The GCT Muon and Quiet Bit System

Design and Production Status

Matthew Stettler[a d], Kostantinos Fountas[b], Magnus Hansen[a],
Gregory Iles[a], John Jones[c]

[a] CERN, 1211 Geneva 23, Switzerland
[b] Imperial College, London, UK
[c] Princeton University, Princeton NJ 08544, USA
[d] Los Alamos National Laboratory, Los Alamos NM 87545, USA

matthew.stettler@cern.ch

Abstract

The CMS Global Calorimeter Trigger HCAL Muon and Quiet bit processing function is being implemented with a micro TCA system. This system is reconfigurable in both logical functionality and data flow, allowing great flexibility to meet processing requirements. The system consists of a processing module based on a Xilinx Virtex 5 FPGA and custom backplane based on a Mindspeed crosspoint switch. The overall progress of the design will be presented. In addition, future application of this technology for the SLHC level 1 trigger will be discussed.

I. Overview

In order to meet the requirements of the CMS Global Calorimeter Trigger, a system is required to route and reassign HCAL muon and quiet bits forwarded by the Regional calorimeter trigger.

A.Requirements

The GCT HCAL Muon and Quiet bit functionality entails the reorganization of the data as collected by the 18 Regional Calorimeter Trigger (RCT) crates, and transfer to the Global Muon Trigger (GMT). The interface to the GMT has changed in the last year due to ongoing problems with the original DC coupled 1.44Gbps serial links (National Semiconductor DS92LV16). The new design, known as the optoGTI, calls for optical links based on the technology used elsewhere in the GCT, and is described in detail in a related paper[1]. The system needs to accept the RCT data on 18, 2.0 Gbps fibers from the RCT, formatted as an 8b/10b serial stream by the GCT source cards[2]. The output data to the GT is on 16, 2.4 Gbps fibers, also formatted as an 8b/10b stream.

In addition to the physical translation, a simple logical transform must also be applied. The RCT data is organized in 40 degree phi, ½ barrel eta slices per crate. The GCT requires that the data be reorganized into 120 degree, full barrel eta segments.

Since the existing GCT modules are not well suited to these requirements, an additional system is under development, based on the existing GCT module designs.

B.Architecture

This function is being implemented utilizing a multi-gigabit switched serial mesh processing topology. It represents an evolution of the current GCT architecture, taking advantage of the lessons learned implementing the optical data transmission and concentration between the Regional Calorimeter Trigger racks and the GCT leaf cards[2]. This topology is realizable in the micro TCA communications equipment standard, with a custom (though spec compliant) backplane, and has been described in detail in a previous paper[3]. Due to the flexibility of the optoGTI, the Muon and Quiet bits function can be implemented with three processing modules in a commercial micro TCA crate. The core concept is that high speed serial links (both fiber and copper) are used for all communications both internally and externally. Analog crosspoint switching technology is used to provide a flexible communications mesh, allowing a regular hardware topology while retaining significant data routing options. Based on extensive experience with FPGAs in many applications, a concious decision was made to provide plentiful link routing, since connectivity remains the primary limiting factor in fully utilizing the logic resources of large FPGAs.

The design is composed of two major elements, a micro TCA processing module that interfaces directly to fiber I/O, and a high bandwidth implementation of the micro TCA backplane. The processing module is currently in production, and proved to be a challenging design, requiring more modern PC board manufacturing techniques than used in the past.

II. Processing Module Design

The processing module provides the data manipulation functionality to implement muon and quiet bit system logic, and directly interfaces to the fiber input from the RCT (through the GCT source cards). It consists of three fiber I/O modules, a Xilinx V5LX110T or FX70T FPGA, a Mindspeed 21141 crosspoint, and an Ethernet enabled micro controller for slow control. The design has been presented previously[3], but is necessary background for the detailed implementation information that follows.

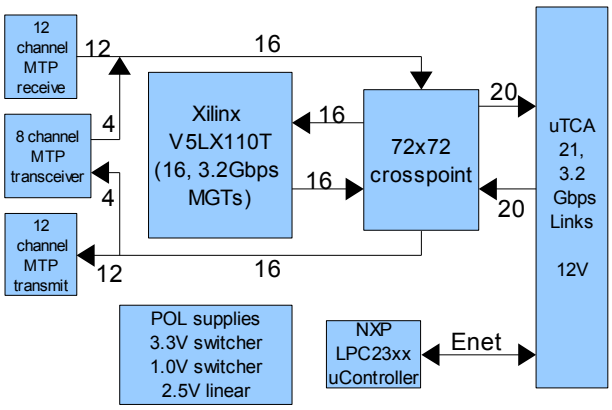

Figure 1: Processing module block diagram

A.Primary Elements

The fiber input modules are the same family of MTP modules used on the GCT leaf cards, and provide the dense packaging required to physically concentrate data to feed the large FPGA. These modules provide either 12 input, 12 output, or 4 in and 4 out. They are currently available rated up to 3.2Gbps.

The processing FPGA is a Xilinx V5LX110T or V5FX70T, which provides 16 3.2Gbps (6Gbps in FXT70) serial links in addition to generous logic and routing resources. This family of FPGAs also provides analog PLLs, which result in more stable frequency synthesis. All control and configuration information for the Mindspeed crosspoint flows through the V5, allowing firmware to directly control local data switching if required.

The Mindspeed 21141 crosspoint is the data hub of the module, routing data to/from the optics, FPGA, and backplane. Since all data flows through the crosspoint, it can be routed, or duplicated, to any destination. The crosspoint switch automatically detects and powers down unused links to reduce power consumption, and includes analog conditioning to clean up degraded signals.

The micro controller is the slow control interface, and includes an integrated Ethernet MAC. This device supports TCP/IP sockets, simple telnet, and http protocols. It also supports I2C, which is the system management interface of uTCA. It performs all the required negotiation with the backplane during module initialization and removal. In addition, it is possible to program the FPGA and configuration memory via the micro controller. The device chosen is the NXP (Phillips) 2368, an ARM-7 based device with 512K of FLASH on chip, and many integrated peripherals in addition the the Ethernet MAC. The selection criteria was maximum integration, and though not impressive, the performance is more than adequate for control and configuration tasks.

B.Power

The module power subsystem, although not as functionally interesting, is a critical part of the module design. The power subsystem consists of two parts, the uTCA mandated power management logic, and the high current analog and digital power required by the FPGA and crosspoint.

The module receives 3.3V management power and 12V payload power from the backplane. The management power is activated first, and powers the micro controller and related logic. When the module is plugged in, or the system is powered up, the micro controller negotiates with the backplane, which then commands the crate power module to energize the payload power. Similarly, when the module is unplugged from a running system a micro switch on the ejector signals the micro controller to shut down the payload. A front panel LED is used to indicate that it is safe to remove the module.

More critical from an engineering standpoint is the low voltage generation scheme. Five voltages (3.3V, 2.5V, 1.8V, 1.2V, 1.0V digital/analog) are required for the various core and I/O loads on the module. These are derived from three switching POL supplies, running at 3.3V, 1.8V, and 1.0V. Analog regulators supply the 2.5V, 1.2V, and 1.0V analog. The 1.2V and 1.0V analog supplies power the crosspoint and FPGA serial links, and require careful attention to achieve reliable link operation. Due to the potentially high power required by the crosspoint (10 watts), this more complex supply was prototyped to verify it's performance

C.Clocking

The module supports a simple clock distribution scheme designed to supply the FPGA with a low jitter reference clock, and general logic clocks. The clock tree is based on a differential 4x4 discrete crosspoint that connects both backplane clock inputs, a local oscillator, and an output from the crosspoint to 4 groups of 2 high speed serial reference clocks, 1 global clock, and 1 crosspoint input. Of these clock sources, the local oscillator and backplane clocks are best suited for serial link references.

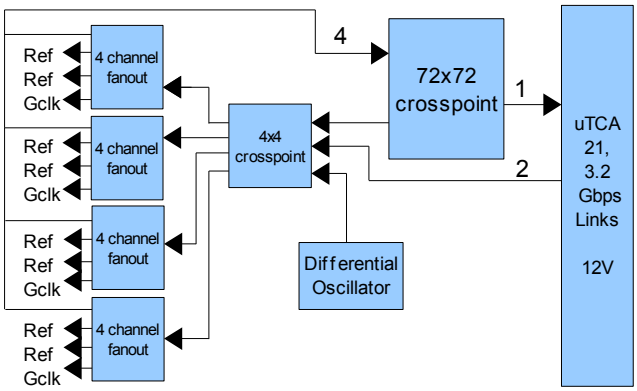

Figure 2: Module Clock tree

D.New PCB Structures

The module implements over 150 multi Gbps differential signal and clock pairs, and required the use of more advanced PCB design and construction techniques than used on the GCT. When preliminary design files were sent to the PCB vendor, they recommended that the construction technique be modified to including new structures to reduce fabrication risk. These techniques are considered main stream by the PCB

vendor, and in their view reduced fabrication risk. The prototype production had a yield of 100% (admittedly on only 5 boards), which supports their assertion. The new technologies used on the module are via in pad, micro vias, and build up PCB manufacture.

With BGA ball pitches of less than 1mm becoming more common, it is becoming more challenging to use traditional escape techniques and maintain required clearances. This is especially true when drill vias are used. The alternative of drilling the via directly through the pad has always carried the risk of compromising the solder joint of the BGA ball. The void in the pad wicks solder away from the BGA ball, starving the joint of solder and creating a weak bond. As the via gets deeper (more layers traversed), the worse this problem becomes. The module utilizes BGAs with .8mm ball pitch, which precluded the standard escape pattern. Instead, a 6 mil laser drilled via is used directly through the pad. The use of such a small via, which only penetrates 2 layers, produces a very small void in the pad, and is considered acceptable for reliable bonding of the BGA.

Micro vias are very small laser drilled structures that span a single PCB layer. Their small size and span makes them ideal for high speed signalling, providing minimal inductance and impedance discontinuity. Obviously, their small span is a limitation, and they need to be used with other construction techniques to be very useful. Typically, they are aligned with an underlying via (stacked), to allow for greater spans. In order to properly join vias in this manner, the underlying via needs to be filled to create a large enough contact surface.

It is typical to combine micro vias with a incremental PCB fabrication technique known as build up manufacturing. This entails depositing one layer of the PCB at a time, and laser drilling the micro vias for that layer before going on to the next. This allows the micro vias to be stacked to provide many more blind and buried spans than previously possible, while providing better registration for small PCB structures.

E.Module Stackup

In comparison with the GCT leaf cards[2] , the processing module used fewer layers with more via spans. The GCT leaf cards were constructed in 14 layers with four via spans in a three layer laminate process. Maintaining proper registration in a three layer laminate was a manufacturing challenge, and resulted in a surcharge for production. In contrast, the uTCA processing module was constructed in 12 layers with 8 spans in a 2 layer laminate with two layers of buildup (both top and bottom). This board was considered standard process, and will be less expensive to produce. In our case, the cost reduction was not the goal, but rather to use the extra via spans to successfully route the many high speed pairs within the overall board thickness allowed.

The detailed stackup of the module is shown in figure 3. It was constructed as two 4 layer boards with through drill vias laminated together. The laminated assembly was then drilled through, and vias filled. A single built up layer was added top and bottom, and micro via spans (1 layer) laser drilled. These vias were then filled, and the final top and bottom layers deposited. The final micro via spans were added, completing the construction.

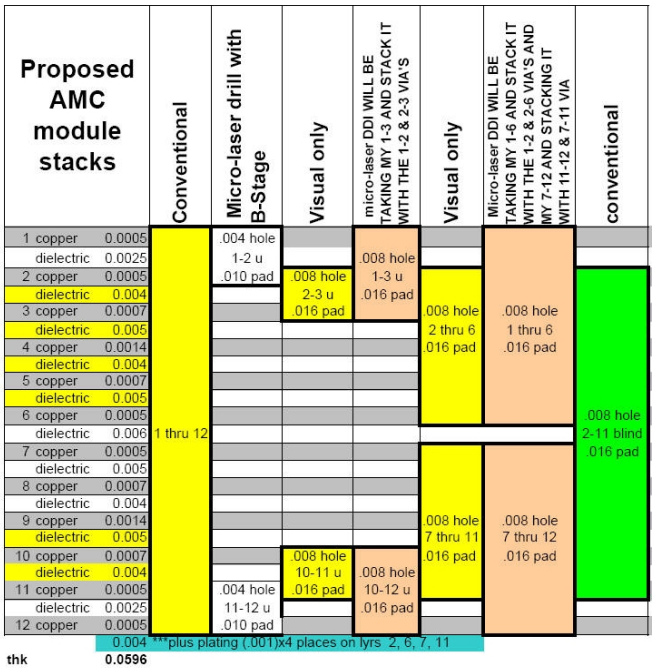

Figure 3: Module stackup

Since the design has many high speed differential pairs, 4 tightly controlled impedance layers were necessary. In this stackup, layers 3, 5, 8, and 10 are impedance controlled at 100 ohms (10% tolerance). Layers 1 and 12 are more loosely controlled (~20% tolerance) as well. Figure 4 shows typical routing density on the controlled impedance layers. Since layers 3 and 10 had direct micro via access with no stubs (via extending beyond the signal layer), these were the preferred routing layers for the most critical signals.

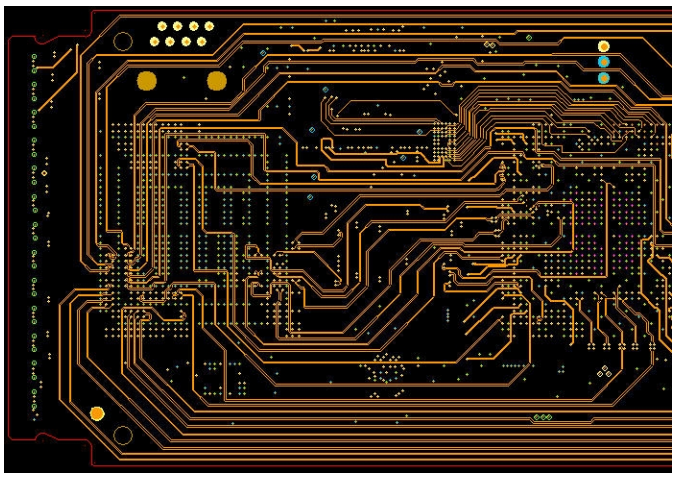

Figure 4: Partial routing on layer 10

F.Signal Integrity

The requirement of maintaining a 100 ohm impedance, and meeting micro TCA board thickness requirements necessitated the use of 3.5 mil trace widths on all stripline differential pairs (layers 3, 5, 8, and 10). Such small dimensions are not optimal for multi GHz signals, and several tests were performed to insure adequate signal amplitude over the trace lengths used in the design. In addition to software simulations, a test was performed on equivalent coax to verify the attenuation characteristics. Figures 5 and 6 show the results of the equivalent coax test. Note that 10-20% pre-

emphasis will be required to maintain full amplitude on the receiving end.

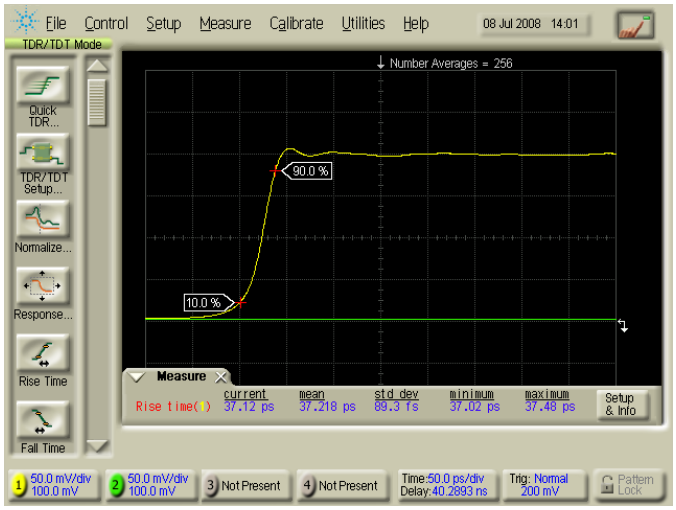

Figure 5: Input test pulse (equivalent coax test)

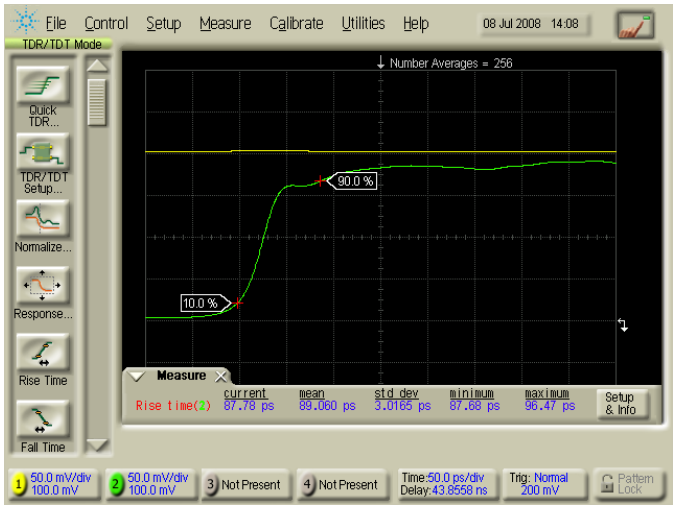

Figure 6: Output pulse (coax equivalent of 14 cm trace)

Recognizing that our testing is really a “best case” scenario, a test board has also been designed using the same stackup and PCB structures to verify the design margins. While unfortunately too late for the processing module, the results will be used to guide the backplane design, which we believe will have even higher routing congestion. The test board utilizes high performance edge mount connectors (over 10GHz analog bandwidth) and is designed for easy connection to network analysers, TDRs, and high bandwidth oscilloscopes. The test board is the same form factor as the processing module, so that in addition to testing board trace properties it can also be used to test backplanes and connect test equipment to running systems. Several connections are available on the front panel for this purpose, while the trace test connectors are designed for bench use only. During assembly a board will be dedicated to one of these purposes, and will have a subset of all possible connectors installed. The test board is now in the process of being ordered, and should be available within a few months.

III. Algorithmic Implications

Experience on the GCT has shown that much more effort is spent on the firmware design and integration than on the initial hardware design and test. While the initial application of the processing module in the GCT Muon and Quiet bit system is rather trivial, using such modules in future trigger applications will not be.

A. *Current State of Trigger Processing*

The GCT implements the jet finding algorithm defined by CMS as large combinatorial functions. The details are a bit more complex, but the goal was to achieve the jet finder with as few clocks as possible. This is due to the fact that the latency budget is limited, and large combinatorial functions avoid the dead time associated with the combinatorial settling time versus the clock period. Currently this is hand tuned to optimize throughput by matching these times as closely as possible while maintaining the required setup time for the pipeline registers. The jet candidates identified by each jet finder are then forwarded to a series of comparator stages which sort the candidates and forward the most energetic three to the Global Trigger. Due to the large number of parallel connections required to transport the jet objects, much of the GCT is composed of I/O bound FPGAs that utilize a small fraction of their logical capacity. The large number of signals that need to be routed between stages drive the design of the system, and favour large boards with extensive cabling. This situation in not unique to the GCT, but is shared by many subsystems of the CMS Trigger.

B. *Future opportunities*

The hard serial nature of the processing module lends itself to a physical concentration of data not possible in current systems. Along with this advantage comes the challenge of processing much larger data sets as efficiently as possible.

The current scheme of hand tuning the combinatorial object finding functions works well, but will run into problems as the number of regions scanned increases, or if several types of data are required to properly discern trigger objects. The root cause of this is that the size of the functions increase geometrically with a linear increase in the number of inputs. Unfortunately these are exactly the algorithmic modifications being discussed for the SLHC upgrade. The result of this situation will be much longer compile/optimization run times, and a loss of synthesis efficiency since the FPGA devices are being used in a mode which does not take full advantage of their functional architecture.

The optimal solution is not obvious, and will depend on the details of the algorithmic modifications deemed appropriate. An obvious implementation option is to utilize a pipeline clocked by a multiple of the LHC clock that approaches the limit of the FPGA. While less efficient than a combinatorial implementation, it will result in much faster optimization and higher synthesis efficiency. The key challenge in this implementation will be to break the algorithm into small stages compatible with a few levels of FPGA logic. It is likely that this approach will lead to FPGA

code that is harder to read, but more efficient and reliable to synthesize.

A related option is to evaluate new technologies that may be better suited to the primary algorithms than what is currently being used. Archronix[4], a relative newcomer to the FPGA business, has developed an FPGA based on a asynchronous handshaking model. In this model the data flows through the processing logic as fast possible, so no dead time accumulates due to setup margins required to accommodate a global clock. Figure 7 illustrates the general idea of this approach. The efficiency gained allows these FPGAs to run at 3 times the speed of comparable synchronous designs – 1.5GHz at 65nM. While conceptually simple, it is obvious that for complex designs the overall data interlocking will not be trivial. While Archronix claims these complexities are handled transparently in their technology mapper, allowing the use of familiar synchronous RTL design entry, it will require serious evaluation before they can be considered as a viable alternative. However, if the tools prove to be robust, these devices will likely provide the most efficient implementation of future trigger algorithms.

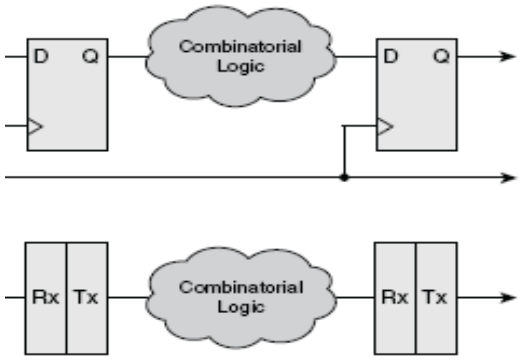

Figure 7: Synchronous vs. Asynchronously interlocked logic

IV. Conclusions

The uTCA architecture lends itself well to the processing requirements of the GCT muon and quiet bit system and provides a path forward to the future. The hard serial nature of the system lends itself well to the physical concentration of data required to implement more sophisticated trigger algorithms currently being discussed. Much work still needs to be done in the definition of SLHC trigger requirements and evaluating the appropriate technology for implementing future systems.

A.Current Status

The processing module PCB has been fabricated and received at Los Alamos. Assembly has been delayed due to problems with finding an appropriate vendor. A vendor has currently been selected and finished boards are expected by the end of October. The test board is currently being quoted, and is expected in the same time frame as the processing module.

B.Future Development

The modular nature of the system, with its considerable data routing flexibility, make it an attractive architecture for future trigger system development on the SLHC. Basing a large trigger system on a high bandwidth fine grained modular commercial standard would allow a degree of standardization not possible with the traditional full custom approach. As it stands, the generic nature of the V5LXT/FXT FPGA and it's built in features suggest that many applications could be addressed. The technology utilized is also compatible with the Archronix asynchronous devices, but would need a board spin to support the pinout and power supply configuration required by these parts. Indeed, the module is little more than a stand alone FPGA carrier with a fiber interface and multi-gigabit serial switching support.

The challenges of firmware development for such a system are beyond the scope of the paper. Experience on the GCT has shown that utilizing many serial links on the same device is not trivial, and moving to a hard serial architecture will highlight these difficulties. Focusing on standard implementations supported by the manufacturers will be key in overcoming this challenge. Merely providing a solid hardware platform will not insure success. Much effort will need to be made in providing a stable, reusable framework of firmware modules to allow designers to be productive and keep integration efforts reasonable.

V. References

[1] G. Iles et al, *Performance and Lessons of the CMS Global Calorimeter*, TWEPP 2008.

[2] M. Stettler et al, *The CMS Global Calorimeter Trigger Hardware Design*, 12th Workshop on Electronics for LHC and Future Experiments, 2006.

[3] M. Stettler et al, *Modular Trigger Processing, The GCT Muon and Quiet Bit System,* TWEPP 2007.

[4] Achronix, *Speedster FPGA product line,* http://www.achronix.com.