GBT link testing and performance measurement on PCIe40 and AMC40 custom design FPGA boards

2016 JINST 11 C03039

(http://iopscience.iop.org/1748-0221/11/03/C03039)

Jinst

Published by IOP Publishing for Sissa Medialab

Received: *November 15, 2015*
Revised: *January 14, 2016*
Accepted: *January 25, 2016*
Published: *March 17, 2016*

Topical Workshop on Electronics for Particle Physics 2015,
September 28th – October 2nd, 2015
Lisbon, Portugal

GBT link testing and performance measurement on PCIe40 and AMC40 custom design FPGA boards

Jubin Mitra,[a,1] **Shuaib A. Khan,**[a] **Manoel Barros Marin,**[b] **Jean-Pierre Cachemiche,**[c] **Erno David,**[d] **Frédéric Hachon,**[c] **Frédéric Rethore,**[c] **Tivadar Kiss,**[d] **Sophie Baron,**[b] **Alex Kluge**[b] **and Tapan K. Nayak**[a]

[a] *VECC, Kolkata, India*
[b] *CERN, Geneva, Switzerland*
[c] *CPPM, Marseille, France*
[d] *Wigner RCP, Budapest, Hungary*

E-mail: jubin.mitra@cern.ch

Abstract: The high-energy physics experiments at the CERN's Large Hadron Collider (LHC) are preparing for Run3, which is foreseen to start in the year 2021. Data from the high radiation environment of the detector front-end electronics are transported to the data processing units, located in low radiation zones through GBT (Gigabit transceiver) links. The present work discusses the GBT link performance study carried out on custom FPGA boards, clock calibration logic and its implementation in new Arria 10 FPGA.

Keywords: VLSI circuits; Trigger concepts and systems (hardware and software); Optical detector readout concepts; Detector control systems (detector and experiment monitoring and slow-control systems, architecture, hardware, algorithms, databases)

[1] Corresponding author.

doi:10.1088/1748-0221/11/03/C03039

Contents

1 Introduction

In High Energy Physics experiment one of the major challenges is to transfer data with very high reliability between the different sub-detectors situated in the harsh radiation zone to the data acquisition electronics located in the non-radiation area. The S-LINK (Simple Link Interface) [1] specification defined in 1995 at CERN, describes a data link for moving data or control word between front-end electronics to the read-out electronics, via GOL (Gigabit Optical Link) serializer chip. The sub-detectors at the same time must able to receive trigger and timing information maintaining a constant latency. A unified approach developed by RD12 for the broadcasting of TTC (Timing, Trigger and Control) [2] signals from the RF generators of the LHC machine to the outputs of the timing receiver ASICs (TTCrx + QPLL) at the experiment and beam instrumentation destinations. Since there wasn't one single chip doing both the functionalities, so the CERN team decided to build one single chip combining both. This initiated the GBT (Gigabit Transceiver) [3] and the versatile link [4] project. GBT is a radiation tolerant error resilient data communication standard with fixed latency support. A single GBT channel link can handle detector data, timing, trigger and control information traffic. Radiation tolerant GBT chipset used for packaging of detector data and transmitting it in GBT standard. While the optoelectronic components and point to point optical link connecting the GBT ASIC with the FPGA (Field Programmable Gate Array) / COTS

(Commercial Off-The-Shelf) is qualified by versatile link project group. A short summary of the GBT protocol is tabulated in table 1 and figure 1 gives the details of GBT frame.

Current work describes *GBT FPGA Core* firmware [5, 6] implementation on FPGA (reconfigurable electronics hardware), inter-FPGA firmware migration, clock calibration strategy and comparative performance study. Two custom DAQ boards AMC40 [7] and PCIe40 [8, 9] based on latest Altera Stratix V [10] and Arria 10 [11] FPGA were used for the detailed comparative study of GBT FPGA Core firmware implementation.

Table 1. Details of GBT protocol standards.

Parameters	**GBT Data Transmission Protocol** [3, 5, 6, 12]
Security	Not Applicable
Channel Data Throughput	4.8 Gbps
Raw data Throughput	3.2 Gbps
Bandwidth Utilization (Data Efficiency)	66.67% (80 / 120)
Bandwidth Utilization (Coding Efficiency)	73.33% (88 / 120)
Forward Error Correction	RS Encoding (15,11) with symbol size of 4 bits
Number of FEC block used	2
Error Detection	32 bits
Error Correction	16 bits
Burst Error Correction	16 bits
Latency optimized data transmission	Supported
Interleaver	Block
Supported Data Type	Idle/Control and Data

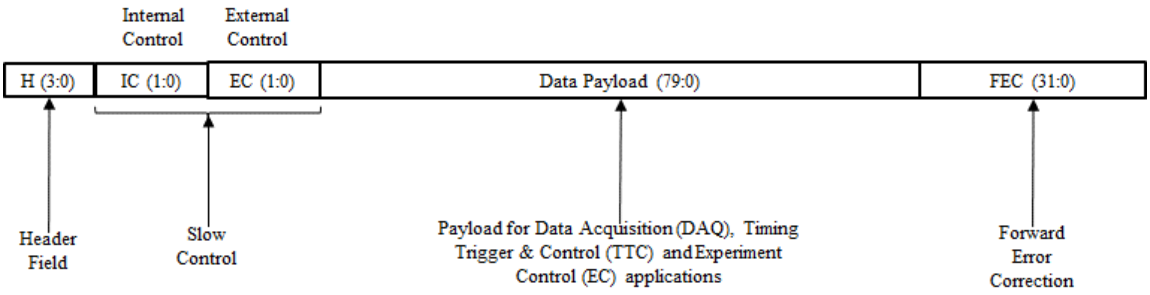


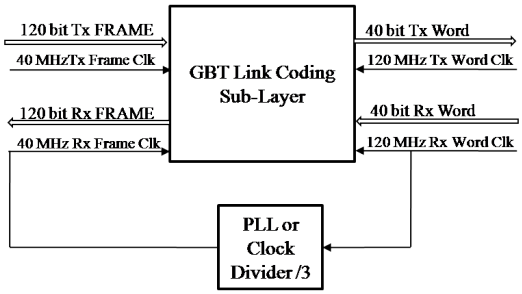
Figure 1. Data format for GBT protocol specific data packet.

2 Design implementation

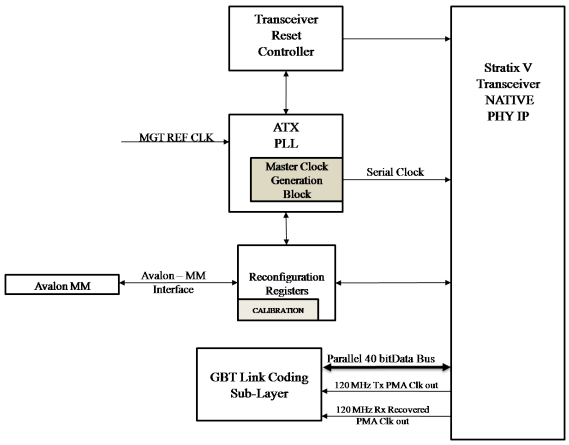
AMC40 and PCIe40 are based on Altera Stratix V and Arria 10 FPGA respectively. GBT FPGA core is composed of two sections: (1) GBT Coding Sub-layer; (2) MGT (Multi-Gigabit Transceiver). GBT coding sub-layer is independent of FPGA architecture chosen. While, MGT (Multi-Gigabit Transceiver) is FPGA dependant. Figures 2a, 2b and 2c shows for GBT firmware implementation

2016 JINST 11 C03039

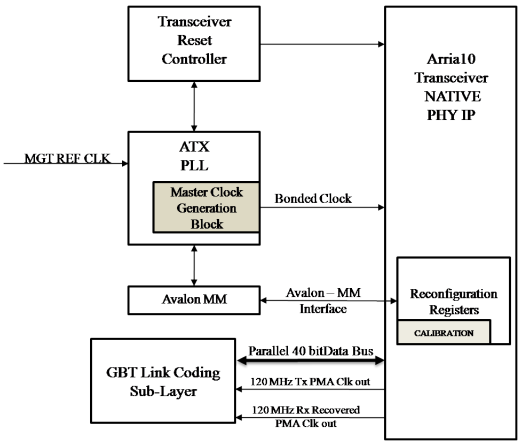
only transceiver connectivity need to be modified based on device migration guideline, rest of the core coding logic can be kept intact.



(a) Basic signal block for GBT Coding Sub-layer



(b) Stratix V Transceiver Design Connectivity



(c) Arria 10 Transceiver Design Connectivity

Figure 2. GBT core design connectivity.

Both the designs use **PMA** (Physical Media Attachment) **bonded mode** [10, 11] configuration, to get same latency through each bonded data-path. Accordingly in Stratix V and Arria 10, GBT PMA links are grouped into 3 and 6 links respectively to form a single **GBT Bank**.

3 Resource estimation

For resource estimation, a standard reference design from CERN GBT team is used having same configurations on both boards. The reference design altogether contains 4 links operating in two different modes as shown in table 2. Comparison of resource utilization for GBT reference design implemented in AMC40 vs PCIe40 boards are shown in table 3.

4 Latency measurement

GBT has got two operational modes namely; Standard mode that uses an elastic buffer or FIFO for Clock Domain Crossing (CDC) and Latency Optimized mode which uses an inelastic buffer or register bank for CDC. Latency measurement is very critical for GBT. In GBT, information content

2016 JINST 11 C03039

Table 2. Reference Design configuration.

Configuration Parameter	Value
No of GBT Bank	2
Bank 1	No of Links = 1 Tx Mode = STANDARD Rx Mode = LATENCY OPTIMIZED
Bank 2	No of Link*s* = 3 Tx Mode = LATENCY OPTIMIZED Rx Mode = STANDARD

Table 3. Shows the comparison of resource utilization in two boards for Standard Reference Design.

Parameters	**AMC40**		**PCIe40**	
FPGA Family	**Stratix V**		**Arria 10**	
Device Part Number	**5SGXEA7N2F45C3**		**10AX115S4F45I3SGE2**	
	Occupancy Ratio	Percentage of Occupancy	Occupancy Ratio	Percentage of Occupancy
Logic utilization (in ALMs)	10,542 / 234,720	4%	9,055 / 427,200	2 %
Total registers	20060		18338	
Total block memory bits	202,752 / 52,428,800	< 1 %	126,464 / 55,562,240	< 1 %
Total RAM Blocks	56 / 2,560	2 %	40 / 2,713	1 %
Total HSSI PMA TX/RX (Serializers/Deserializers)	4 / 48	8 %	4 / 72	6 %
Total PLLs	11 / 92	12 %	10 / 176	6 %

can be data, timing, and control format, each comes with its latency boundary condition. Most stringent latency boundary conditions apply for Timing and Trigger distribution.

Latency. Latency occurs in both transmission and receiving directions, and is different in each direction, depending on the media and path involved. Latency gives information of the logic path delay, and whether the path contains an elastic or an inelastic buffer. It directly contributes to the calculation of round trip delay. In our observation, we have used round trip delay to get a quick estimation of aggregate latency.

Roundtrip delay. The round trip delay corresponds to the length of time it takes for a signal to be sent plus the length of time it takes for an acknowledgment of that signal to be received. It includes serialization, deserialization time along with propagation delay.

The round trip delay are formulated in these following given equations for easy understanding of the path delay or logic delay (D) involved.

1. Serial Loopback or Electrical Loopback

$$D_{\text{Serial Loopback Round-Trip}} = (D_{\text{GBT Tx Encoder}} + D_{\text{Serialization}} + D_{\text{Deserialization}} + D_{\text{GBT Rx Decoder}})^{\text{FPGA}}$$

2. Optical Loopback

$$D_{\text{Optical Loopback Round-Trip}} = D_{\text{Serial Loopback Round-Trip}} + D_{\text{Optical Fibre Length}}$$

2016 JINST 11 C03039

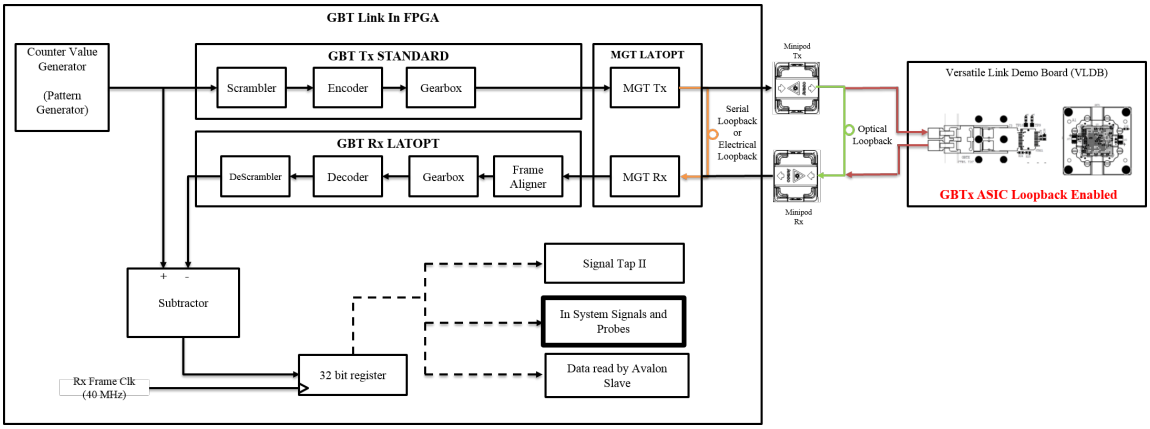


Figure 3. Test Setup for 3 types of round trip measurement, each loopback arrangement are marked by different colours.

3. GBTx ASIC Loopback

$$D_{\text{GBTx ASIC Loopback Round-Trip}} = D_{\text{Optical Loopback Round-Trip}} + (D_{\text{GBT Tx Encoder}} + D_{\text{Serialization}} + D_{\text{Deserialization}} + D_{\text{GBT Rx Decoder}})^{\text{ASIC}}$$

The round-trip delay can be considered as round-trip latency ($\mathscr{L}$), provided there are no hold ups in the path, like elastic buffer, queuing, congestion control. Such conditions exists only under special constraints. GBT ASIC by default operates in latency optimized mode on both Tx and Rx side, while GBT FPGA core need to be forced to operate in such mode.

$$\mathscr{L}_{\text{GBTx ASIC Round-Trip}} = D_{\text{GBTx ASIC Loopback Round-Trip}} - D_{\text{Optical Loopback Round-Trip}} \quad (4.1)$$

$$\mathscr{L}_{\text{GBT FPGA Core Round-Trip}} = D_{\text{Serial Loopback Round-Trip}} : \text{when Tx and Rx latency optimized} \quad (4.2)$$

$$\Delta\mathscr{L} = (\mathscr{L}_{\text{GBT FPGA Core Round-Trip}} - \mathscr{L}_{\text{GBTx ASIC Round-Trip}}) \quad (4.3)$$

By design GBT ASIC takes lowest round-trip latency. So, lower the $\Delta\mathscr{L}$ better the GBT FPGA Core performance.

Resolution. Measurement of latency within FPGA using firmware is always dependent on the data rate, and at which step of data parallelism it is sampled. In our case, it is sampled at the beginning of sending data frame (120 bit)@40 MHz. So, the resolution is 25 ns, which is the rate of LHC bunch crossing.

From the above table we can infer:

$$\mathscr{L}_{\text{GBTx ASIC Round-Trip}} = 4$$
$$\mathscr{L}_{\text{GBT FPGA Core Round-Trip}} = 6$$
$$\Delta\mathscr{L} = 2$$

5 Bit Error Rate (BER) analysis

The Transceiver Toolkit (TTK), an on-chip debugging tool provided by Altera for BER monitoring is used for this analysis. The test values are listed in table 6. TTK full functionality is not available for Arria 10 ES1. So, different test setup is used for two boards. The, results for PCIe40 1st version DAQ engine are preliminary.

2016 JINST 11 C03039

Table 4. Round Trip delay meausurements for multi-level loopback in PCIe40.

GBT Protocol		**ROUND TRIP DELAY** (In terms of LHC bunch crossing Rate = 25 ns)		
Transmission Side	**Receiver Side**	**Serial or Electrical Loopback**	**Optical Loopback**	**GBT ASIC Loopback**
Latency Optimized	Latency Optimized	6	7	11
Latency Optimized	Standard	13	14	18
Standard	Latency Optimized	8	9	13
Standard	Standard	18	18	22

Table 5. GBT Serial Loopback Round Trip Measurement comparison for AMC40 and PCIe40.

	Different Mode of operations			
Tx Side →	**Standard**	**Standard**	**Latency Optimized**	**Latency Optimized**
Rx Side →	**Standard**	**Latency Optimized**	**Standard**	**Latency Optimized**
AMC40	600 ns	350 ns	350 ns	150 ns
PCIe40	450 ns	350 ns	200 ns	150 ns

Table 6. BER Measurement.

Parameters	AMC40	PCIe40
Test Setup	AMC40 (Tx) - optical cable - AMC40 (Rx)	PCIe40 (Tx) - optical cable - AMC40 (Rx)
Tx configuration Parameter		
V_{OD} control	50	29
Pre-emphasis 1st post-tap	0	0
Pre-emphasis pre-tap	0	0
Pre-emphasis 2nd post-tap	0	0
Rx configuration Parameter		
DC Gain	1	1
Equalization Control factor	0	0
DFE Enabled	off	off
PLL reference Clock Frequency	120 MHz	120 MHz
Test Pattern	PRBS31	PRBS31
Results of TTK		
BER (Bit Error Rate)	1.115E-12	5.0723E-10 #
Eye Width/Eye Height	26/127	23/89

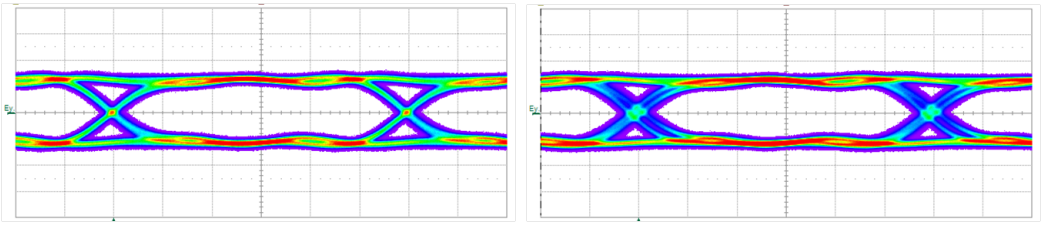
This result is preliminary, will subject to change from new versions

6 Jitter and eye measurement

Data from the FPGA transceivers are transmitted to the onboard MiniPOD^{TM} [13] optical modules. They are coupled with flat ribbon cable using a Prizm connector and terminated on the other side with industry standard MTP connector. A Lecroy Serial Data Analyser (SDA) oscilloscope is used for analyzing MiniPOD^{TM} signal quality.

2016 JINST 11 C03039

An eye diagram is a common indicator of the quality of signals in high-speed digital transmissions. The signal to noise ratio of this high-speed data signal is directly indicated by the amount of eye closure or Eye Height.



(*a*) (*b*)

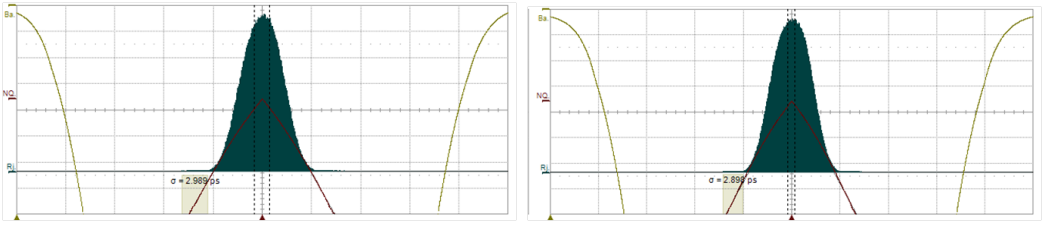
Figure 4. Showing the Eye Diagram for Tx Optical using GBT encoded data (a) AMC40 and (b) PCIe40.

Table 7. Shows the comparison of Optical Minipod^{TM} performance in two boards.

	AMC40	PCIe40
Eye Height	378 mV	373 mV
Eye Width	190.4 ps	176.8 ps

	AMC40	PCIe40
Deterministic Jitter (Dj)	5.503 ps	13.125 ps
Periodic Jitter (Pj)	6.75 ps	7.24 ps
Data Dependant Jitter (DDj)	11.228 ps	21.647 ps
Inter Symbol Interference (ISI)	11.095 ps	21.657 ps
Duty Cycle Jitter (DCD)	2.000 ps	1.912 ps
Random Jitter(Rj)	3.245 ps	3.204 ps
Total Jitter (Tj)	51.148 ps	58.185 ps
Standard Deviation(σ)	2.989 ps	2.898 ps

Jitter spectrum is the measurement of timing variation of a signal edge from its ideal values. Contributing factors include power load variation, thermal noise, line attenuation and interference coupled from nearby devices. It is very important that the clock jitter is within a tolerable hold time of the pipeline registers for error free data transmission.



(*a*) (*b*)

Figure 5. Showing the Jitter Spectrum for (a) AMC40 and (b) PCIe40.

7 Calibration

High-Speed Serializer/Deserializer (SerDes) or Multi-Gigabit Transceivers (MGT) as well as internal PLLs include both analog and digital blocks that require calibration to compensate for process,

2016 JINST 11 C03039

voltage, and temperature (PVT) variations [11]. MGT in fixed-latency mode do not have the circuitry to maintain same latency for data path with each power or reset cycle [14]. In time synchronization protocol, like *GBT in Latency optimized mode*, **Automated Transceiver Calibration** (ATC) must be followed by **Tx Latency Calibration** (TLC) and **Rx Latency Calibration** (RLC) during both power and reset cycle. In the case of incomplete or improper calibration, ATC contributes to random jitter while TLC and RLC are responsible for deterministic jitter. The latency error measured in Unit Intervals (UI) which corresponds to one pulse duration of GBT data stream (1 UI = 208.33ps@4.8Gbps).

7.1 Tx Latency Calibration (TLC)

The fabric-transceiver clock interface in GBT design consists of signals: input reference clocks (refclk) and transceiver data path interface clocks (tx_clkout, rx_clkout) [10, 11]. Transceiver forwards *tx_clkout* to be is used in FPGA fabric as *tx parallel word clock*. This clock is used to drive the user logic data into the transmitter DDR (Double Data Rate) PISO (Parallel Input Serial Output). This *tx_clkout* (120 Mhz) is obtained from the serial clock (2400 MHz or 2 UIs) by dividing it with serialization factor of 20. However, this clock divider used to obtain *tx parallel word clock* (tx_clkout) introduces a phase difference uncertainty ($\delta\phi_p$) with *external parallel word clock* (refclk). Equation (7.1) shows $\delta\phi_p$ can take any value within a set of 20 elements in $\Delta\Phi_p$.

$$\Delta\Phi_p = \{2n \text{ UI} \ : \ n \in \mathbb{I},\ 0 \leq n \leq 19)\}, \text{where 1 UI} = 208.33\text{ps@4.8 Gbps} \tag{7.1}$$

$$\text{Max}[\delta\phi_p] = 38 \text{ UI} = 7.916 \text{ ns} \approx 8 \text{ ns} \tag{7.2}$$

$$\text{Phase Difference Error or Latency Error} = \pm\text{Max}[\delta\phi_p]/2 = \pm 4 \text{ ns} \tag{7.3}$$

This phase difference error goes completely undetected by GBT PCS (Program Coding Sublayer) unless it falls in metastability region and flags data frame error. The phase (ϕ_p) transition from one value to other with each reset cycle is 2 UIs, which is sequential, predictable and has exhaustive number of states. We have used this property to develop the proposed calibration solution to get always the same value for $\delta\phi_p$ and therefore get a reproducible fixed latency or just find a value that avoids metastability.

7.2 Proposed solution

Forwarded Tx Word Clk takes some initialization time after each reset cycle, so for proper alignment between the frame and the word in (1:3) Gearbox, the gearbox sync reset is delayed by few clock cycles from the global reset. Synchronization register chain or synchronizer is used to minimize the failures due to metastability [15]. The calibration trigger can be based on metastability detection (Logic 1) or phase error detection (Logic 2). If the requirement of the design allows a latency error tolerance margin of ±4 ns then metastability detection logic is suitable because of its lower resource usage. Otherwise, the phase error detection logic is more suitable for better phase predictability design. The entire solution is implemented by modification in the latency optimized gearbox in GBT PCS.

7.2.1 Phase error detection logic

Tx FrameClk has a synchronous phase relationship with reference clock (refclk). So, the phase calculation is computed between the Tx word and Tx frame clock. This digital calculation of phase

2016 JINST 11 C03039

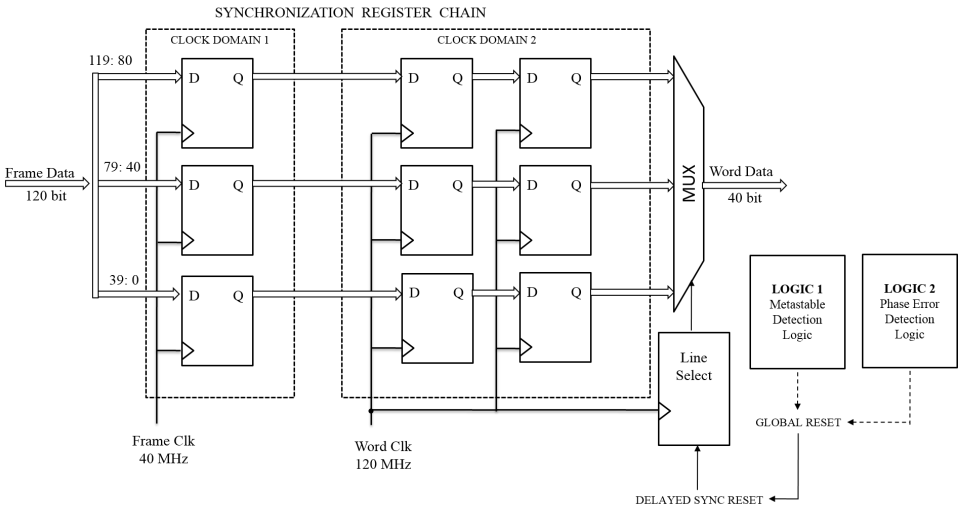


Figure 6. Synchronization register based clock domain crossing with metastability or phase error detection logic.

difference within FPGA fabric is not very accurate, yet sufficient to detect a $\delta\phi_p$ phase difference. A look-up table is prepared manually, and a particular phase value is pre-defined with a safe margin of the metastable zone. Figure 7 shows the detected phase difference is compared against phase calibration table values to assert global reset.

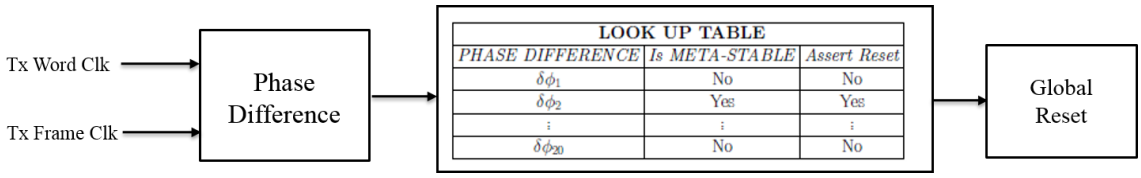


Figure 7. Calibration trigger from manually tabulated phase error look-up table.

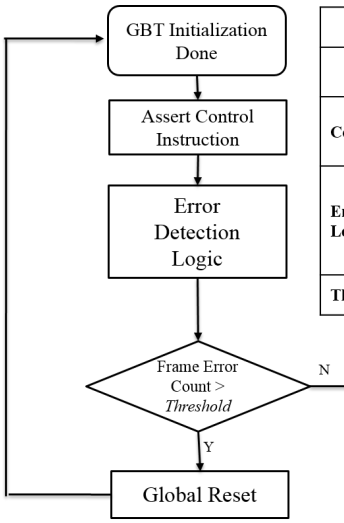
Phase selection for temperature variation tolerance. Proper phase synchronization between frame clock and word clock is necessary to avoid metastability during data transmission. However, with temperature variation phase drift happens. If the synchronized region lies close to the metastable region, phase drift due to temperature causes it to get unstable. So, the calibrated phase should be chosen wisely. We have varied the temperature by controlling the on-board cooling system and noted for any data frame errors. Temperature Sensing Diode (TSD) internal to FPGA is used for temperature reading.

7.2.2 Metastability detection logic

To ensure data write reliability the input to a register must be stable for a minimum time before the clock edge (register setup time) and for data read reliability, the output of a register must be stable a minimum time after the clock edge (register hold time) [15]. For proper data frame to word conversion, it must avoid metastable regions. However, due to the phase transition of forwarded

2016 JINST 11 C03039

tx_clkout, it sometimes falls into this zone and causes data error. Figure 8 shows with flowchart 3 types of data error detection logic.



Parameters	Solutions		
	Pattern check	***RS Syndrome check***	***Tx Side Header Lock***
Control Instructions	•Serial Loopback •Tx→ Pattern Generator •Rx→ Pattern Checker	•Serial Loopback	N/A
Error Detection Logic	Input / Output pattern equality	Syndrome Check of Reed Solomon Forward Error Correction in Rx Side	•Loss of Header Lock
Threshold	1 Frame Error	1000 Frame Errors	100 Frame Errors

Figure 8. Calibration trigger on data frame error or metastability.

7.3 Stratix V and Arria 10 FPGA

A user recalibration of the FPGA is required after every power cycle showing a loss of lock of the MGT reference clock during Automatic Transceiver Calibration (ATC) [10, 11]. Arria 10 FPGA uses hardened Precision Signal Integrity Calibration Engine (PreSICE) to perform calibration routines at every power-up before entering user-mode. After transceiver calibration is over, latency calibration logic is followed. Tx Latency Calibration (TLC) is done as discussed in section 7.1. The Arria 10 device used in this test procedure can work without TLC. Rx Latency Calibration (RLC) uses a new type of barrel shift logic that does bit slip scanning to lock Frame Alignment Word (FAW) at first hit[3]. The maximum number of bits slipped is equal to the FPGA fabric-to-transceiver interface width minus 1, i.e. 19 (= 20 – 1). After completing the calibration sequences, the control is transferred to user logic.

8 Conclusions

With this comparative analysis, we have proved that the GBT protocol can indeed be implemented with success in both the 28nm Stratix-V and 20 nm Arria-10 Altera FPGAs. The source code of developed firmware are archived in CERN espace, and more resources are available to users on requests.

References

[1] O. Boyle, R. McLaren and E. van der Bij, *The S-LINK interface specification*, *ECP division CERN* (1997).

[2] B. Taylor, *Timing distribution at the lhc*, *CERN European Organization for Nuclear Research-Reports-Cern* **3** (2002) 63.

2016 JINST 11 C03039

[3] P. Moreira et al., *The GBT project*, CERN-2009-006.342.

[4] F. Vasey et al., *The versatile link common project: feasibility report*, 2012 *JINST* **07** C01075.

[5] S. Baron, J. Cachemiche, F. Marin, P. Moreira and C. Soos, *Implementing the GBT data transmission protocol in FPGAs*, in proceedings of the *Topical Workshop on Electronics for Particle Physics (TWEPP-09)*, Paris, France, 21–25 Sep. 2009, pp. 631–635, CERN-2009-06.

[6] M. Barros Marin et al., *The GBT-FPGA core: features and challenges*, 2015 *JINST* **10** C03021.

[7] J.P. Cachemiche, P.Y. Duval, F. Hachon, R. Le Gac and F. Rethore, *Recent developments for the upgrade of the LHCb readout system*, 2013 *JINST* **8** C02014.

[8] M. Bellato et al., *A PCIe Gen3 based readout for the LHCb upgrade*, *J. Phys. Conf. Ser.* **513** (2014) 012023.

[9] F. Alessio et al., *LHCb: Clock and timing distribution in the LHCb upgraded detector and readout system*, CERN-Poster-2014-461, presented at *Topical Workshop on Electronics for Particle Physics 2014*, Aix En Provence, France, 22–26 Sep. 2014,

[10] Altera Corporation, *Transceiver Clocking in Stratix V Devices*, in *Statix V Device Handbook* **3** (2012).

[11] Altera Corporation, *Arria 10 Transceiver PHY User Guide*, in *Arria 10 Device Handbook* (2015).

[12] A. Caratelli et al., *The GBT-SCA, a radiation tolerant ASIC for detector control and monitoring applications in HEP experiments*, 2015 *JINST* **10** C03034.

[13] D. Vaughan, R. Hannah and M. Fields, *Applications for Embedded Optic Modules in Data Communications*, in *Avago Technologies White Paper* (2011).

[14] R. Giordano and A. Aloisio, *Fixed-latency, multi-gigabit serial links with Xilinx FPGAs*, *IEEE T. Nucl. Sci.* **58** (2011) 194.

[15] J. Stephenson, D. Chen, R. Fung and J. Chromczak, *Understanding metastability in FPGAs*, in *Altera Corporation white paper* (2009).

2016 JINST 11 C03039