Performance and lessons of the CMS Global Calorimeter Trigger

G. Iles [a], J. Brooke [b], C. Foudas [a], R. Frazier [b], G. Heath [b], M. Hansen [c], J. Jones [d], J. Marrouche [a], A. Rose [a], G. Sidiropoulos [a], M. Stettler [c], A. Tapper [a]

[a] Blackett Laboratory, Imperial College London, UK
[b] Bristol University, Bristol, UK
[c] CERN, 1211 Geneva 23, Switzerland
[d] Princeton University, New Jersey, USA

g.iles@imperial.ac.uk

Abstract

The CMS Global Calorimeter Trigger (GCT) has been designed, manufactured and commissioned on a short time schedule of approximately two years. The GCT system has gone through extensive testing on the bench and in-situ and its performance is well understood. This paper describes problems encountered during the project, the solutions to them and possible lessons for future designs, particularly for high speed serial links. The input links have been upgraded from 1.6Gb/s synchronous links to 2.0Gb/s asynchronous links. The existing output links to the Global Trigger (GT) are being replaced. The design for a low latency, high speed serial interface between the GCT and GT, based upon a Xilinx Virtex 5 FPGA is presented.

I. Introduction

This paper is devoted to the challenges faced and lessons learnt during the development and commissioning of the GCT system and refers to the architecture of the design and the implementation of high speed serial links. Both are likely to be used in future systems and are of value to the larger LHC trigger community.

A detailed description of the GCT is beyond the scope of this paper and is covered in detail in the CMS Trigger TDR [1] and several subsequent CMS internal notes and conference proceedings [2,3].

The main challenge with the GCT and with most trigger systems is the high bandwidth requirements coupled with the fact that data often needs to be shared or duplicated, and done so with low latency. The GCT uses a mixture of high speed serial links and wide parallel busses. The high speed serial links are necessary to concentrate the data into a single FPGA, thus reducing data sharing requirements and making the processing efficient. The latency cost of these links is not negligible and thus wide parallel busses operating conservatively at 80MHz are used for the rest of the system.

The GCT is modular, which allowed multiple design teams to work in parallel in the initial stages of the product. It also simplified each board, thus reducing the layout and design time. It allowed the GCT-to-GT links to be replaced without requiring complex changes to the main 9U VME data processing card.

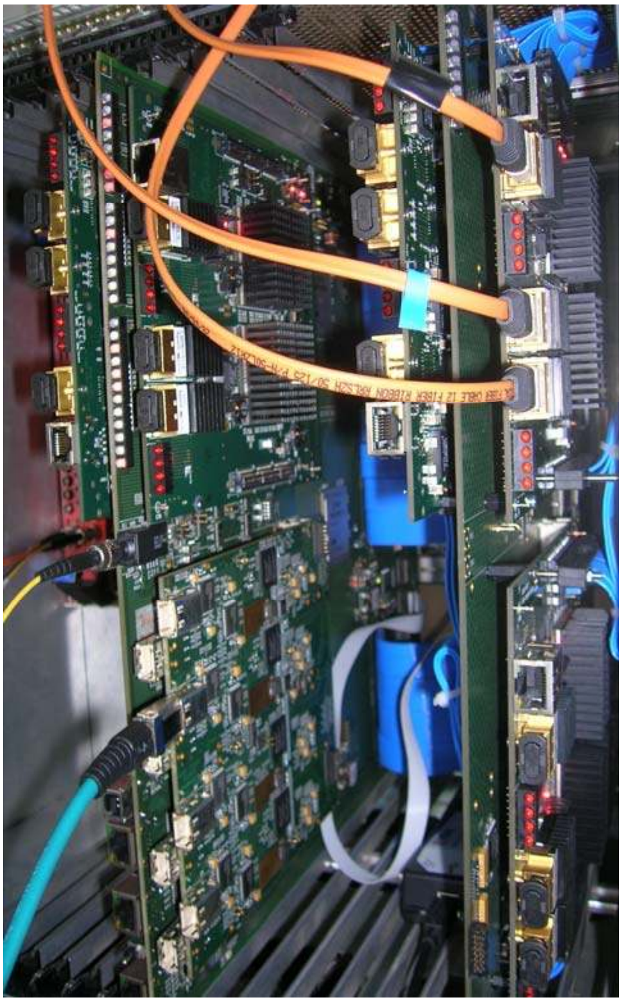

Figure 1: The GCT under test in the laboratory. The card on the left is the Concentrator card with 2 Leaf cards mounted on either side to process incoming electron data. The output to the Global Trigger (GTI card) is mounted at the base of the Concentrator. On the right is a Wheel card with 3 Leafs which process incoming jet data. In normal operation a second Wheel card would be mounted to the left of the Concentrator card to process jet data from the negative η region of the detector. Wide parallel LVDS cables connect the Concentrator and Wheel cards. They are just visible towards the back of the crate.

II. SYSTEM OVERVIEW

The following description is simply to provide a brief overview of the GCT. A far more comprehensive guide is given elsewhere [2,3].

The GCT input interface with the Regional Calorimeter Trigger (RCT) consists of 63 Source cards. Each of these receive 2 cables with 32bit wide, 80MHz, differential ECL data. The data is retransmitted on 4 optical links, each with a data rate of 1.28Gb/s, 8B/10B encoding and CRC check. The links themselves were originally designed to run synchronously with the TTC (Timing Trigger & Control) clock at 1.6Gb/s, however they were subsequently modified to use a local oscillator and operate at 2.0Gb/s, asynchronously to TTC.

There are significant benefits of using optical links. The GCT is electrically isolated from the high power, ECL technology of RCT. The electron, jet and muon data arriving from RCT can be sorted into separate optical links and reassembled at an optical patch panel into a more appropriate grouping and form factor (12 way fibre ribbons) for GCT.

The electron data, transmitted on 54 optical links, are received by two Leaf cards mounted on the Concentrator card. The links are split across the 2 Leaf cards depending on whether they come from the positive or negative η region of the experiment. The Leaf cards determines the 4 highest rank isolated and non-isolated electrons and transmit the result to the Concentrator, which then performs the same task before transmitting the data to the Global Trigger.

The jet data, transmitted on 180 optical links, are received by 6 Leaf cards distributed across 2 Wheel cards (one for each η polarity). The jet data processing is more complex because substantial amounts of data must be shared between Leaf cards. The 3 Leaf cards are connected in a circular fashion so that each card processes data from a 120 degree φ segment, and can share data with the neighbouring Leaf cards. The same sharing requirement also arises at the boundary between positive and negative η because each half of the detector is processed by Leaf cards mounted on different Wheel cards. In this instance the data at the boundary is duplicated in the Source cards so that the Leafs on both Wheel cards have access to boundary condition data. After the jet clusters have been formed they are sorted in the Wheel and Concentrator card before transmission to the Global Trigger.

III. INPUT LINKS

During commissioning in USC55 it was noted that occasionally one of the links on each Leaf card was generating CRC errors. This was a surprise given that there had been substantial testing in the laboratory before deployment to USC55. The main difference between the two tests had been that the laboratory system had used a local oscillator rather than the TTC clock. Furthermore, the TTC clock specification of less than 50ps peak-to-peak jitter was just outside the specification limit for the Xilinx Virtex II Pro.

Consequently, the original hypothesis was that the CRC errors were due to the quality of the TTC clock. The links were therefore modified to use low jitter 100MHz local oscillators on the Source cards and operate asynchronously to the TTC clock. A low latency clock bridge shifted the incoming parallel data on the Source card from the 80MHz TTC clock to the 100MHz local oscillator. Dummy words were inserted where necessary. The link speed jumped from 1.6Gb/s to 2.0Gb/s. Despite these measures the problem was not resolved.

The fault was eventually traced to firmware tools incorrectly routing the recovered Multi Gigabit Transceiver (MGT) clocks despite constraints to the contrary in the User Constraints File (UCF). Normally these clocks would not be used outside the MGT hard IP (Intellectual Property) block, but to achieve a fixed, low latency design, the elastic buffer, which bridges from the recovered serial link clock and the FPGA fabric clock had to be placed in the FPGA fabric [4,5,6].

The tools default to using global clocks when possible; however, there are only 8 true global clocks in a Xilinx Virtex II Pro and our design required up to 16 serial links, each with their own recovered clock, in addition to the main TTC clock in the FPGA fabric. In this situation, the FPGA can route small parts of the design using local clock routing in a dedicated part of the fabric.

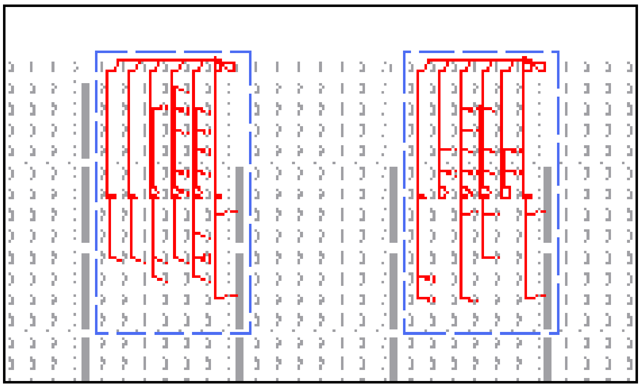

Figure 2: The correct local clock routing adjacent to two MGTs (not visible) along the top edge of the Virtex II Pro. Occasionally the local clock routing exceeded the boundary (dashed blue line). The local clock (solid red lines) connects to the SelectRAMs (large rectangles) which are used as FIFOs to bridge between the two clock domains. The arrays of Configurable Logic Blocks (CLBs) that are used for general purpose logic are just visible (small grey marks).

The tools should have constrained the clock to the local clock region as shown in figure 2. The MGT local clock route is a 5 x 12 Configurable Logic Block (CLB) array on the top of the device and a 5 x 11 CLB array on the bottom. There are also two block SelectRAMs within each MGT local clock domain.

The local clock routing stayed within the routing boundary after two changes were made. The first was the removal of an asynchronous signal clocked in the TTC clock domain, but used in the local clock domain. The second was CRC check in the local clock domain. It is not understood why these changes made a difference, however the local clock routing now seems to respect the boundary. A similar problem was seen much later with the MGTs that were routed with global clocks. This issue was simply fixed by forcing the use of a local clock with a constraint within the VHDL file. The firmware has now been synthesised, placed and routed several times and the problem has not recurred.

The system continues to operate with asynchronous links which has the benefit that we can use a very low jitter clock source and the latency is not affected because the increase in latency due to the clock domain bridge on the Source card is cancelled by the internal logic in the SERDES units operating faster.

IV. Output links

The original GCT-to-GT interface was based on National Semiconductor DS92LV16 [7] electrical high speed serial links operating just beyond specification of 1.6Gb/s. In the revised GCT design, these legacy links were placed on a dual CMC daughter card; the Global Trigger Interface (GTI) card. The links are DC coupled and connect to the GT via 100Ω impedance InfiniBand cables with HSSDC2 connectors. The DS92LV16 chips serialize a 16bit word at up to 80MHz, bounding it with start/stop bits.

The interface was successfully tested with 3.0m cables manufactured by LEONI [8]; however, it was not possible to procure more from this company. An alternative supplier, Amphenol Interconnect [9], provided 1.5m cables. However, when new shorter cables were used for the GCT-to-GT links it was noticed that the SERDES links occasionally lost lock. This was traced to reflections from the receiver rather than any issue with the cable itself.

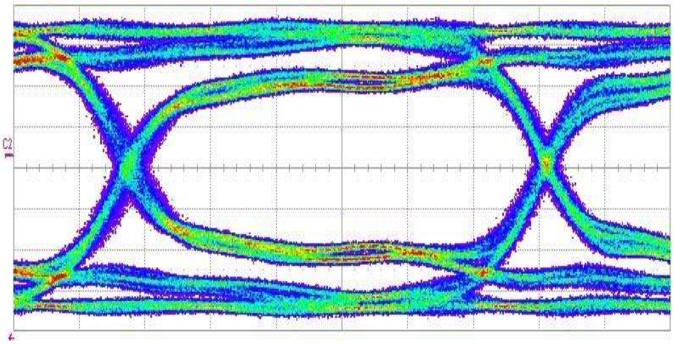

Figure 3: The eye diagram of the signal transmitted by the GTI card after it had traversed an HSSDC2 connector, 1.5m InfiniBand cable, a second HSSDC2 connector and then terminated with 100Ω. Horizontal scale = 100ps/div. Vertical scale = 150mV/div. Receiver switching threshold = +/- 100mV.

A good quality eye diagram (fig. 3) is measured when the signal transmitted from the GTI card is measured without the receiver, but with cable, connectors and 100Ω termination. This is not the case when the card is placed in loop back mode and the signal measured across the termination resistor immediately prior to the receiver (fig. 4). To rule out any PCB issue the eye diagram was measured in the same location, but on a separate unpopulated PCB (except for 100Ω termination resistor). The results were very similar to those in fig. 3.

It was suspected that the degradation in the eye diagram was caused by the signal being reflected of the receiver, which was estimated to be ~7mm from the differential oscilloscope probe.

To confirm the hypothesis a signal consisting of just the start mark (defined as ‘1’), payload of ‘0x00’ and stop (defined as ‘0’) was repeatedly transmitted and measured across the receiver input termination.

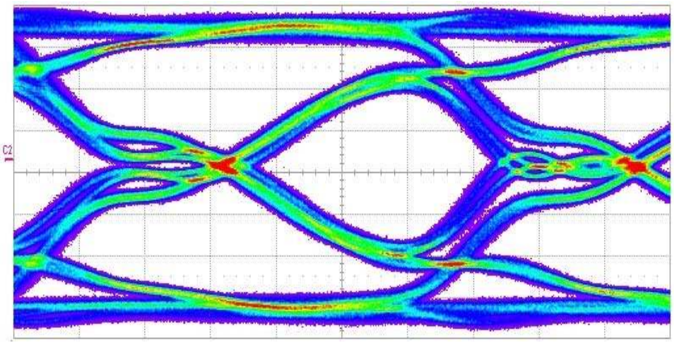

Figure 4: The eye diagram measured across the 100Ω termination resistor immediately prior to the receiver. The transmission path includes 1.5m InfiniBand cable and two HSSDC2 connectors. Horizontal scale = 100ps/div. Vertical scale = 150mV/div. Receiver switching threshold = +/- 100mV.

The start pulses are visible of the far left and right of fig. 5. A suspected reflection is visible approximately 3 divisions or ~4.5ns after the start pulse. The propagation delay of the 0.5m LEONI cable used here is unknown, but the nominal propagation delay of the comparable Amphenol cable is ~4.25ns/m. Consequently, the conclusion is that a reflection has travelled back to the transmitter and has been reflected again and thus when we measure it has traversed 1.0m. Soldering a 0201 package 100Ω resistor directly across the pins of the receiver did not improve the signal quality. The current system in USC55 is precarious and the intention is to replace it with the new GCT-GT interface described below as soon as possible.

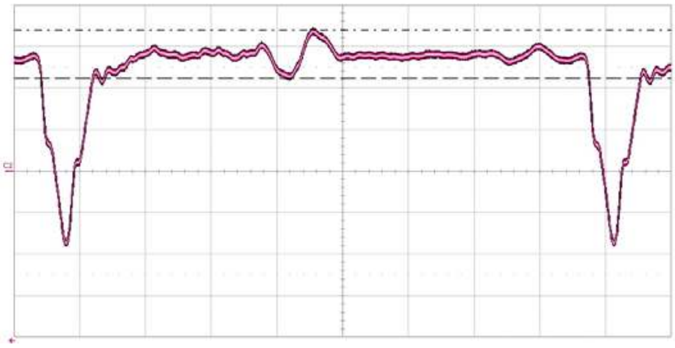

Figure 5: The signal measured across the 100Ω termination resistor immediately prior to the receiver. The transmission path includes 0.5m, InfiniBand cable and two HSSDC2 connectors. Horizontal scale = 1.5ns/div. Vertical scale = 200mV/div. Receiver switching threshold = +/- 100mV.

V. Optical Global Trigger Interface

The new interface to the Global Trigger is being built around a Xilinx Virtex 5 FPGA and 16 bidirectional optical links based on 4 POP4 transceivers. The Optical Global Trigger Interface (OGTI) will use the same dual CMC form factor as the original GTI card and will be capable of both transmitting and receiving data and thus be used at both GCT and GT end of the link. The GT will require a new motherboard to provide an interface to their custom backplane.

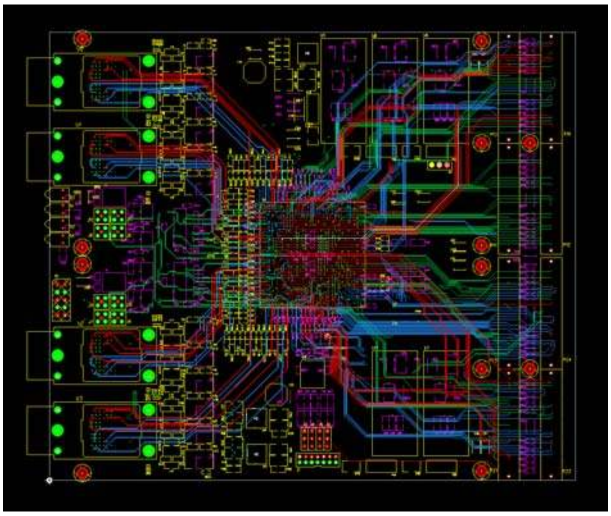

Figure 6: The layout of the OGTI card. The 4 POP4s (optical transceivers) are visible on the left hand side. The Virtex 5 is situated in the middle with clock distribution in the centre of the left hand side. The right hand side is filled with CMC headers.

The Xilinx Virtex 5 (XC5VLX110T-3FF1136C) will operate up to 3.75 Gb/s, however the parallel POP4 optics manufactured by AvagoTech (HFBR-7934Z) and Zarlink (ZL60304) are specified for use up to 3.2Gb/s. Each POP4 contains 4 multimode transceivers operating at 850nm with multimode fibre interface. The baseline design is to run these links in the same asynchronous mode as those in the GCT-GT links, but at 2.4Gb/s rather than 2.0Gb/s to reduce latency. The latency from just the SERDES itself falls from 5.0bx at 1.6Gb/s to 3.3bx at 2.4Gb/s, and 2.5bx at 3.2Gb/s. It would therefore be useful to run the links as fast as possible. Initially the links will be filled with dummy words; however the possibility remains of being able to potentially transmit extra information to GT. Board layout is complete and it will be submitted for manufacture within the next few weeks.

VI. Conclusions

The high bandwidth available from high speed serial links and the integrated SERDES blocks within FPGAs make them attractive for high energy physics electronics; however, it should be noted that the two main hardware problems faced by the GCT project were both related to high speed serial links. This may, at least in part, be because the technology used was not as mature as it is now.

Operating the link in a semi-synchrous mode, in which data synchronised to the experiment wide TTC (Trigger Timing & Control) system is sent over an asynchronous, fixed and low latency link is not completely trivial. It has the advantage that the link reference clock can can be provided by a low jitter local oscillator. The disadvantage is the complexity of the firmware, which must contain buffers to bridge the data from the link clock domain to the main TTC clock domain with a fixed and low latency. As local clock resources become a standard feature of FPGA fabric this should become easier.

VII. Acknowledgements

We would like to thank Sarah Greenwood (Imperial) for layout of the Optical Global Trigger Interface card; PPARC for financial support; the PH/ESE group at CERN for providing a high bandwidth sampling scope to debug the output links.

VIII. References

[1] The Trigger and Data Acquisition Project, Vol. I, The Level-1 Trigger, CERN/LHCC 2000-038, CMS TDR 6.1, 15 December 2000.

[2] M. Stettler et al., "The CMS Global Calorimeter Trigger Hardware Design", 12th Workshop on Electronics For LHC and Future Experiments, Valencia, Spain, 2006, pp.274-278

[3] G. Iles et al., "Revised CMS Global Calorimeter Trigger Functionality & Algorithms", 12th Workshop on Electronics For LHC and Future Experiments, Valencia, Spain, 2006, pp.465-469

[4] Matt Dipaolo & Lyman Lewis, "Local Clocking for MGT RXRECCLK in Virtex-II Pro Devices", Xilinx Application Note: XAPP763 (v1.1), 2004

[5] Emi Eto & Lyman Lewis, "Local Clocking Resources in Virtex-II Devices", Xilinx Application Note: XAPP609 (v1.2), 2005

[6] Jeremy Kowalczyk, "Minimizing Receiver Elastic Buffer Delay in the Virtex-II Pro RocketIO Transceiver", Xilinx Application Note: XAPP670, 2003

[7] DS92LV16 DataSheet: 16-Bit Bus LVDS Serializer/Deserializer – 25 – 80 MHz, National Semiconductor, Feb. 2002

[8] LEONI Special Cables GmbH, Eschstraße 1, 26169 Friesoythe, Germany

[9] Amphenol Interconnect, Products Corporation, 20 Valley St.Endicott, NY 13760, USA