huge challenges, such as bandwidth. In the CMS detector [3], the Level-1 trigger output rate is limited to 100 kHz by readout bandwidth. Without a trigger hardware upgrade at the new luminosity, there will be a substantial rise in trigger threshold, which is not physically acceptable, especially in the study of newly discovered Higgs-like boson. In the ATLAS detector, the TDAQ system [4] will allow the ATLAS experiment to efficiently trigger and record data at a higher instantaneous luminosity that is up to three times that of the original LHC design, while the trigger threshold should keep at the same level as the initial run. Obviously, the bandwidth needs to be improved to satisfy these demands at high luminosity [5, 6].

In order to provide solutions for experiments that demand such high-bandwidth data as mentioned above, we propose a novel general purpose high throughput module which could be used not only in LHC experiments but also in future experiments, such as the CEPC (Circular Electron Positron Collider). This high throughput module has strong data IO ability with 480 Gb/s bandwidth in the input direction and 240 Gb/s bandwidth in the output direction, with a line rate of 10.0 Gb/s per channel. The Virtex-7 [7], a recent FPGA from Xilinx, is used to provide large logic and memory resources suitable for sophisticated algorithms. This FPGA allows high throughput and fast data processing. Parallel-optical modules [8] from Avago are used to provide high bandwidth data transmission. In the following sections of this paper, the details will be explained.

Received 28 October 2015, Revised 28 January 2016

∗ Supported by National Natural Science Foundation of China (11435013, 11461141011)

1) E-mail: liuza@ihep.ac.cn

2 Hardware description

Figure 1 shows the hardware block diagram of this electronics module (later refered to as the board), which mainly consists of two FPGAs (Virtex-7 and Kintex-7), two Avago MiniPOD transmitters [8], 4 Avago MiniPOD receivers [8], and one Atmel microprocessor [9].

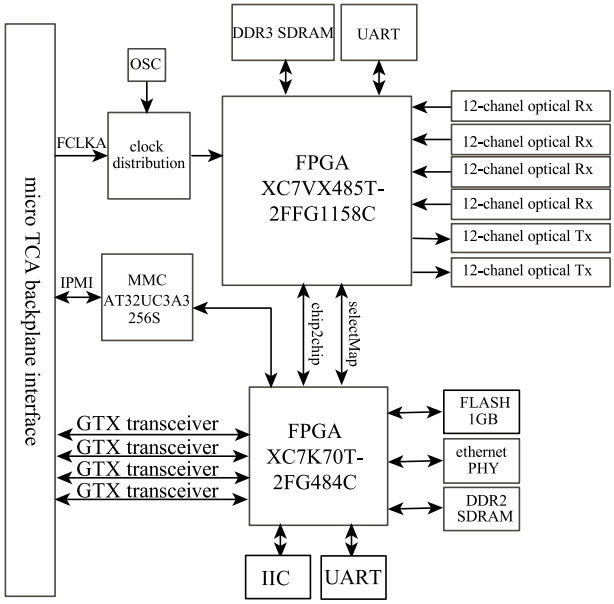

Fig. 1. Block diagram of the high throughput electronics module.

2.1 FPGA and microprocessor

On this board, two Xilinx FPGAs are installed, of which one is a XC7VX415T in FFG1158 package, and the other a XC7K70T in FBG484 package. The former functions as the data processor and is called the core-chip, supporting up to 48-channel GTH transceivers. The latter functions as the board controller and is called the control-chip.

The purpose of using two FPGAs is to separate the processing function and controlling function, which simplifies the management and upgrade of the firmware in the individual chip. One SDRAM is connected to each of the two FPGAs individually for data buffering; one Ethernet interface is provided for external communication, such as with a supervision computer; one flash drive is for the firmware storage, connected to the control-chip.

Besides the FPGAs, there is a micro-controller AT32UC3A512 [9] from Atmel for communication with the MicroTCA [10] backplane aimed at IPMI (Intelligent Platform Management Interface) [11] functionality and configuration of the sensor readout. The AT32UC3A512 chip is a high-performance, low-power 32-bit Atmel AVR Microcontroller, which is based on the AVR32 UC RISC processor running at a frequency up to 66 MHz. It provides on-chip flash and memory for secure and fast access as well as two-wire interface, universal synchronous/asynchronous receiver transceivers (UART), Analog-to-Digital Converter (ADC) and so on. All of these are suitable for board monitoring and control for management through the backplane in the uTCA crate.

2.2 Optical interface

In order to provide high bandwidth in a limited space on the front panel of a double width AMC (Advanced Mezzazine Card) module, 2 MiniPOD transmitters and 4 MiniPOD receivers from Avago, shown in Fig. 2(a), are used as optical interfaces. Each MiniPOD transceiver provides 12 optical links, running up to 10.3125 Gb/s, so a total of 24 transmitting optical links give 240 Gb/s bandwidth, and a total of 48 receiving optical links give 480 Gb/s bandwidth. The AFBR-XXRxy [8] from Avago, which is a twelve-channel, pluggable, parallel-optical transmitter and receiver, is used in our design. These high density optical modules operate over multi-mode fibers using a nominal wavelength of 850 nm. Adapted to MiniPOD as mentioned before, 12-fiber PRIZM [12] is deployed, shown in Fig.2(b), giving the pathway from the MiniPOD to the front panel on the board. The 1.80 mm bare-fiber-ribbon cable mates perpendicularly to the top of the optical modules, providing simple assembly and optimum airflow on the PCB (Printed Circuit Board), as shown in Fig. 2(c).

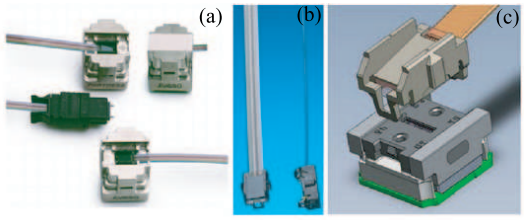

Fig. 2. (color online) (a) MiniPOD transmitter and receiver modules with a flat cable; (b) Prizm optical fiber; (c) MiniPOD with Prizm optical fiber.

Outside the board, ribbon fiber with an MTP connector, as shown in Fig. 3, is used.

Fig. 3. (color online) Ribbon fiber with MTP connector.

2.3 PCB layout and simulation

A data exchange rate of 10.0 Gb/s is an unprecedented challenge in our design experience. In the PCB

layout process, several measures are taken to ensure the signal integrity. First, Rogers 4350b [13] material is chosen instead of the normal FR4 because of the former's high frequency characteristics: its smaller dielectric constant (ξ)of 3.48 ± 0.05 at 10 GHz/23℃and its low dissipation factor (δ) of 0.0037 at 10 GHz/23℃. Its laminates are as easily fabricated into printed circuit board as FR4. A 14-layer stack-up is used to form the PCB in our design as shown in Fig. 4. Second, attention is paid to the via configuration and trace route. The signal lines are routed in signal layers as close to the surface layer as possible to reduce via stub size, as shown in Fig. 5(a). In addition, ground vias are placed next to each signal via of the differential pair to form a G-S-S-G configuration [14], shown in Fig. 5(b). For differential high speed trace routes, the length in the differential pair is tuned equally, and lines are drawn in symmetrical fashion in order to keep the differential pair balanced, surrounded by a ground polygon as shown in Fig. 6.

With all these careful measures in the laying out, a simulation should be made to evaluate the signal quality to keep signal integrity. We used Ansys SIwave 7.0 [15] to extract the PCB module and resize it under the consideration of reducing simulation time, shown in Fig. 7(a). The longest differential pair of transmission lines in the receiving side with DC-blocking capacitors were used as a worst example to analyze the characteristics, with a path with the sequence: top-layer, via, bottom-layer, capacitor, layer3, via and top-layer, as shown in Fig. 7(a). If this pair could satisfy the requirements, so could all the others. The differential S-parameter Sdd21 [16] in Mixed-mode is shown in Fig. 7(b). The S-parameter describes the signal characteristics on the transmission network. At the same time, the S-parameter results of the selected differential pair are saved as the touchstone file of a 4-port network for the link simulation. Finally, a simplex link simulation in the model of IBIS-AMI [17] by ADS12 (Advanced Design System)[18] was launched as shown in Fig. 8 and the result is shown in Fig. 9. The performance of the eye-diagram, as shown in Fig. 9, in the receiver side without the receiver package model, is better than the result in Ref. [17]. The conclusion is therefore drawn that the signal quality can support 10.0 Gb/s data rate.

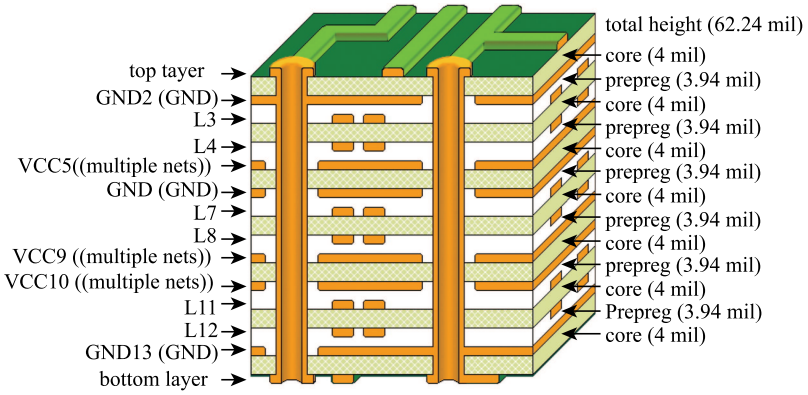

Fig. 4. (color online) Layer stack-up for PCB board. Thickness is measured in thousandths of an inch.

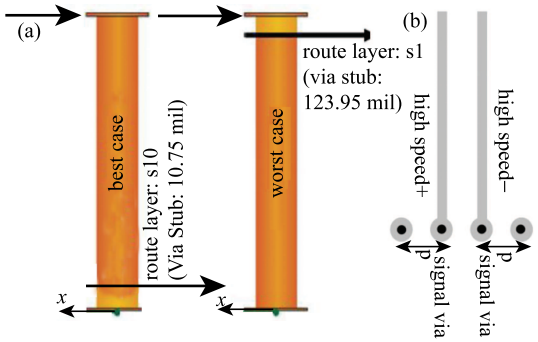

Fig. 5. (color online) (a) Via stubs; (b) G-S-S-G configuration.

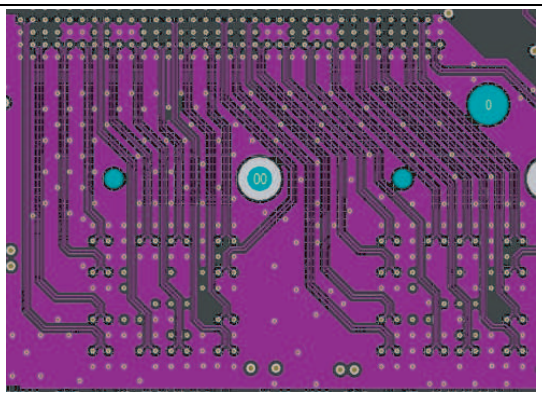

Fig. 6. (color online) Differential pairs in layer-12 in the board.

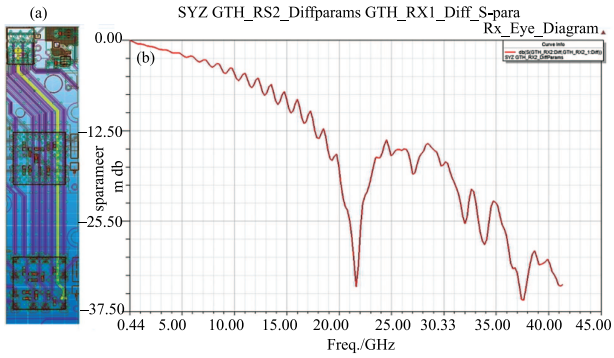

Fig. 7. (color online) (a) Resized PCB and a selected longest differential pair; (b) Sdd21 parameter of the selected pair.

2.4 Clock distribution

In this board, a multi-clock source mode is provided for flexibility. When a common clock coming from the backplane clock path, such as FLKA [11], is used, it is called synchronous mode, because the 12 AMC [11] boards in one crate can use the same clock source. When an on-board oscillator is used for individual clock sources, it is called asynchronous mode, which means the clock on each board is independent.

The reference clocks are distributed as shown in Fig. 10. Both these clock sources go through a jitter cleaner and PLL [19] to generate different frequency clocks, which work as the reference clock for both the system and transceivers. The jitter cleaner and PLL can be programmed through the IIC bus.

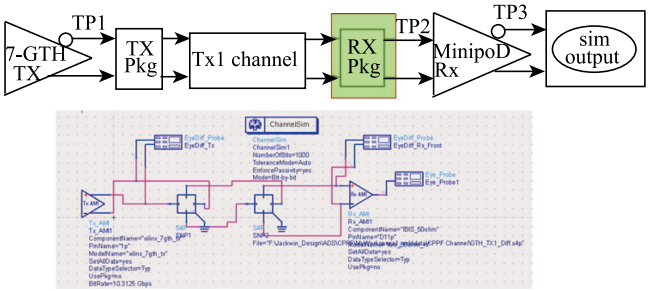

Fig. 8. (color online) Link simulation based on IBIS-AMI model.

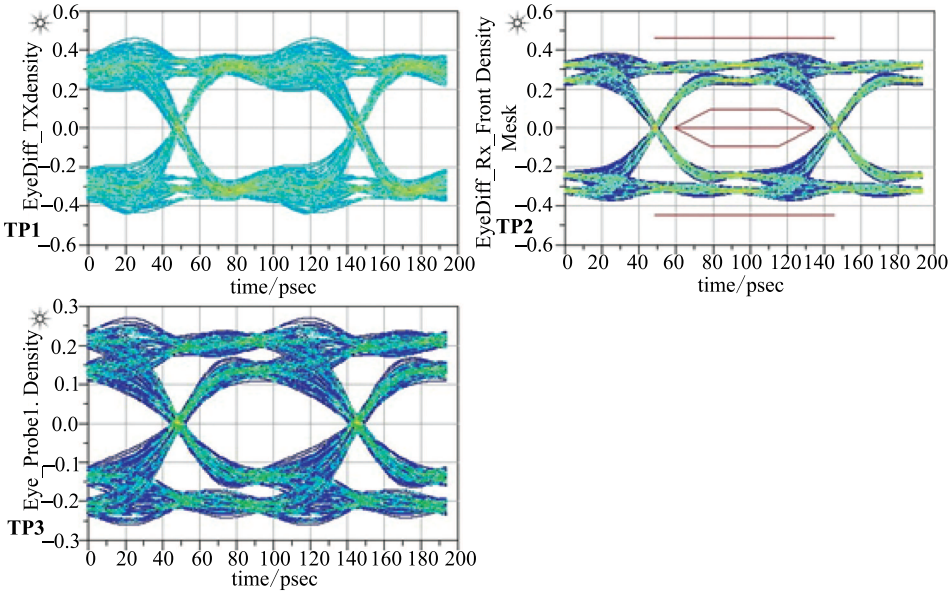

Fig. 9. (color online) Simulation results at 3 testing points.

3 Software design

As in the hardware architecture, the software consists of two parts, one for the board management and control (control-chip) and the other for the application (core-chip).

3.1 Module management and control

In the control-chip, a Microblaze soft processor [20] is implemented to control the whole board. The Microblaze embedded processor soft core is a Reduced Instruction Set Computer (RISC) optimized for implementation in Xilinx FPGA. It provides advanced architecture options like AXI (Advanced Extensible Interface) or PLB (Processor Local Bus) interface, MMU (Memory Management Unit), instruction and data cache, and FPU (Floating-Point Unit). Around the Microblaze processor are peripherals including the following: DDR2 SDRAM, Flash, IIC Interface, universal synchronous/asynchronous receiver transceivers (UART), GPIO, Chip2chip [21], and SelectMap [22] Interface. The software block diagram in the control FPGA is shown in Fig. 11. Specially, Chip2Chip functions like a bridge to seamlessly connect the control-

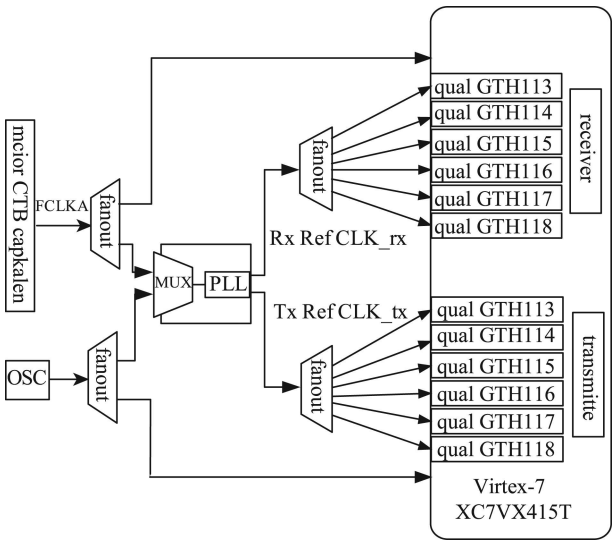

Fig. 10. Clock distribution.

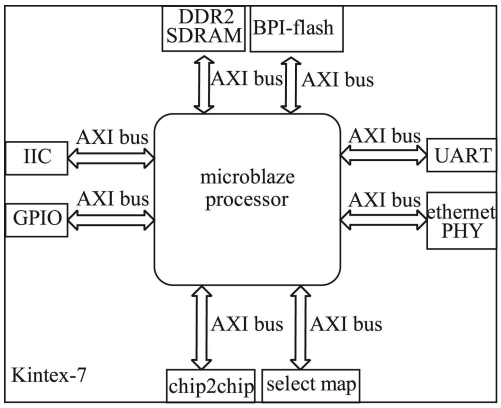

Fig. 11. Simplified software architecture in the control-chip.

chip and core-chip over an AXI interface; SelectMap configuration mode is one of the five configuration modes supported by Xilinx, and behaves in such a way that it uploads the object firmware from flash and downloads it flexibly to the core FPGA.

3.2 Application software

The core-chip is mainly responsible for data receiving, processing and transmitting at high speed. A GTH transceiver in the Virtex-7 FPGA could support up to 13.1 Gb/s data rate, and practically, 10.0 Gb/s is set for output in our design.

As an example application, a concentration algorithm that merges 6 1.6 Gb/s data channels into a single 10.0 Gb/s channel for the CMS Level-1 trigger phase upgrade has been developed and implemented in the core-chip. The purpose of this concentration is to receive low speed data from the radiation tolerant fiber transmitter ASIC of CMS legacy detector electronics and merge them into the new high speed trigger (upgraded) system. For example, the Gigabit Optical Link (GOL) chip [23], with a data transmission rate of 1.6 Gb/s, is used in the CMS legacy muon electronics. After concentration, the data is transmitted at 10.0 Gb/s and processed later for the tracking trigger, for example.

Figure 12 shows the block diagram of the concentration algorithm. The main steps can be listed as: (1) receive and buffer 6 channels of 1.6 Gb/s fiber data in 6 FIFOs individually, with clock domain of 40 MHz; (2) read the data from 6 FIFOs sequentially at the 240 MHz clock in the concentration logic; (3) use a FIFO to bridge the clock domain between 240 MHz and 250 MHz, with the key idea here being to insert an idle code when the FIFO becomes empty. The data width is kept at 32 bits,

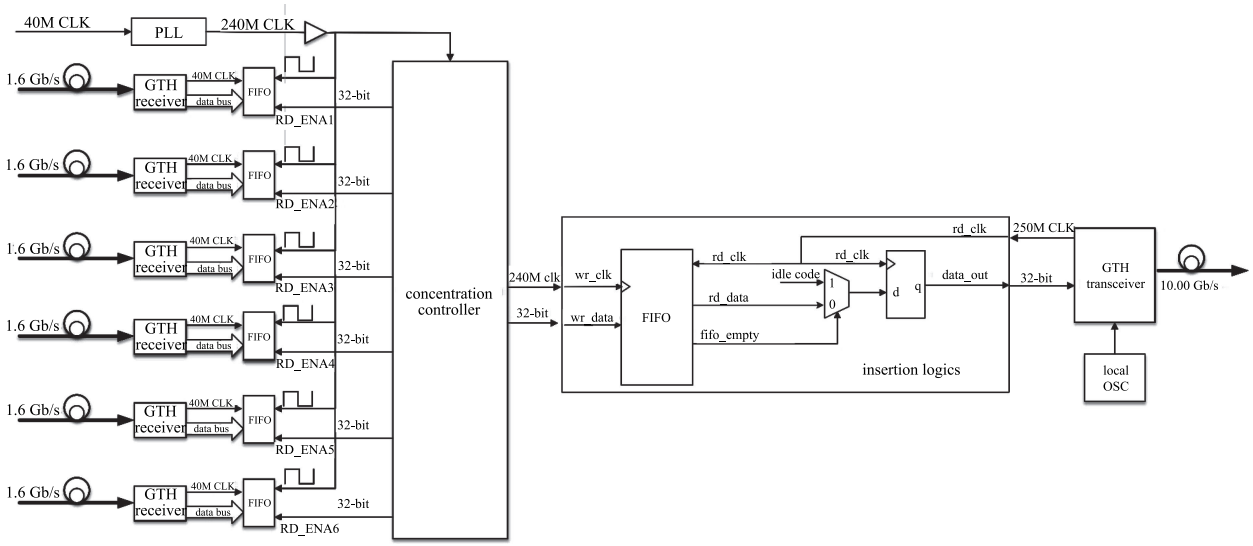

Fig. 12. Diagram of the concentration algorithm.

and after an 8 B/10 B encoder it changes to 40 bits, hence the line rate is 10.0 Gb/s with a 250 MHz clock.

4 Test and verification

A photograph of the board is shown in Fig. 13. For the purpose of functionality testing and performance evaluation, a test demo system has been built to demonstrate the reliability of transmission between two individual boards at 10.0 Gb/s. The reliability of the transmission at 10.0 Gb/s can be expressed in bit-error rate (BER). The BER is the percentage of bits that have errors relative to the total number of bits received in a transmission. To obtain a bit-error rate, two tests were made. First in the loopback test the self-test data are sent out to a GTH transceiver, which then goes through the optical fibers, shown in Fig. 14(a), back to the receiver side on the same board. In the second cross-board test, 24 outputs from one board are connected to 24 inputs on another board with the test data, as shown in Fig. 14(b). The BER test and evaluation are based on ChipScope IBERT [24] from Xilinx.

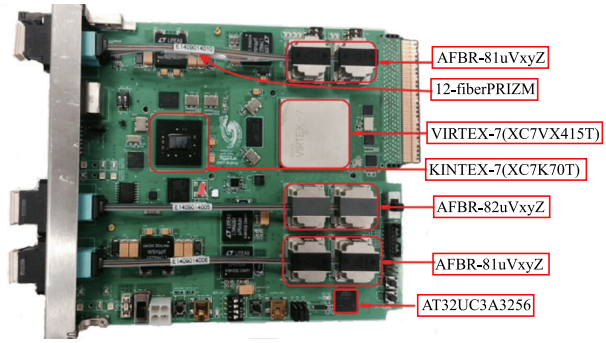

Fig. 13. (color online) A photograph of the processing board.

For the clock, asynchronous mode was chosen. We used an on-board 156.25 M oscillator as the clock source to PLL, and the 156.25 M clock was taken as the reference clock of the GTH transceiver after PLL.

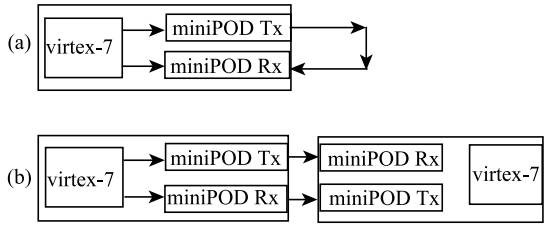

Fig. 14. (color online) (a) Loop-back test in a single board through optical fibers; (b) Cross-board test with two boards connected with optical fibers.

Ahead of core-chip data processing functionality, the basic control-chip tasks should be checked. In the control-chip, PLL configuration, temperature monitoring, voltage and humidity monitoring and firmware downloading to the core-chip through SelectMAP were all implemented and checked to be functioning. All of the software was developed using the Xilinx SDK platform in the C language.

In the single board test, the 10G-BASER protocol was set, of which the line rate is 10.3125 Gb/s. A bit-error rate of 2×10^{-14} was obtained, as shown in Fig. 15. In the cross-board test, no protocol was set with line rate at 10.0 Gb/s, aimed at availability without any protocol guarantee, and the bit-error rate was 9.5×10^{-15}, as shown in Fig. 16. Figure 17 shows the test setup in the lab.

To demonstrate the concentration algorithm as mentioned in Section 3, a testing system in the VT892 uTCA shelf from VadaTech was established as shown in Fig. 18. Two boards were inserted, of which one was used as a data source to send 6 channels of data at 1.6 Gb/s, and the other used to concentrate the received data into 10.0 Gb/s. The firmware was developed and debugged using the Vivado tool from Xilinx, by which the performance can be analyzed in the waveform.

Figure 19(a) shows the data sent from the 6 channels, which are labeled as aa, ab, ac, ad, ae and af respectively. The received data are shown in Fig. 19(b), where the

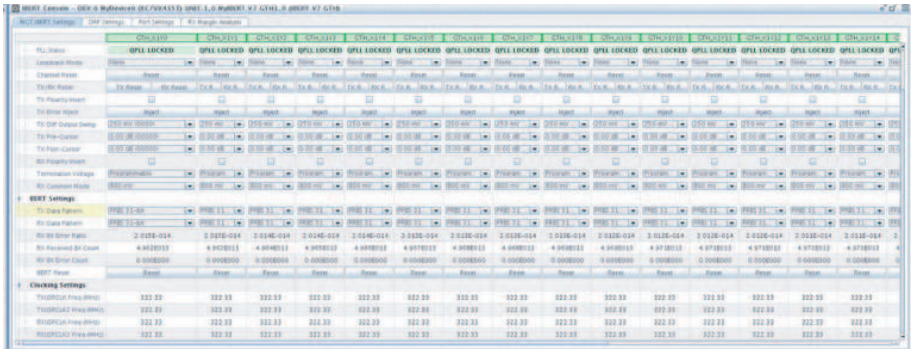

Fig. 15. (color online) Single board loop-back test results. The bit-error rate is 2×10^{-14}. Only 16 channels are shown.

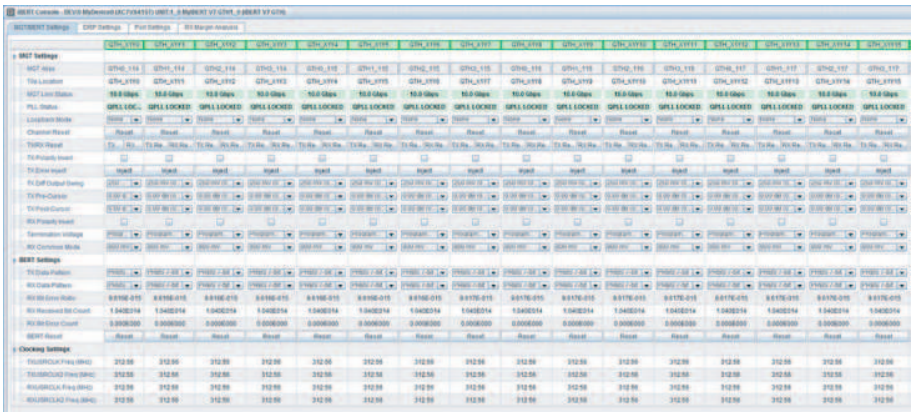

Fig. 16. (color online) Board to board test results. The bit-error rate is 9.5×10^{-15}. Only 16 channels are shown.

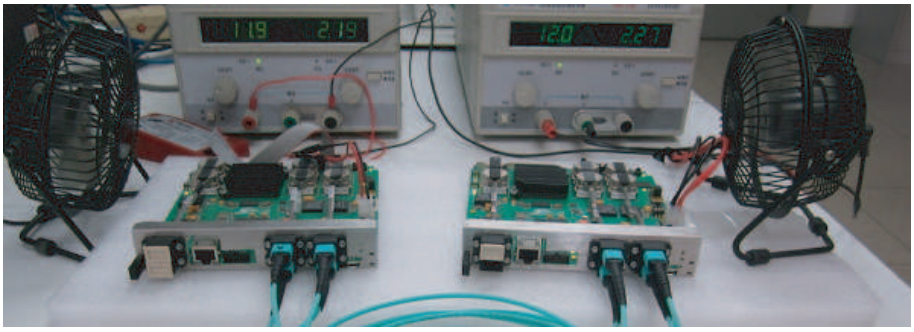

Fig. 17. (color online) Two boards under testing.

Fig. 18. (color online) The testing system in VT892 shelf

byte sequence is different from the sent data because the comma alignment is chosen on even bytes. After being swapped, these data are buffered into a FIFO at the 40 MHz clock and then are read from the FIFO when the clock is in the 240 MHz domain. Figure 20 gives the process of using FIFO empty symbols to switch from the 240 MHz to the 250 MHz domain. Once the FIFO becomes empty, an idle code is inserted into the data stream and this will repeat again and again. Finally, when the 32-bit data under the 250 MHz clock goes to the 8 B/10 B encoder in the transceiver, the line rate is 10.0 Gb/s.

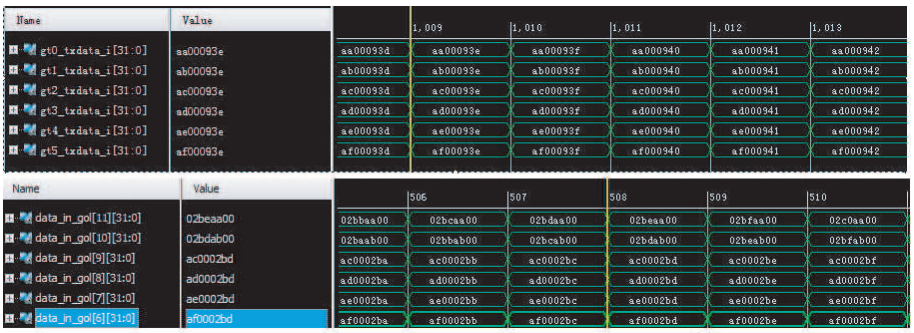

Fig. 19. (color online) (a) 6-channel sent data. (b) The received data without being swapped.

Fig. 20. (color online) The process of switching from 240 MHz clock domain to 250 MHz clock domain.

5 Conclusions and discussion

A high throughput uTCA module based on Virtex-7 with 480 Gb/s input and 240 Gb/s output bandwidth capability has been presented in this paper. Test results show that the board meets the expected requirements and has the ability of data throughput at ultra high speed. It is suitable for particle physics experiments that require high data exchange bandwidth, now and in the future. Further evaluations will be made with experiments such as CMS.

References

1 ATLAS Collaboration, Phys. Lett. B, 716: 1–29 (2012)
2 CMS Collaboration, Phys Lett. B, **716**: 30–61 (2012)
3 Acosta Darin and Ball Austin. CMS Technical Design Report for the Level-1 Trigger Upgrade. CERN-LHCC-2013-011, CMS-TDR-12: 1, 2
4 ATLAS Collaboration, Technial Design Report for the Phase-I Upgrade of the ATLAS TDAQ System. CERN-LHCC-2013-018, ATLAS-TDR-023-2013: 1–2
5 H Xu, ZA Liu et al, An ATCA-based high performance compute node for trigger and data acquisition in large experiments. Physics Procedia **37**: 1849–1854
6 Zhen-An Liu. *Basic idea on Concentrator*, CMS L1 Trigger Meeting (July 9 2013)
7 http://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf, retrieved 1st October 2015
8 http://www.avagotech.com/docs/AV02-2869EN, retrieved 1st October 2015
9 http://www.atmel.com/Images/32058S.pdf, retrieved 1st October 2015
10 PICMG, Micro Telecommunications Computing Architecture Base Specification (July 6, 2006), p.1–8
11 PICMG. Advanced Mezzanine Card Base Specification (November 15, 2006), p.3–1
12 http://www.mouser.cn/pdfdocs/Molex_106267.PDF, retrieved 1st October 2015
13 https://www.rogerscorp.com/documents/726/acm/RO4000-Laminates-Data-sheet.pdf, retrieved 1st October 2015
14 http://www.avagotech.co.jp/docs/AV02-0725EN, retrieved 1st October 2015
15 http://www.ansys.com/Products/Simulation+Technology/Electronics/Signal+Integrity/ANSYS+SIwave, retrieved 1st October 2015
16 David E and Bockelman, IEEE Transactions on Microwave Theory and Techniques, **43**: 7 (1995)
17 http://www.xilinx.com/support/documentation/white_papers/wp424-7Series-GTX-IBIS-AMI-Models.pdf, retrieved 1st October 2015
18 http://cp.literature.agilent.com/litweb/pdf/5991-1516EN.pdf, retrieved 1st October 2015
19 https://www.silabs.com/Support%20Documents/TechnicalDocs/Si5326.pdf, retrieved 1st October 2015
20 http://www.xilinx.com/support/documentation/sw_manuals/xilinx2014_2/ug984-vivado-microblaze-ref.pdf, retrieved 1st October 2015
21 http://www.xilinx.com/support/documentation/ip_documentation/axi_chip2chip/v4_2/pg067-axi-chip2chip.pdf, retrieved 1st October 2015
22 http://www.xilinx.com/support/documentation/application_notes/xapp502.pdf, retrieved 1st October 2015
23 P. Moreira and T. Toifl, Gigabit Optical Link Transmitter Maual, Version 1.9 (2005)
24 http://www.xilinx.com/support/documentation/sw_manuals/xilinx13_3/ug811_ChipScopeUsingIBERTwithAnalyzer.pdf, retrieved 1st October 2015