

# THE CMS HCAL DATA CONCENTRATOR: A MODULAR, STANDARDS-BASED IMPLEMENTATION

E. Hazen, J. Rohlf, S. Wu, Boston University, USA  
 A. Baden, T. Grassi, University of Maryland, USA

## Abstract

The CMS HCAL Upper Level Readout system processes data from 9300 detector channels in a system of about 26 VME Crates. Each crate contains about 18 readout cards, whose outputs are combined on a Data Concentrator Card, with real-time synchronization and error-checking and a throughput of 200 Mbytes/s. The implementation is modular and based on industry and CERN standards: PCI bus, PCI-MIP and PMC carrier boards, S-Link and LVDS serial links. A prototype system including front-end emulator, HTR cards and Data Concentrator has been prototyped and tested. A VME motherboard provides a standard platform for the data concentrator. Implementation details and current status are described.

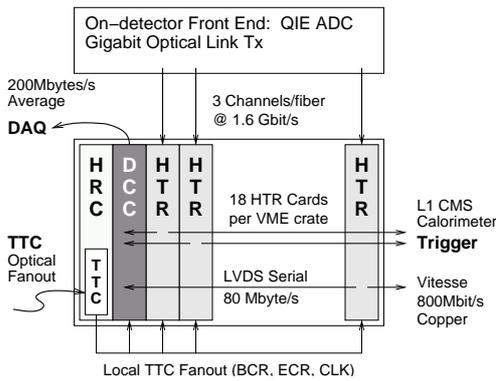


Figure 1: HCAL DAQ Crate

## 1 OVERVIEW

The CMS HCAL trigger/DAQ system consists of about (26) 9U VME64xP crates (Fig. 1) with up to 18 HCAL Trigger Readout (HTR) modules, one Data Concentrator Card (DCC), and one HCAL Readout Controller (HRC). Front-end data is carried from the on-detector front-end electronics to the crate by 100m optical fibers, each carrying 3 front-end channels. LVDS data links are used to transfer data from the HTR modules to the DCC and for local fanout of TTC (Trigger, Timing, Control) signals. The primary DAQ output is via an S-Link/64[1] carrying an average data volume of 200 Mbytes/s from each crate.

## 2 HCAL TRIGGER READOUT CARD

The HTR module is a 9U VME module (Figure 2) equipped with optical receivers, TTCrx circuitry, outputs on serial

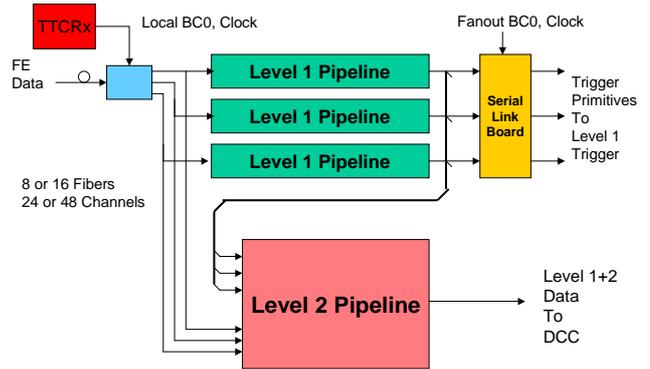


Figure 2: HCAL Trigger Readout Card

LVDS (Channel Link) and a custom mezzanine card. The optical inputs receive data from the HCAL front-end electronics, with one charge sample per bunch crossing (BX). The high-speed serial inputs require special board layout techniques. The CMS HCAL is a trigger detector, thus the HTR includes two data pipelines: the trigger pipeline, which assigns Front-End data to a BX and sends them to the CMS regional trigger, and the DAQ pipeline where the FE-data are pipelined, triggered and sent to the Data Concentrator Card.

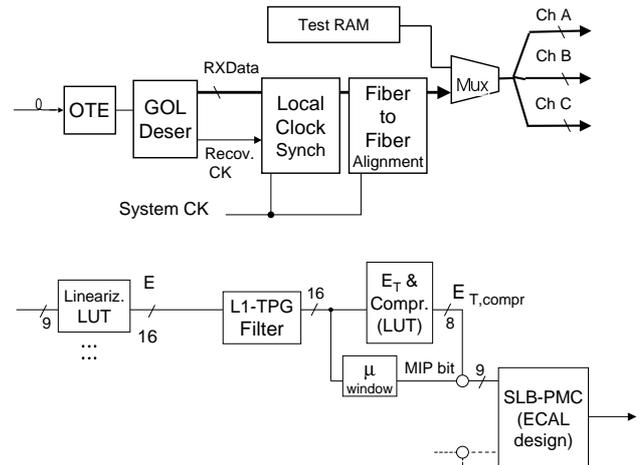


Figure 3: HTR Input and Level 1 Pipeline

The HTR input processing and Level 1 Pipeline is shown in Figure 3. The raw fiber data stream is deserialized, then synchronized to the local clock. A programmable delay of up to a few clocks is used to align data from different input fibers. A test RAM can substitute for the input data stream. Finally, the 3 channels carried on one fiber are demulti-

plexed. Each channel is then fed to a linearizing look-up table which converts raw input data to a 16-bit linear energy value. Next a finite-impulse response (FIR) filter is used to subtract the pedestal and assign all the energy to a single bunch crossing. This performs the same function as a traditional analog shaper, but has the advantage of being easily reprogrammable. Finally, the energy is converted to  $E_T$  and compressed to 8 bits according to a non-linear transformation specified by the CMS level 1 calorimeter trigger, and a comparison is done to see if the signal may represent a muon. This compressed output plus a muon ID bit is sent to level 1. The final synchronization and serial transmission is performed by a Synchronization and Link Board (SLB) described in detail elsewhere[2]. The latency of the level 1 pipeline is critical; it must be less than  $\approx 23$  BX periods. Currently the theoretical minimum for the HTR implementation is 16 BX periods.

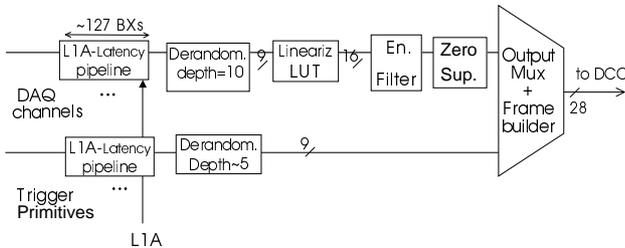


Figure 4: HTR Level 2 (DAQ) Pipeline

The HTR Level 2 Pipeline is shown in Figure 4. First is a pipeline of programmable depth which stores data during the CMS level 1 latency period (a fixed value). Then comes a “derandomizer” buffer into which data is copied at each level 1 accept. The derandomizer can hold up to 10 charge samples (one per BX) per event although currently we anticipate only processing 5 samples. Note that a given charge sample can in principle participate in multiple events, so the pipeline-to-derandomizer copy logic must handle overlapping events. From the derandomizer, data is linearized by a LUT, filtered by an FIR filter similar to that in the level 1 pipeline, and a threshold is applied for zero-suppression. At this point either the output of the filter, the raw data or both may be inserted into the output data stream.

A similar pipeline is used to store the level 1 trigger primitives, synchronized with the corresponding level 2 data. Finally the data is packaged in a variable-length block format along with any error information from the input links and transmitted using an LVDS serializer to the data concentrator.

### 3 DATA CONCENTRATOR CARD

The Data Concentrator Card is composed of a VME motherboard, six LVDS link receiver boards and a PMC-type logic board. The motherboard is a VME64x 9Ux400mm single-slot module. The motherboard[3] (Fig. 5) supports VME access up to A64/D32, and contains three bridged

PCI busses. Six PC-MIP[4] mezzanine sites are arranged in groups of three on two 33MHz 32-bit PCI busses. A third 33MHz 64-bit PCI bus is bridged to the VME bus using a Tundra Universe II VME-to-PCI bridge.

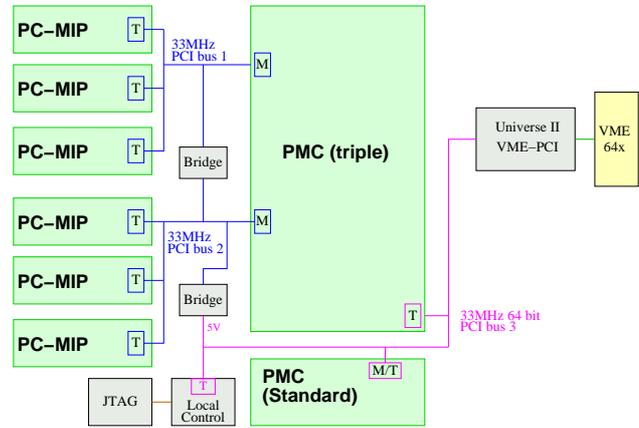


Figure 5: VME 9U Motherboard

A single large logic mezzanine board has access to all three PCI busses for high-speed application-specific processing, and an additional standard PMC site is available. A local control FPGA on the motherboard provides access to on-board flash configuration memory, a programmable multi-frequency clock generator, and JTAG.

The LVDS link receiver boards[5] (Fig. 6 use Channel Link[6] technology from National Semiconductor. Each board contains three independent link receivers which can operate at 20–66MHz (16-bit words). Buffering for 128K 32-bit words is provided for each link with provision to discard data if buffer occupancy exceeds a programmable threshold. Event building, protocol checking, event number checking and bit error correction are performed independently for each link. A PCI target interface provides single-word and burst access to the data stream, plus numerous monitoring registers. A single PCI burst read serves to build an event from fragments found in each of the three input buffers. The expected event number (low eight bits) is provided as part of the PCI address, and a mis-match causes an error bit to be set in the link trailer.

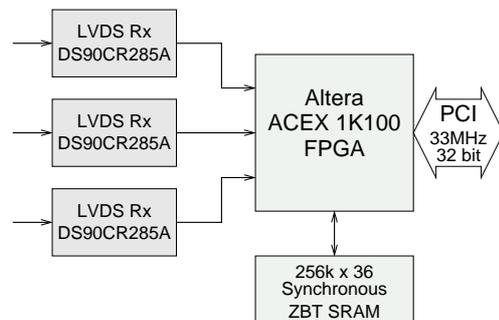


Figure 6: PC-MIP 3-Channel Link Receiver Board

The logic mezzanine board (Fig. 8) contains the core

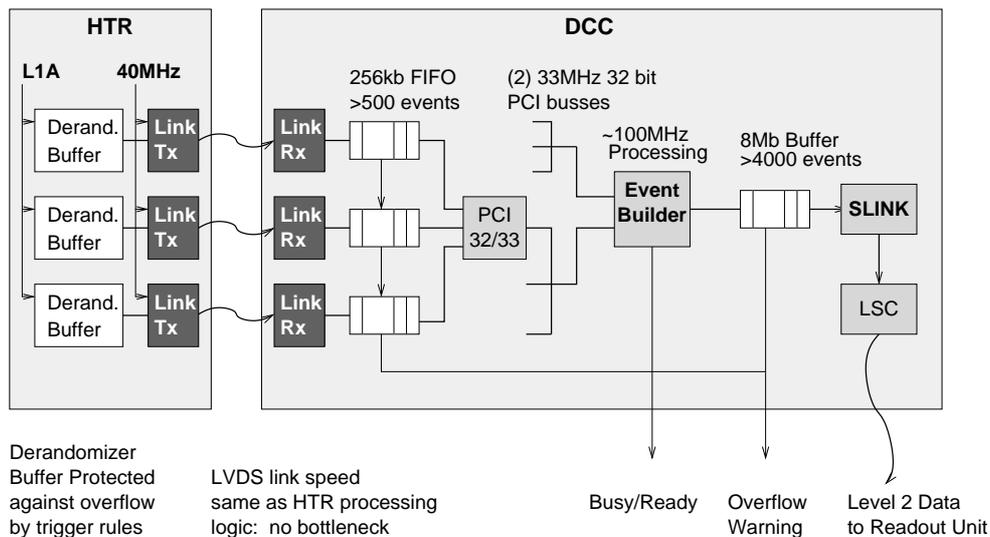


Figure 7: HCAL DAQ Buffering

data concentrator logic. The prototype was implemented using a Xilinx XC2V1000 for the logic, plus three Altera EP1K30 for three PCI bus interfaces.

The event builder logic merges two data streams from the two PCI busses, and re-orders the incoming data so that the various sub-types (Level 1, Level 2...) are in contiguous blocks in the output stream. An on-board TTCrx stores level 1 accepts (L1A) into a FIFO which drives the event builder. For each L1A, the *data decoder* triggers a PCI burst read on the PCI-1 and PCI-2 interfaces simultaneously. As data is transferred it is sorted into various sub-types and summary and monitoring information is collected. Each sub-type is pushed into a unique FIFO. After the end of the event has been processed (block trailer received from LRB) an end-of-event marker is pushed into each of the FIFOs. The *event builder* reads data from each of the sub-type FIFOs in turn, inserting protocol words as needed. The DCC logic is designed to operate continuously at the full speed of the two input PCI busses, namely  $33\text{MHz} \times 32 \text{ bits} \times 2$ . The event builder and output logic must thus run at an average rate of at least  $66\text{MHz}$  (32-bit words) or 264MBytes/sec.

The event builder output is sent in parallel to several destinations. Each output path contains a filter which can be programmed to select specific portions of events or a specific subset of events (prescaled, specially marked, etc.).

1. The DAQ Output. Every event is sent via SLINK-64 to the CMS DAQ. The detailed contents of each event may be controlled by configuration registers.
2. The Trigger Data Output. The trigger primitives sent to the CMS L1 trigger are also sent to via SLINK-64 to a special "trigger DAQ" system for monitoring of the trigger performance.
3. The Spy Output. A selected subset of events is sent to

a VME-accessible memory for monitoring and diagnostics.

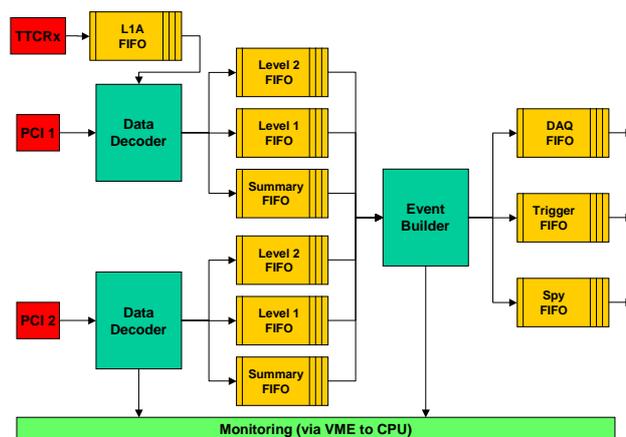


Figure 8: DCC Logic PMC

Error detection and recovery are a primary consideration in a large synchronous system and the DCC contains logic dedicated to this purpose. Figure 7 shows the main DAQ data pipeline and buffering in the HCAL readout system. Hamming error correction is used for the LVDS links between the HTR and DCC. All single-bit errors are corrected and all double-bit errors are detected by this technique. Event synchronization is checked by means of an event number in the header and trailer of each event, which are checked by the LRB logic against the TTC event number. Buffer overflow is avoided by the expedient of discarding the data payload and retaining only header and trailer words when the LRB buffer occupancy exceeds a programmable level. Additionally, an "overflow warning" output is provided which is delivered to the CMS trigger throttling system to request a reduction in the rate of L1A. Data transfers from the LRB to DCC logic are protected by

parity checks on the PCI busses. The event builder operates at a processing speed sufficient to handle 100% occupancy of the two PCI busses. After the event builder is a large memory, which can contain several thousand average-size event.

The main bottleneck (speed limitation) in the DCC is the two 32/33 PCI busses through which all data must flow. The theoretical maximum bandwidth for one of these busses is 33MHz x 4 or 132 Mbytes/s per bus. In practice we expect to achieve about 100Mbytes/s, for a total of 200Mbytes/s throughput on the two busses. This is exactly the maximum average data volume permitted on one input of the CMS DAQ switch.

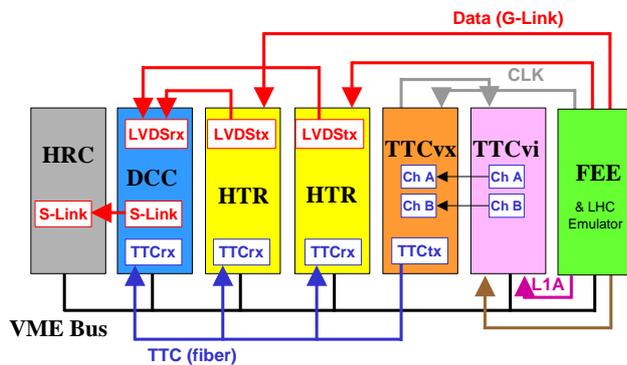


Figure 9: HCAL Readout Demonstrator

## 4 PROTOTYPE TESTING

A “demonstrator” (first prototype) of the entire system is being built (see Figure 9). The HTR demonstrator is a 6U VME module with 4 G-Link receivers running at 800Mbyte/s and an Altera APEX family FPGA for the processing logic. The (second) prototype and production HTR modules will be 9Ux400mm VME modules using CERN GOL links. The DCC demonstrator is built on the 9U VME motherboard as described above, and is quite close in hardware configuration to the anticipated production design. A custom front-end emulator (FEE) which simulates LHC timing and produces dummy front-end data is used to provide simulated input data to the HTR for testing. A G-Link based optical S-Link is used to transport data from the DCC demonstrator to a VME CPU for verification.

As of this writing, a simplified demonstrator using one FEE, one HTR, one DCC and S-Link to CPU has been successfully tested for use in a high-rate radioactive source test at Fermilab. Data was transferred through the entire chain without error at a continuous rate of 80 Mbytes/s. The S-Link data is received on the CPU in a large DMA buffer (400+ Mb) and when full written to disk for off-line analysis.

We expect to complete the full demonstrator shortly, though only highly simplified FPGA code will be implemented in the HTR and DCC.

## 5 SUMMARY

A demonstrator of the CMS HCAL DAQ has been assembled and testing has begun. The data concentrator makes extensive use of standard interfaces and busses, and was assembled from “multifunction” components developed separately. This resulted in significant savings by sharing development costs between multiple projects. The design of the full-fuction prototypes will continue through the remainder of 2001, with a working prototype system expected in 2002.

## 6 REFERENCES

- [1] “The S-LINK 64 bit extension specification: S-LINK64”, A. Racz, R. McLaren, E. van der Bij, EP Division, CERN. see <http://hsi.web.cern.ch/HSI/s-link/>
- [2] See <http://cmsdoc.cern.ch/carlos/SLB/>.
- [3] See [http://ohm.bu.edu/%7Ehazen/my\\_d0/mb9u/](http://ohm.bu.edu/%7Ehazen/my_d0/mb9u/)
- [4] “PC\*MIP Specification”, VITA 29. VMEbus International Trade Association Standards Organization
- [5] See [http://ohm.bu.edu/%7Ehazen/my\\_d0/TxRx/](http://ohm.bu.edu/%7Ehazen/my_d0/TxRx/)
- [6] The National Semiconductor family of LVDS point-to-point serial links. See for example the transmitter data sheet at: <http://www.national.com/pf/DS/DS90CR285.html>