Modular Trigger Processing,
The GCT Muon and Quiet Bit System

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 system's HCAL Muon and Quiet bit reformatting function is being implemented with a novel processing architecture. This architecture utilizes micro TCA, a modern modular communications standard based on high speed serial links, to implement a processing matrix. This matrix is configurable in both logical functionality and data flow, allowing far greater flexibility than current trigger processing systems. In addition, the modular nature of this architecture allows flexibility in scale unmatched by traditional approaches. The Muon and Quiet bit system consists of two major components, a custom micro TCA backplane and processing module. These components are based on Xilinx Virtex5 and Mindspeed crosspoint switch devices, bringing together state of the art FPGA based processing and Telcom switching technologies.

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). In addition, the serial encoding of the data needs to be changed to provide compatibility with the GMT. While computationally fairly straightforward, the number of channels (18, 1.6Gbit in, 24, 1.44Gbit out) is significant.

The system needs to accept the RCT data on 18, 1.6GHz fibers from the RCT, formatted as an 8b/10b serial stream by the GCT source cards[1]. The output data to the GT is on 24, 1.44GHz DC coupled cables, compatible with the National Semiconductor DS92LV16 (16 bit NRZ with two frame bits).

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[1]. This topology is realizable in the micro TCA communications equipment standard, with a custom (though spec compliant) backplane. 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. In addition, a simple reformatting card is required to buffer the output for transport to the GT over copper cables.

Such a modular design is not only well suited to the GCT muon and quiet bit system, but can also be of use for general trigger processing. The combination of fine grained processing modularity and flexible data routing make it an attractive choice for many high bandwidth computational tasks. Basing the system on a commercial standard brings the typical advantages of standard modules and infrastructure support.

II. Short Background on Micro TCA

A. ATCA and uTCA

The micro TCA standard is derived from the ATCA (Advanced Telecommunications Architecture) standard,

developed under the auspices of the PICMG[2] group. ATCA is targeted to large scale switching and routing applications. It provides generous high speed data throughput, and advanced system control capabilities designed to facilitate robust, high performance implementations. Another interesting feature is that it is a hard serial standard, utilizing no parallel buses for any global functions. Ethernet is used for traditional command and control. In the recent past, ATCA has been suggested as a standard worthy of consideration for HEP implementations[3]. The electrical standard of the backplane is AC coupled CML logic, usable for Gig/10G Ethernet, PCIExpress, serial ATA, serial RapidIO, and numerous custom protocols.

Micro TCA (or uTCA) was developed as a stand alone version of an AMC (Advanced Mezzanine Module), which in turn was derived from the CMC/PMC standard. The AMC modules were envisioned to be hosted on ATCA carrier boards, and implement a subset of the ATCA system management functions. The primary target application of micro TCA is small form factor switching and routing gear, as is common in cell phone base stations.

The uTCA standard calls for system management functions, which enable hot swapping, module compatibility checking, and redundancy. Similar to ATCA, backplane power is separated into management and payload power to allow intelligent power management. Special ejector hardware is required that contains a switch that provides an indication when a module needs to be shut down upon extraction. The modules are 75x180mm (single width), and support 21 full duplex high speed serial links running up to 10Gbps. Each backplane has at least one power module, whose function is to provide and switch payload power to each slot. Redundant power schemes are encouraged, but not required.

Figure 1: micro TCA crate with single high backplane

B.Advantages of uTCA

An interesting feature of uTCA is that it provides a much denser high bandwidth solution than ATCA, which brings several advantages to trigger processing systems. The small size of the modules and number of links closely match the minimum module size required by today's large high speed serial link enabled FPGAs, such as the Xilinx V5LXT/SXT series. In addition, the front panel area is sufficient to mount enough high density optical interfaces (such as SNAP-12 and

POP-4), to fully link to the backplane. This symmetry allows one to construct a flexible high bandwidth system of unprecedented density. Also, the improved airflow and larger envelope allowed by uTCA supports a power density of 80 watts per slot, an important consideration when designing with high performance devices. It is for these reasons that uTCA was chosen as the base architecture for the GCT muon and quiet bit system.

While the potential raw bandwidth of both uTCA and ATCA is very large (500Gbps for single high uTCA crate utilizing 3.2Gbps links), this is impossible to attain with existing commercial backplanes. Both uTCA and ATCA rely on "hub", or "switch" cards to provide the routing of the high speed links, creating a bottleneck that severely limits the flexibility of the system due to pin constraints. However, the uTCA specification leaves open the possibility of an active backplane – which can be designed to provide much greater connectivity, and avoid dataflow bottlenecks.

C.Related Developments

A technology closely related to the development of the ATCA standards is high capacity crosspoint ICs. These devices, which support up to 144x144 full duplex channels, have been designed to support the same large scale switch applications that the ATCA architecture targets. These are non-blocking, asynchronous, protocol agnostic devices, which allow a mix of data rates and support all protocols usable on the uTCA/ATCA backplanes. Using such devices on an active uTCA backplane results in a system that provides both the raw bandwidth and routing flexibility to compliment the processing capability of the largest modern FPGAs. Our implementation of the uTCA backplane includes such a switch, and dedicates one of the 21 links from each slot to 10/100 Ethernet as the slow control interface. Since the backplane itself acts as the hub, no hub slots are provided in the design.

III. 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 FPGA, a Mindspeed 21141 crosspoint, and an Ethernet enabled micro controller for slow control.
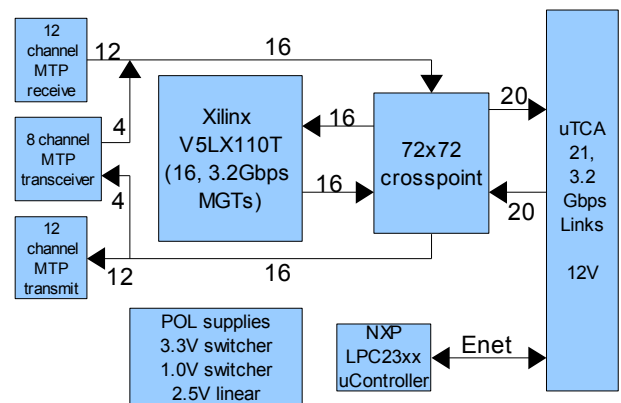

Figure 2: Processing module block diagram

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, which provides 16 3.2Gbps 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, possibly allowing direct generation of unusual protocols such as that used by the National Semiconductor DS92LV16 required by the GT. 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 (12 watts), this more complex supply is being 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.
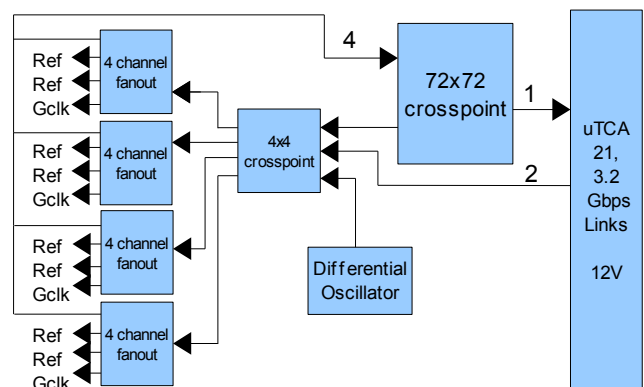

Figure 3: Module Clock tree

IV. BACKPLANE DESIGN

While it is always unfortunate to design a custom backplane, especially for a standard bus, the performance gain in the case of uTCA is potentially very significant. The uTCA standard has been out for approximately one year now, and hopefully higher performance commercial units will be available in the future, but presently the commercial support is focused on the high volume lower performance applications.

A critical decision in the design of a point to point serial backplane is the connectivity model. Generally, there are two choices, mesh or star. In a mesh, each slot is connected to a maximum number of other slots, in a star, all slots are connected to a single slot. While a mix of these approaches is obviously possible, on a passive backplane there is no avoiding the hard wired nature of the chosen architecture. Another approach is to use an active switching device to create the desired connectivity. In addition to providing far greater flexibility in point to point data routing, an active buffering switching device can also duplicate data, allowing broadcast to many destinations simultaneously.

Current commercial backplanes perform this function in a special hub slot, which is a variation of the star topology, but are quite limited in the total number of switched links due to pinout limitations of the hub slot.

These issues have driven the decision to build an active switched backplane, with a large number of switched links per slot. With the addition of the switching resources on the processing module, this provides complete freedom in system data routing.
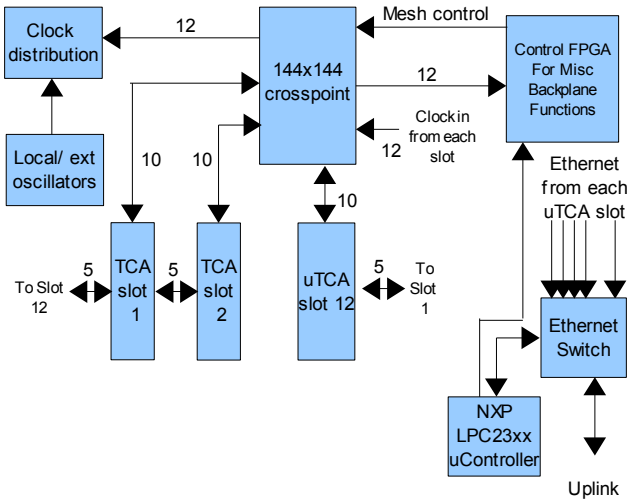

Figure 4: Backplane block diagram

A. Link Allocation

Each uTCA slot has 21 full duplex serial links available, which are distributed as follows. A total of ten links, 5 in each direction, are hard wired to each neighbouring slot. Ten links are wired to a Mindspeed M21161 crosspoint switch. One link is used for Ethernet, the slow control interface. Each slot also provides 3 clock pairs, 2 dedicated input and one dedicated output. One of these input pairs is wired to a dedicated low skew fanout tree, intended for use as a serial clock reference. The other clocks are wired to the switch as a general logic clocks. Twelve remaining links are wired to the control FPGA, to allow each slot access to global configuration functions.

B. Control Functions

The global control functions required by uTCA are provided by a Xilinx spartan3 FPGA and NXP 2368 micro controller. The spartan3 provides JTAG and I2C interfaces to each slot, as well as crosspoint switch control. The micro controller provides an Ethernet interface for external slow control, and uses the spartan3 to access the I2C and JTAG.

The spartan3 FPGA controls access to the configuration and control ports of each uTCA slot (JTAG and I2C) - which are arranged in a star configuration, performs module presence detection, and interfaces to the power controller. It acts as a peripheral of the micro controller, electrically isolating it from the slots. In addition, it interfaces directly to the crosspoint. This allows a high performance (to the limit of spartan3 I/O) connection to the uTCA modules. This connection is intended to be used to directly access crosspoint

reconfiguration functions for dynamic load switching or other advanced schemes.

The micro controller performs the required uTCA crate control functions, as well as providing a means to configure the crosspoint and access the crate JTAG chains remotely. The plan is to implement a minimum set of crate control functions, mainly power sequencing. The uTCA specification defines many command types and functions, but only the minimum required to detect and power the modules will be initially implemented. Similar to the processing module, the backplane micro controller supports TCP/IP sockets, a simple telnet server, and simple http service.

C. Clock Distribution

The backplane supports two clock distribution systems, a dedicated low litter discrete fanout similar to that used on the module, and a general purpose distribution based on the crosspoint switch. The precision tree is sourced by a small, low jitter, 4x4 crosspoint which is fed by up to three discrete reference oscillators and external coaxial inputs. This tree feeds one clock input per backplane slot (uTCA clk#1). The clock output from each slot (uTCA clk#2), and the remaining clock input (uTCA clk#3) are connected to the crosspoint. While the jitter specification of the crosspoint is not tight enough for a high speed serial reference clock (according to the Xilinx V5 specification), it is easily adequate for general purpose logic clocking.

V. LINK CONVERTER

In order to communicate with the GMT, the 8b/10b serial encoding scheme typically used in the high speed serial links needs to be translated into a 16 bit NRZ scheme used by the GMT, which was imposed by an earlier version of the GCT design. Currently two designs are being considered for this conversion process. It is possible to directly generate the NRZ 16 code in the FPGA, and simply buffer the output to drive several meters of cable. This scheme is the simplest, but requires a DC coupled signal path from the FPGA to the buffer output. Another advantage of this scheme is that it provides the lowest possible latency. An alternative approach is to recode the 8b/10b into NRZ16 using a decoder/encoder pair. This approach has the advantage of retaining AC coupling until the last possible moment, and uses the built in encoding logic of the FPGA serializers. A link converter using an encoder/decoder is planned to reduce risk in the overall design.

VI. CONCLUSIONS

The uTCA architecture lends itself well to the processing requirements of the GCT muon and quiet bit system. In addition, it provides a path forward to more modular high performance processing systems in the future.

A. Current Status

The initial design work on the processing module and backplane has been completed for some months. The processing module is currently in layout at Los Alamos Laboratory in the US, with initial hardware expected by the

end of the year. The backplane layout will begin as soon as the processing module is released for fabrication.

B. Larger Processing Arrays

Additional funding for this development has been provided by Los Alamos Laboratory, which is planning to use the architecture as part of it's signal processing research. The planned use of uTCA is in building a larger module array for video processing. Initial testing will be done with the same modules designed for the GCT system, but more specialized modules are planned. The new modules will include multi-FPGA (V5LXT/SXT) processing engines and DSP based secondary processing boards. The system will also require a fully switched double height backplane capable of a sustained data rate of >500Gbps.

C. Other Features

Several other features of the processing module are worthy of note. In addition to the high speed serial interfaces, the Xilinx Virtex 5 FPGA contains a Gigabit Ethernet MAC and a PCIExpress endpoint. At the request of Los Alamos, provisions have been made in the design for 512MB of DDR2 SDRAM, organized as two 128Mx16 banks. While these are intended for coefficient storage and waveform buffers, they can obviously put to more general use. These hardware features compliment the serial mesh architecture and optical I/O, enabling powerful new applications.

Commercial PCIExpress fiber extenders are becoming available[4], and will allow standard PCs to directly access the PCIExpress endpoint in the processing modules via fiber patch panels. This opens the door for a new class of high performance control and data acquisition capabilities. Since the fiber links can easily span more than 200 meters, remote PCs could easily interface with multiple crates of uTCA over robust, electrically isolated, standard links.

Similarly, the embedded Gigabit Ethernet MAC allows a direct connection to high performance network based data acquisition and control. While it entails more protocol overhead than PCIExpress, the advantage is that the module becomes an independent node. While in many ways this duplicates the function of the micro controller, there is no comparison in potential performance. The inclusion of a soft core processor such as the Xilinx Microblaze would make this an even more powerful solution. With the SDRAM installed, there would be no problem running reasonably sophisticated software on the module.

In addition to the built in features of the V5, the extra ports of the crosspoint have been wired to standard V5 differential I/O. This gives one the option of using simpler (or slower) data protocols than those supported directly by the dedicated high speed serializers. These I/O are rated to 1.2Gbps, and although this speed may be practically difficult to obtain, should be usable for simple or unusual protocols at up to 1Gbps rates without undue effort. It should be noted that the I/O is AC coupled to the switch, so some form of DC balanced encoding must be used.

D. Future Possibilities

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. An additional significant benefit would be that a standard backplane interface would allow more efficient collaboration between institutions and facilitate greater sharing of module designs. As it stands, the generic nature of the V5 FPGA and it's built in features suggest that many applications could be addressed. Indeed, the module is little more than a stand alone FPGA carrier with a fiber interface and multi-gigabit serial switching support. The intent of the architecture is to allow custom processing arrays to be built easily from standard hardware with a minimum of housekeeping overhead (from at least a hardware standpoint). One only needs to select compatible optical standards to interface directly to sensor front ends and DAQ hardware.

VII. REFERENCES

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

[2] PCI Industrial Computer Manufacturers Group, http://www.picmg.org, 401 Edgewater Place, Suite 600 Wakefield, MA 01880 USA

[3] R. S. Larson, *High Availability Electronics Standards*, 12th Workshop on Electronics for LHC and Future Experiments, 2006.

[4] ADNACO, *Sirius-4 PCI Express extension,* http://www.adnaco.com/products.html,