Front-End / DAQ Interfaces in CMS

G. Antchev, E. Cano, S. Cittolin, S. Erhan, W. Funk, D. Gigi, F. Glege, P. Gras, J. Gutleber, C. Jacobs, F. Meijers, E. Meschi, L. Orsini, L. Pollet, A. Racz, D. Samyn, W. Schleifer, P. Sphicas, C. Schwick

CERN, Div. EP, Meyrin CH-1211 Geneva 23 Switzerland

Abstract

After reviewing the architecture and design of the CMS data acquisition system, the requirements on the front-end data links as well as the different possible topologies for merging data from the front-ends are presented. The DAQ link is a standard element for all CMS sub-detectors: its physical specification as well as the data format and transmission protocol are elaborated within the Readout Unit Working Group where all sub-detectors are represented. The current state of the link definition is described here. Finally, prototyping activities towards the final link as well as test/readout devices for Front-End designers and DAQ developers are described.

I. Introduction

In the case of CMS, there will be about 9 different detectors providing ~1 MB of data per trigger to the DAQ (see Fig. 1). Interfacing these sources with the DAQ is a critical point given the overall size and complexity of the final system (on-detector electronics, counting room electronics and DAQ).

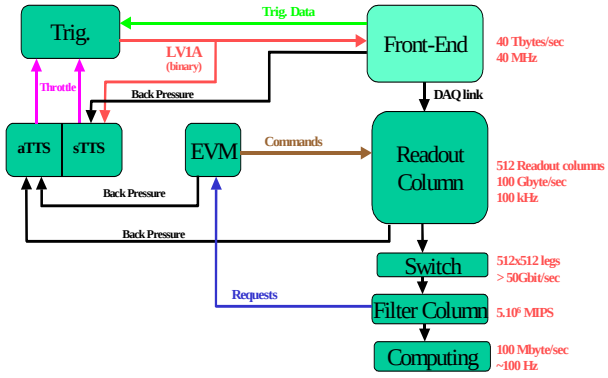

Fig. 1 CMS DAQ block diagram

The Front-End (FE) operation is synchronous with the machine clock and is located in the underground areas (detector cavern and counting rooms). The distribution of fast signals (LV1A, machine clock, resets, fast commands) is carried out by the Timing, Trigger Control system (TTC) [1]. The TTC provides to the FE the signals needed to signal the presence of data on every bunch crossing and send the trigger selected data to the Readout Column (RC). In turns, the FE can throttle back the trigger by providing fast binary status signals to the synchronous Trigger Trottling System (sTTS) [2].

The FE modules are read out by the RC (see Fig. 2) which is running asynchronously w.r.t. the machine clock, the RC being "trigger driven". For every event, the FE pushes its data as soon as possible through the data transportation devices towards the RC located at the surface. The event data are then buffered in the Readout Unit Input (RUI). The RC receives its control messages through the Event Manager (EVM). The EVM is sub-divided into a Readout Manager (RM) and a Builder Manager (BM). The RM enables the data integrity check in the RUI and the writing of the event fragment into the Readout Unit Memory (RUM). The BM enables the Readout Unit Output (RUO) to send an event fragment to a requesting Builder Unit (BU) sitting on the other side of the switch network. As for the FE, the DAQ can also throttle the trigger by means of messages provided to the asynchronous TTS (aTTS) through a control network.

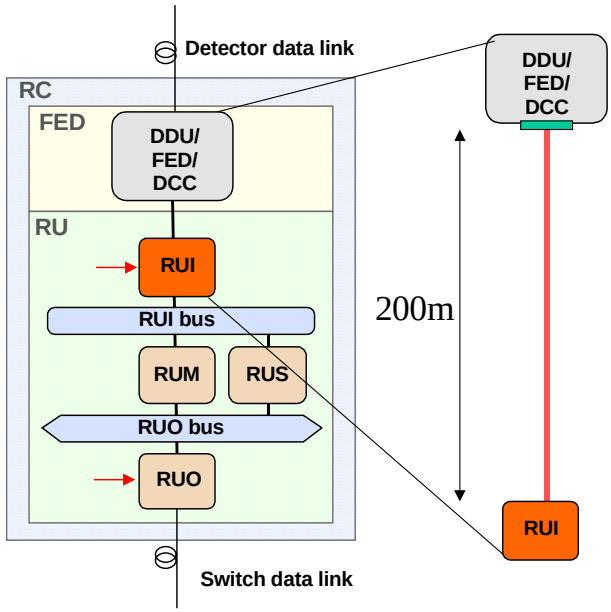

Fig. 2 CMS Readout column block diagram

II. Front-End Data Sources

As mentionned in the introduction, 9 sub-detectors will provide a total of 1 MB of data for every trigger. The central DAQ is designed to acquire this 1 MB of data at a maximum trigger rate of 100 kHz.
According to the most up-to-date information, the data are provided as follows:

- Pixel: 32 sources @ [850..2100] bytes
- Tracker: 442 sources @ [300..1500] bytes
- Preshower: 50 sources @ 2 kByte
- ECAL: 56 sources @ 2 kByte ± 10-20%
- HCAL: 24 sources @ 2 kByte
- Muon-DT: 60 sources @ ~170 bytes

- Muon-RPC: 5 sources @ ~300 bytes
- Muon-CSC: 36 sources @ ~120 bytes
- Trigger: 4 sources @ 1kByte

This makes a total of 709 sources with individual data sizes ranging from 120 bytes to ~ 2kByte. In order to use efficiently the nominal bandwidth of the DAQ hardware, a minimum packet size must be achieved by the front-end data sources (see Fig. 3). Given the current situation, the Pixel detector, the Tracker detector and the Muon detectors may need an additional concentration layer to match this requirement.

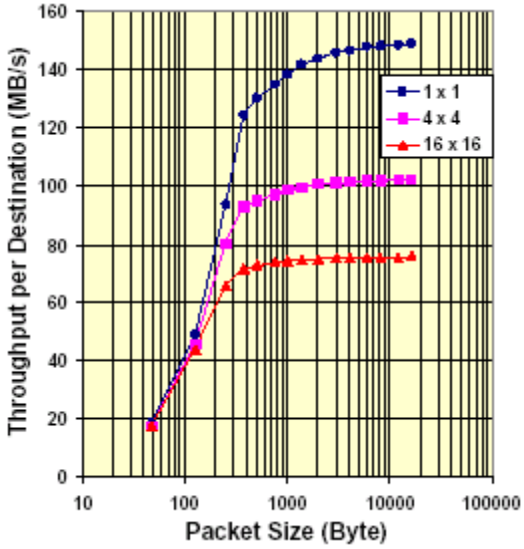

Fig. 3 Effective data throughput versus packet size

III. READOUT INTERFACE

The DAQ is the natural convergence point of the data produced by the sub-detectors. Reducing the diversity in the electronic devices is highly desirable if not outright necessary, in order to facilitate the system integration (especially during the initial debug phase) and also the maintenance operations. Therefore, the decision to use a common interface for all sub-detectors was made at a very early stage of the DAQ design. The interface is defined and elaborated within the Readout Unit Working Group (RUWG) [3] where all data providers and data consumers are represented. A common functional specification document [4] is adopted by all CMS data producers.

A. *Detector Dependent Unit*

The Detector Dependent Unit (also known as Front End Driver) is hosting the interface between the DAQ and the sub-detector readout systems. No sub-detector specific hardware is foreseen after the DDU in the readout chain. If the event size at the FED/DDU level is far from 2 KBytes, an intermediate Data Concentrator Card (DCC) merges several FEDs/DDUs in order to reach the 2 KB per event. This element is not needed for all sub-detectors. When a DCC is present in the sub-detector data flow, the DCC is seen by the DAQ as the interface between the DAQ and the sub-detector.

The task of the DDU is to deal with the specificities of each sub-detector and make available the data to the DAQ transportation hardware according to the specifications [4]. The specifications include the minimum functionalities to be performed by the DDU (header generation, alignment checking…) and the description of the DAQ slot which is located on the DDU where the DAQ transportation hardware is plugged.

B. *DAQ slot*

The DAQ slot is an S-LINK64 port [5]. S-LINK64 is based on S-LINK[1] [6] which has been extended to match CMS needs (64 bits @ 100 MHz). The extension is implemented through an additional connector, hence allowing the usage of standard S-LINK product until the availability of the final DAQ hardware.

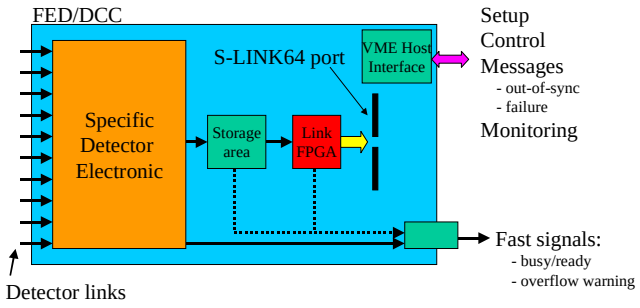

Fig. 4 DAQ slot on the FED/DCC

S-LINK as well as S-LINK64 specifies a set of 2 connectors (a sending and a receiving one) but not the physical link in between. The design and the implementation of the S-LINK64 port on the FED is under the responsibility of the sub-detector.

IV. DAQ DATA TRANPORTATION

A. *Data transportation requirements*

The required data throughput is 200 MB/s (2 KB @ 100 KHz) over a distance of 200m (cable path between the underground areas and the surface DAQ building). The data transportation hardware must be able to absorb stochastic fluctuations on the event size and provide enough contingency to cope with large uncertainties on the LHC luminosity and the detector occupancy/noise. However, the available bandwidth will clearly have an upper limit that cannot be exceeded. It is assumed that data sources in need of higher bandwidth will have some of their channels readin by an additional FED.

In order to have a good working efficiency, the event builder must receive a balanced traffic through its input ports and destination clashes must be avoided as much as possible. As shown in section II. on page 1, some detectors feature an important data size spread at the output of their data sources Therefore, the data transportation hardware must be able to average the traffic over several FEDs by appropriatly grouping FEDs with low and high data volumes per event.

Regarding the staging policy, at day 0, the trigger rate and the event size will be far from the nominal one: the full capacity of the DAQ will be needed only after LHC luminosity ramping-up and nominal CMS detector efficiency. A capacity of 25% of the nominal one is planned to be available on day 0, doubling after 6 months of data-taking to reach 100% after one year of operation. This staging strategy is also requested to match the expected funding profile. Therefore, the data trans-

1.Generic FIFO interface featuring 32 bits @ 40 MHz specified at CERN by R. McLaren and E. Van Der Bij

portation architecture must allow a progressive deployment of the DAQ.

B. *Data transportation architecture*

The data transportation architecture (see Fig. 5) is strongly driven by the event builder features and the staging strategy as well.

The constituting elements of the data transportation are:

- DAQ short reach link: transfers the FED data to the Front-End Readout Link card (FRL)
- Front-End Readout Link card: receives the FED data via the short link and houses the long reach DAQ link
- DAQ pit-PC: hosts the FRL and performs its configuration/control
- DAQ long reach link: moves the data from the FED/DCC into the intermediate data concentrator located at the surface (200m cable path).
- Intermediate data concentrator (FED Builder): implemented with an N x N crossbar switch. Each of the inputs is connected to a data source and depending on the deployment phase, up to N Readout Unit Inputs (RUI) are connected to the switch outputs. At LHC startup, only one RUI is connected and process the event fragments of N sources. Hence, by connecting hot and cold data sources, traffic balancing is performed *de facto* by the FEDB
  Whatever the deployment scenario is, the data transportation from pit to FEDB is never modified. Later, when higher bandwidths will need to be deployed, this will be achieved by connecting more RUIs in the system.

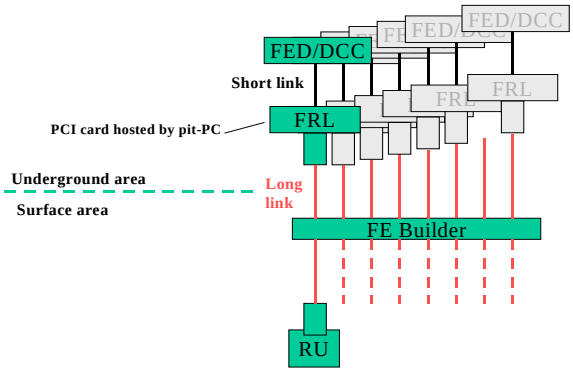

Fig. 5 Data transportation architecture

The technologies used to build the central DAQ system are clearly those used in the telecommunication world. Hence, both the performance and the cost of the system profit from the evolution of this dynamic market. By adopting popular telecom/computer standards (i.e. PCI or Infiniband), custom developments can co-exist with commercial products. Custom developments are at the time of this paper, the only way to achieve the required performances (200 MB/s sustained throughput through all the RC). As the performance of commercial products approach the requirements, these ones will be considered at procurement time or as replacement for the custom implementation. Therefore, the use of popular standards in custom design is a necessity given the most likely evolution and the upgrades.

V. PROTOTYPES

The prototyping phase will extend until Q4 2002 (DAQ Technical Design Report submission). At this time, implementation choices will be frozen and the pre-series production/procurement phase will start.

A. *Short reach link prototype*

The current prototype is based on LVDS technology:

- S-LINK64 compliant
- max. cable length: 10m
- max. throughtput: 869 MB/s
- BER < 10^{-15}

The sender card plugs into a FED and the receiver card is hosted by a multi-purpose PCI card (called Generic III or GIII). This forms the Hardware Readout Kit (HRK) provided to FED developers for laboratory work and beam test activities.

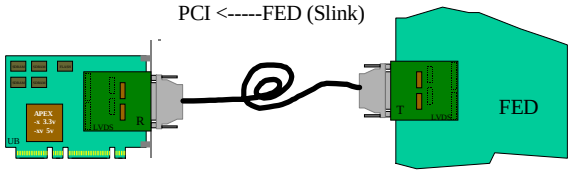

Fig. 6 Hardware Readout Kit usage

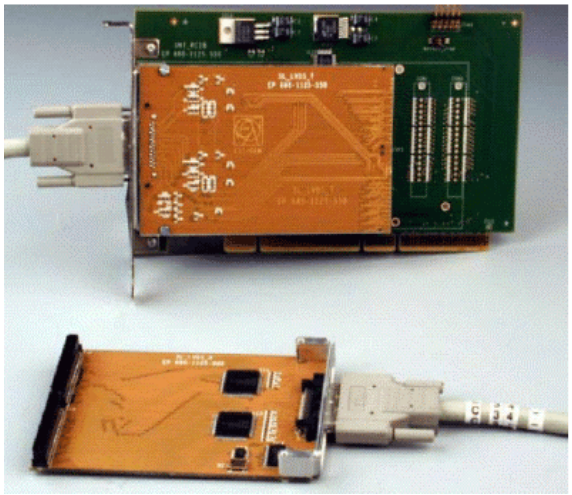

Fig. 7 Hardware Readout Kit picture

B. *FRL prototype*

The prototyping activities around the FRL are based on the third generation of a generic, multi-purpose PCI card (GIII). These cards have been developed within the CMS DAQ group for multiple applications [7][8][9].

The GIII features a single FPGA (200k or 400k logic gates equivalent) along with a 32 MB SDRAM memory, 1 MB flash RAM and two sets of connectors for extensions. One of these connector sets is compliant with the S-LINK64 pinout. The logic needed for interfacing the FPGA with the host PCI bus (64bit/66MHz), the memory devices and the connector sets leaves ~80% of the device free for user applications when the 200 kgate FPGA is soldered.

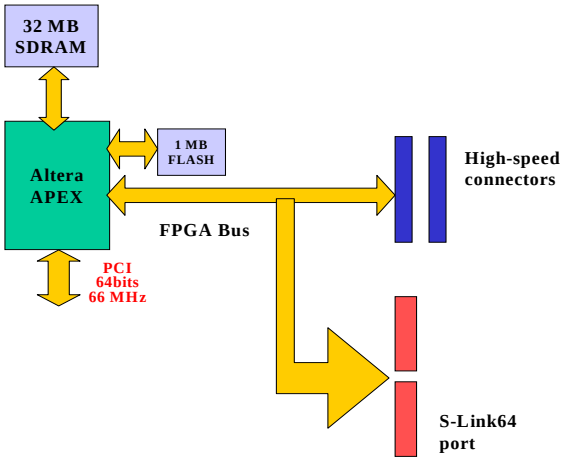

Fig. 8 Generic III block diagram

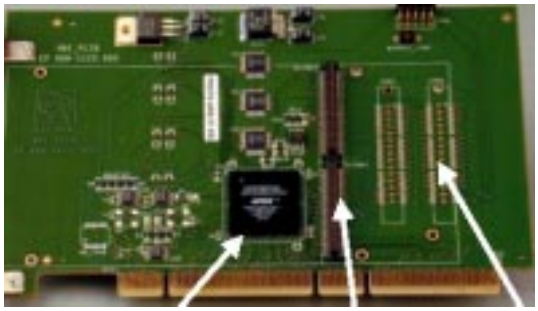

Fig. 9 Generic III picture

C. *Long reach link prototype*

As one end of this link is connected to the FED builder, its technology will be identical to the switch technology. Currently, Myrinet [10] is considered as the baseline technology for the FED Builder. The possible options for implementing such a link are the following:

- Off-the-shelf PMC hosted by the FRL card
- Myrinet protocol/core implemented in FPGA
- FRL with embedded Myrinet processor (Lanai-10)

These options are currently evaluated and discussed with Myricom.

The final decision will be taken for DAQ TDR submission (Q4 2002). Meanwhile, prototyping activities continue.

VI. INFRASTRUCTURES

A. *FRL housing*

As presented above, there is one FRL per data source and the FRL is a PCI card. Therefore, PCI slots must be available in the underground counting rooms. Using PCs for hosting the FRLs would require much more space than PCI cages. Such cages have 13 PCI slots and an interface with the control PC. Rack-mounted PCs with 7 free PCI slots (4U) allow to control 91 data sources within a standard 42U computer rack. A total of height racks is needed for the entire set of front-end data sources.

Fig. 10 A PCI cage with 13 slots

VII. CONCLUSION

An important fraction of the DAQ system (~90%) will be based on standard commercial components from the telecom and computing industries. The breathtaking improvments in speed, capacity and cost of these industries is well established and also expected to continue. Clearly, the benefits from delaying design choices have to be balanced against the constraints of providing readout capability to the Front End electronics which are already in production now.
The plan described in this paper addresses both of these constraints, by both providing hardware prototypes to the current developers and also delaying final technology choices upstream in the DAQ system.

VIII. REFERENCES

[1] http://ttc.web.cern.ch/TTC/intro.html

[2] http://cmsdoc.cern.ch/cms/TRIDAS/horizontal/docs/tts.pdf

[3] http://cmsdoc.cern.ch/cms/TRIDAS/horizontal/

[4] DDU design specifications A. Racz
CMS note 1999-010

[5] The S-LINK 64 bit extension specification: S-LINK64 A. Racz, R. McLaren, E. van der Bij

[6] The S-LINK Interface Specification
O. Boyle, R McLaren, E. van der Bij

[7] http://cmsdoc.cern.ch/~dgigi/uni_board.htm

[8] http://cmsdoc.cern.ch/cms/TRIDAS/horizontal/DDU/content.html

[9] Generic hardware for DAQ applications
G. Antchev et al.
LEB 1999 Proceedings

[10]http://www.myri.com