FED-Kit design for CMS DAQ system

E.Cano[1], A Csilling[1], S. Cittolin[1], S. Erhan[2], B. Giacomo[1], D. Gigi[1], F. Glege[1], J. Gutleber[1], C. Jacobs[1], F. Meijers[1], E. Meschi[1], S. Murray[1], A. Oh[1], L. Orsini[1], L. Pollet[1], A. Racz[1], D. Samyn[1], P. Scharff-Hansen[1], C. Schwick[1] , P. Sphicas[1,3]

[1]CERN, CH1211 Geneva 23, Switzerland
[2] University of California Los Angeles, USA
[3]University of Athens, Athens Greece
Email:dominique.gigi@cern.ch
URL: http://cmsdoc.cern.ch/cms/TRIDAS/html/Documents.html

Abstract

For the development of the CMS Data Acquisition (DAQ) system several test benches have been built. They are based on three hardware modules referred to as the FED kit. This article describes the architecture, applications and performance of these modules.

The first module is a generic PCI card for moving data in and out of a PC containing a FPGA that can be programmed to customise the operation of the card. It is shown that the custom designed PCI interface transfers data at 500 MB/s, which is close to the theoretical bandwidth of 528 MB/s for the 66 MHz/ 64 bits PCI bus.

The other two modules form a data transfer link based on LVDS technology with a maximal data throughput of 800 MB/s and a cable length of 7.5 meters.

1. INTRODUCTION

The CMS DAQ system has been designed to collect event fragments from approximately 650 data sources at a first level trigger rate of 100 kHz. Fragments are assembled into entire events by the Event Builders and transferred to the Filter farm where higher-level trigger decisions are made by software. Accepted events are permanently stored for further physics analysis.

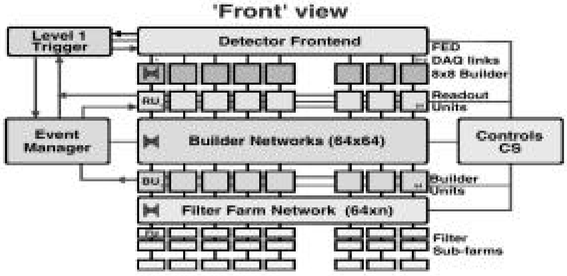

Figure 1: CMS DAQ block diagram

With a mean event size of 1MB the system has to sustain a data throughput of 100 GB/s. This can be achieved with the parallel architecture shown in figure 1.

The first stage Builder Network consisting of 64 8x8 builders collecting and distributing data to eight-64x64 Builder Networks of which one is shown in figure 1. The complete event data is collected in the Builder Units (BU's) and distributed to the Filter Farm for analysis.

The first stage builder network is described in [6].
The basic building blocks of the first stage of the DAQ are:

- An interface to the detector: the Front End Driver (FED)
- An interface to the 8x8 Builder Network: the Front End Readout Link (FRL)

Both components are prototyped with a common hardware platform, the Generic III board, and will be discussed in detail in the following sections.

2. Hardware elements

Three boards were developed to prototype the link between the FED and the FRL. These are the Generic III, the CMC (common mezzanine card) transmitter and CMC receiver boards. The two last boards compose the LVDS link.

2.1 The Generic III card

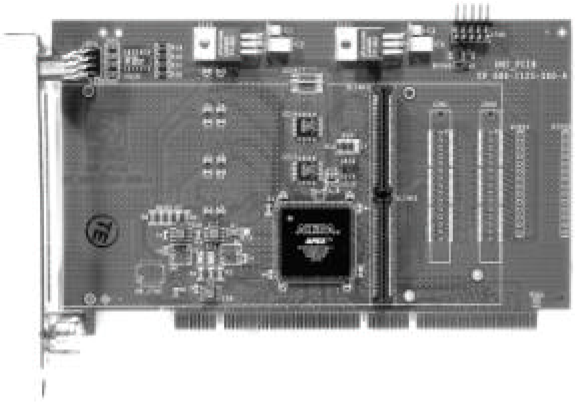

Figure2: Generic III PCI card

In the CMS DAQ development project the Generic III is used for multiple applications, the three majors ones are: FED emulator, FRL emulator and FED-readout kit. Other applications developed on the Generic III board shall not be described here include: the TTC distributor interface, RUI emulator…

The board is built around a single FPGA (APEX20K from ALTERA [4]). The user has a choice between two FPGA capacities (200K or 400K usable gates).

The FPGA is the core of the board and acts as a switch between four IO ports (see figure 3):

- PCI interface for 32-bit or 64-bit data width, up to 66 MHz
- 32MB SDRAM with 64-bit data width, up to 133 MHz
- IO's routed to 4 connectors
- 1MB flash memory

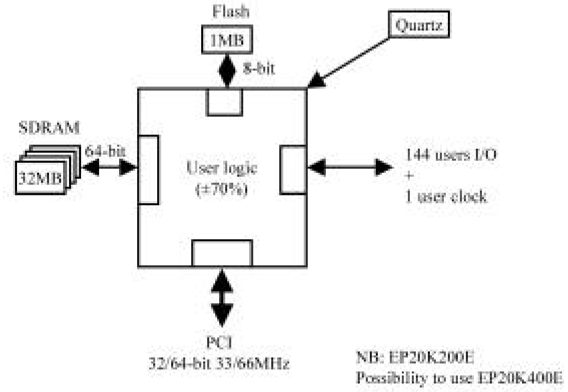

Figure 3: Generic III block diagra m.

The three first ports are synchronous. They can use the same clock (PCI frequency) or a quartz mounted on the board (for SDRAM and connectors)

The Flash memory is a one MByte of asynchronous memory accessed by an 8-bit data bus.

The IO's are routed to two kinds of connectors:

- S-Link64 [2,3] connectors
- High speed connectors (130 IO's from the FPGA)

The S-Link64 connectors are used with the CMC cards detailed below. The function of IO's going to the high-speed connectors is defined by the firmware loaded inside the FPGA.

2.2 S-Link64 implementation

The S-Link64 protocol is used to transfer data (event fragments) from the FED (or FED emulator described below) to the FED Readout kit or to the FRL. A cable with LVDS signals is used as media. This technology has the advantage to be simple and low price. The drawback of a cable instead of fiber is the limited length and rigidity of the cable.

To validate the cable length, extensive tests to measure the bit error rate were carried out. The successfully tested cable lengths are 2, 7.5, 10 and 15 meters. The LVDS link using a 2-meter cable was tested for 2 months with a 6.4Gbit/s data rate. During this test 10^{15} bits were transferred without error.

A similar test is currently being performed for a cable of 15 meters (the maximum length to be used in the DAQ system).

2.2.1 The CMC transmitter card

The CMC transmitter converts the S-Link64 signals to LVDS format. And vice-versa for the few signals coming back (Back-pressure and some reserved signals).

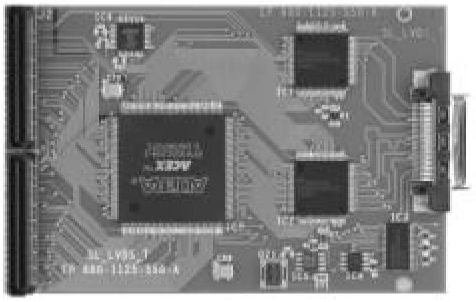

Figure 4: The LVDS CMC transmitter card

The frequency at the S-Link64 connector is up to 100 MHz (64-bit data width). The media used to transfer data is a cable of 18 twisted pairs that are shielded individually. The frequency used on the cable is 60 MHz for a cable length of up to 15 meters. An FPGA is used between the S-Link64 connectors and the LVDS transmitter to convert the transmitter frequency of the data.

It can also generate a test pattern like that specified in the S-Link64 specification [3].

This allows testing of the media and the link without any transfer of data coming from the FED.

2.2.2 The CMC receiver card

The CMC receiver card converts the LVDS signals to S_Link64 signals and in the opposite direction the backpressure and reserved signals to LVDS. It can receive up to two LVDS cables coming from two different FED's.

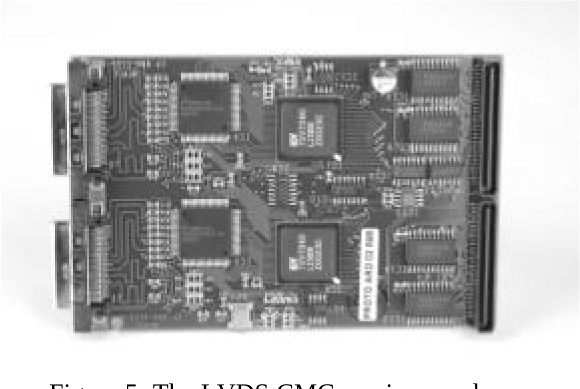

Figure 5: The LVDS CMC receiver card

For each link an onboard FIFO (32 Kbytes) buffers the incoming data (already converted to TTL levels by the LVDS receiver).

The FRL (see below) will control the buffers to concatenate up to two event fragments into one record.

3. Applications

The Generic III card as explained before, is a board where the FPGA firmware can be changed to address different applications. Described below are three of these applications that were developed to test and validate the functionality of the link between the FED and FRL.

3.1 FED emulator

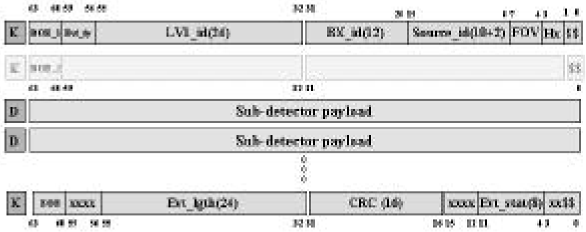

Figure 6: Event fragment format

The FED emulator is a tool used to generate event fragments in the case where there is no real FED available. There are three ways to generate data (according to the S-Link64 Specification [2][3] and the event data fragment format, see figure 6):

1. Event fragments can be generated on board with pseudo random data. The event fragment parameters; FED number, trigger number, size and seed (to generate pseudo random data), are transferred as three words through the PCI bus.
   These parameters are used to create the event fragment header and trailer.
   The data pseudo random generator is implemented with a formula that can be used to check the event fragment data received.

2. The event fragment can be located in memory outside of the Generic III. A software driver will write two words to the Generic III that specify the address of the fragment and its size. As soon as the Generic III has those parameters, it starts a DMA read operation and sends the data to the S-Link64 connectors. The first word will be interpreted as the header and the last as the trailer.

3. The last way is to write a fragment to the Generic III is through single or burst accesses PCI executed by an external DMA engine. The data will be directly sent to the S-Link64 connectors. A header and a trailer have to be written as the first and the last words respectively. An address is dedicated to distinguish between data and header/trailer. These PCI write accesses can be 32-bit or 64-bit words.

The three modes are implemented in the same FPGA firmware; the user distinguishes them through three different PCI base addresses.

3.2 FRL emulator

The FED Readout Link (FRL) interfaces up to two S-Link64 links to a commercial PCI optical link. If two S-Link64 data sources are used, the FRL concatenates the data fragments from the two links into one fragment.
In our FRL prototype, the Generic III sends data to a (network interface card) PCI Myrinet [5] NIC via a dedicated PCI bus.
The custom firmware of the NIC is programmed such that the memory is divided into pages (the page size is a programmable parameter). Each time that a page is available, the NIC sends its pointers to the FRL. When data are coming from the S-Link64, the FRL writes them (DMA, burst) to a free page. As soon as the page is filled up, the FRL uses a new one. A fragment always starts on a new page.
For each page used, a header is built at the beginning of the page by the FRL to indicate characteristics of the data. Five words are written:

- Size: the total number of bytes written inside the page.
- Page number: If a fragment is bigger than the page size, it will use multiple pages. For each, it will increase this number by one.
- Fragment number corresponding to the event number coming with the data
- Source number: the origin of the event fragment.

The NIC has to poll the size values inside this header, to know if the page is filled or not. As soon as the page is transmitted, the NIC frees the page by sending it to the FRL for future use.
The FRL also has to do multiple checks mainly in the case where the FRL has two functions: interface and merger. The main check is the time-out: if a FED source blocks, then the FRL has to ignore the S-link64 input after a time-out to avoid blocking the other FED with back-pressure.
The other important check is the synchronization between the two fragments to be merged. Their event numbers have to be the same.
This is not yet implemented in the current prototype.

3.3 FED Readout kit

The FED-readout kit architecture is similar to that of the FRL (see figure 7). The role of the NIC is performed by the PC running the FED-readout kit software.
At the beginning, the FED-readout kit software allocates memory blocks called ‘free memory blocks’. The free memory block addresses are in a ring buffer in the PC memory from which the software feeds a hardware FIFO.
As soon as data are coming from the S-Link64 or generated inside the Generic III (for the stand-alone mode), they are sent to memory block(s). The Generic III uses blocks as it needs, according to the block size. At the end of each event

fragment, the Generic III sends the fragment size to a software FIFO (FIFO word-count)(see figure 7).

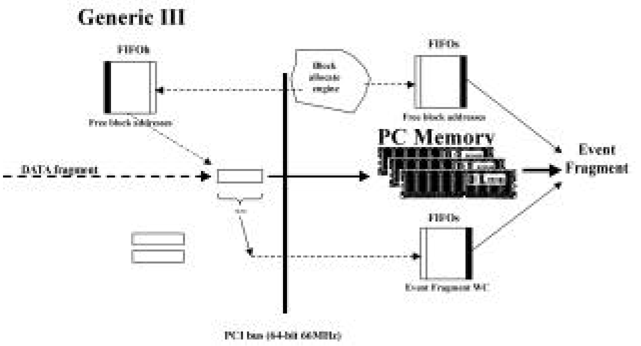

Figure 7: FED-kit block diagram

From the fragment size received in the word-count FIFO, the FED-Kit software knows the number of block(s) used for the fragment.

Figure 8 shows the performance reached by the FED readout kit with two different PC's for a block size set to 4kBytes.

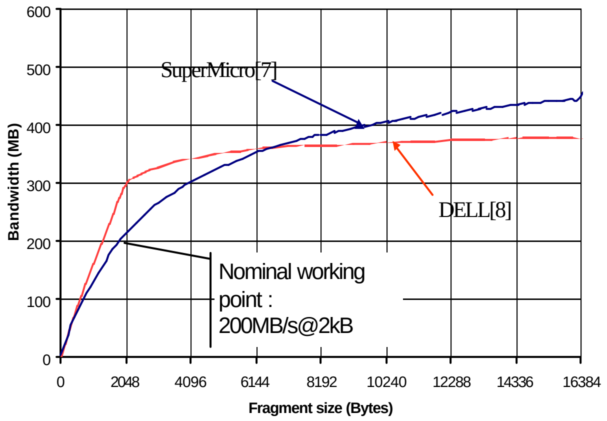

Figure 8: FED readout kit performances

4. CONCLUSIONS

Many applications were developed for the Generic III module:

- -FRL (FED Readout Link) that merges data from one or two FED's through an LVDS up to 480 Mbytes/s
- -FED readout kit: that emulates a FED. Data can be generated on board (random), read by DMA, written to the board by external DMA engine or test mode (data generated as mentioned in the S-Link64 specification).

Details about hardware and software of the FED-readout kit can be found at:
http://cmsdoc.cem.ch/cms/TRIDAS/html/Documents.html

5. REFERENCES

1. PCI local bus specification Revision 2.1, June 1 1995.
2. The S-Link Interface Specification, March 23 1997.
3. The S-Link64 bit extension specification: S-Link64, December 11 2000.
4. ALTERA component http://www.altera.com
5. Myrinet company http://www.myricom.com
6. CMS Data to surface transportation architecture (Attila Racz presentation B42).
7. 370DLE SuperMicro
8. DELL PC PowerEdge