Development and testing of an upgrade to the CMS level-1 calorimeter trigger

Development and testing of an upgrade to the CMS level-1 calorimeter trigger

M. Baber,[a] **M. Blake,**[e] **J. Brooke,**[c] **M. Cepeda Hermida,**[e] **S. Dasu,**[e] **T. Durkin,**[d] **S. Fayer,**[a] **E.K. Friis,**[e] **T. Gorski,**[e] **G. Hall,**[a] **K. Harder,**[d] **G. Iles,**[a] **S. Ives,**[a] **J. Jones,**[b] **P.R. Klabbers,**[e] **A.G. Levine,**[e] **C. Lucas,**[c] **R. Lucas,**[a] **D. Newbold,**[c] **J. Marrouche,**[a] **S. Paramesvaran,**[c] **T.M. Perry,**[e] **A. Rose,**[a,1] **D. Sankey,**[d] **W. Smith,**[e] **A. Tapper,**[a] **A. Thea,**[d] **and T. Williams**[c]

[a]*Blackett Laboratory, Imperial College,*
Prince Consort Road, London, SW7 2BW, U.K.

[b]*Iceberg Technology,*
Helland Mill, Mabe Burnthouse, Penryn, Cornwall, TR10 9JQ, U.K.

[c]*H.H. Wills Physics Laboratory, University of Bristol,*
Tyndall Avenue, Bristol, BS8 1TL, U.K.

[d]*Rutherford Appleton Laboratory, STFC,*
Harwell, OX11 0QX, U.K.

[e]*Dept. of Physics, University of Wisconsin-Madison,*
1150 University Ave., Madison, WI 53706, U.S.A.

E-mail: awr01@imperial.ac.uk

ABSTRACT: When the LHC resumes operation in 2015, the higher centre-of-mass energy and high-luminosity conditions will require significantly more sophisticated algorithms to select interesting physics events within the readout bandwidth limitations. The planned upgrade to the CMS calorimeter trigger will achieve this goal by implementing a flexible system based on the μTCA standard, with modules based on Xilinx Virtex-7 FPGAs and up to 144 optical links running at speeds of 10 Gbps. The upgrade will improve the energy and position resolution of physics objects, enable much improved isolation criteria to be applied to electron and tau objects and facilitate pile-up subtraction to mitigate the effect of the increased number of interactions occurring in each bunch crossing. The design of the upgraded system is summarised with particular emphasis placed on the results of prototype testing and the experience gained which is of general application to the design of such systems.

---

[1]Corresponding author.

Contents

1 The CMS level-1 calorimeter trigger upgrade proposal

To cope with the high-luminosity conditions of the upgraded LHC, many subsystems of the CMS experiment require upgrading and others are using the LHC shutdowns as an opportunity simply to improve performance [1]. Beyond the obvious desire to improve the performance of the trigger, the CMS calorimeter trigger must be upgraded for two technical reasons: first, the existing Regional Calorimeter Trigger (RCT) is ASIC-based and is, therefore, limited in its ability to adapt to a luminosity scenario for which it was not designed. Secondly, the readout electronics for the CMS Hadronic Forward Calorimeter (HF) and Hadronic Calorimeter (HCAL) are being replaced and will include optical data-interconnects instead of the existing copper links.

The CMS trigger upgrade [1] is reliant on upgrades of both the ECAL and HCAL Trigger Primitive Generator (TPG) systems. The ECAL TPG system must be updated to provide an optical, rather than copper, output and the HCAL TPG system will be completely replaced as part of the stage-1 upgrade of the HCAL back-end electronics. The full trigger upgrade has a two-layer architecture: layer-1 consists of so-called "pre-processor" cards and layer-2 of "main processors" or "processing nodes". The roles of the two layers are discussed in section 4.

On the current schedule, the new HCAL electronics will not be ready for operation in 2015 and so an interim trigger for use in the 2015 run is being designed. This will involve duplicating the output of the existing RCT and sending it to both the new and existing trigger hardware. Any of the upgraded TPG systems which are available at that time, such as the new HF, would be used as additional inputs.

The architectures of both the full trigger upgrade and the interim trigger for 2015 are shown in figure 1.

2 Hardware

Three boards, shown in figure 2, are required to modify the existing trigger: the oSLB (Optical Synchronization and Link Board), the oRM (Optical Receiver Mezzanine) and the oRSC (Optical
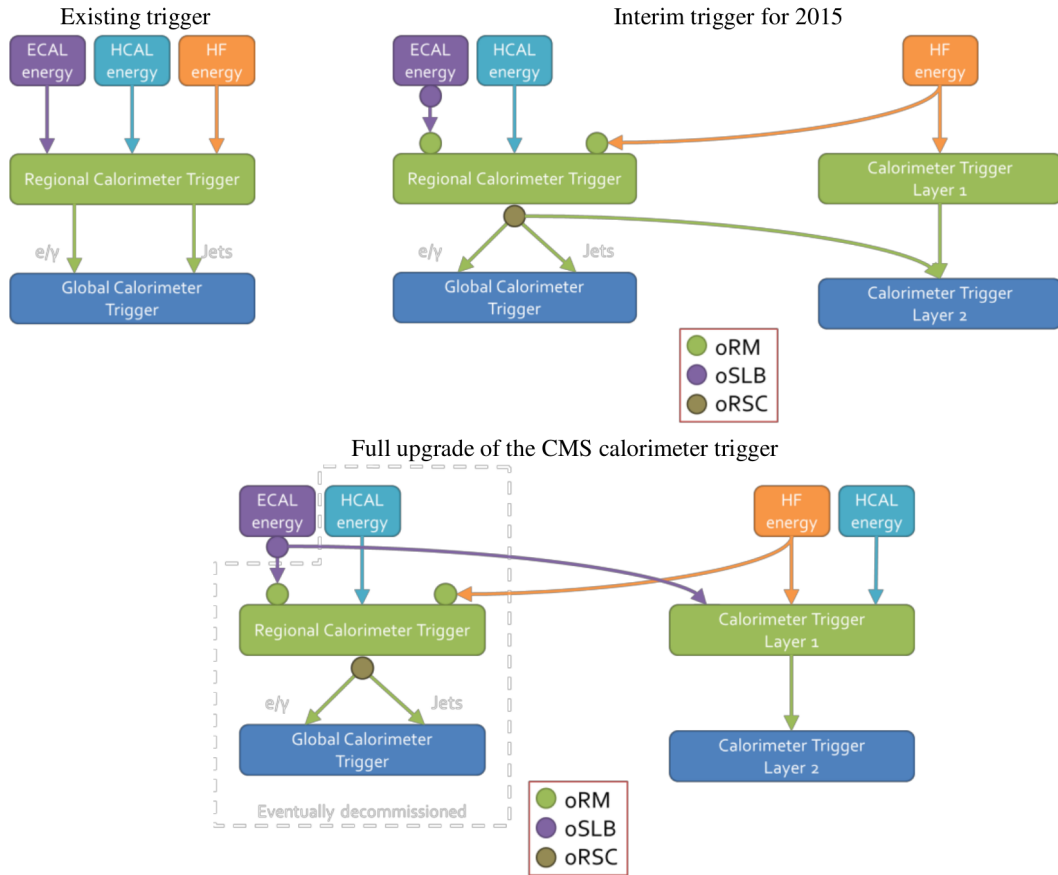
Figure 1. Top-left shows the existing trigger architecture, top-right shows the interim trigger for 2015, and the bottom figure shows the full trigger upgrade as specified in the Trigger upgrade TDR. In the full upgrade, the old and new trigger systems are fully decoupled and operable in parallel. The old trigger will be decommissioned in long shutdown 2. The meaning of "Layer 1" and "Layer 2" in the new calorimeter trigger are discussed in section 4.

Regional Summary Card). The oSLB and oRM are mezzanine cards, designed as direct replacements for the existing SLB and RM cards, but with optical outputs replacing the existing copper links. The oSLB allows the existing ECAL Trigger Primitive Generator (TPG) logic to remain intact and be used as a source for the new trigger, whilst continuing to feed the existing trigger. The oRM is required purely to receive the new optical links from the oSLB. The oRSC is a 9U-VME card, based around a Xilinx Kintex-7 FPGA [2] and is a replacement for the existing GCT source-cards. The oRSC provides two outputs: an output exactly matching the existing GCT to ensure continuity of the trigger, and a higher data-rate output which can be sent to the new trigger. The purpose of the oRSC is purely to provide an interim solution for the 2015 run, since the HCAL upgrade, which is needed as an input for the full trigger upgrade, will not be ready at that point.

The baseline hardware for the layer-1 of the new calorimeter trigger is the CTP7 (formerly Calorimeter Trigger Processor, Virtex-7) card by the University of Wisconsin. This card is based on the Xilinx XC7VX690T FPGA [2], uses a Xilinx Zynq processor for service tasks, and an initial prototype is now in manufacture.
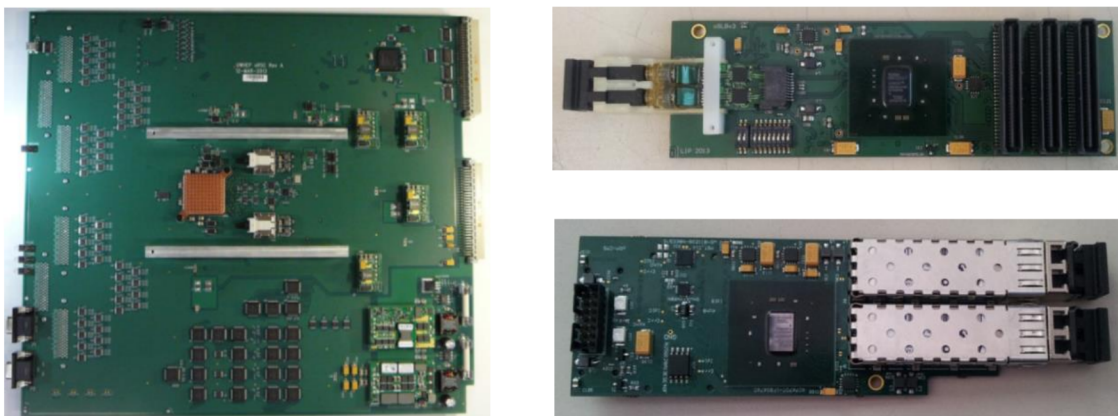
Figure 2. The oRSC (left), oSLB (top-right) and oRM (bottom-right). The oSLB and oRM are direct replacements for existing mezzanine cards, providing optical links, rather than copper links. The oRSC is a 9U VME card and sits in the existing RCT crates, one card per crate. The oRSC is a replacement for the existing GCT source-cards, with the capability of sending data to both the existing and the new trigger system.
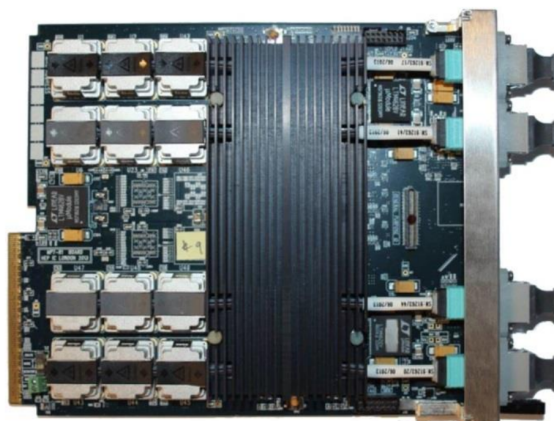

Figure 3. The Imperial MP7 processor. The MP7 is based on the Xilinx Virtex-7 FPGA (under the custom-designed heatsink) and Avago MiniPOD optical modules. The MP7 provides $72 + 72$ optical serial links capable of operating at over $10\,\mathrm{Gbps}$ giving a total bandwidth of $0.75 + 0.75\,\mathrm{Tbps}$.

The baseline hardware for the layer-2 of the new calorimeter trigger is the MP7 (Master Processor, Virtex-7) processor card by Imperial College, London [3, 4] (see figure 3). The MP7 is a $0.75\,\mathrm{Tbps} + 0.75\,\mathrm{Tbps}$ all-optical signal processor based around a Xilinx XC7VX485T or XC7VX690T FPGA [2] and Avago MiniPOD optics [5]. The MP7 also offers many advanced features such as an on-board firmware repository and a full self-test on start-up. This card has been under continuous testing for over a year and in all areas of testing has performed to our expectations. Having had the board under test for such a long period of time and in a very large number of different configurations also makes us confident that the board is thoroughly understood and that it is sufficiently flexible to adapt to unforeseen scenarios. Extensive software and firmware libraries have also been written for the board, based on the CMS-standard uHAL hardware access library and the IPbus control bus firmware [6].
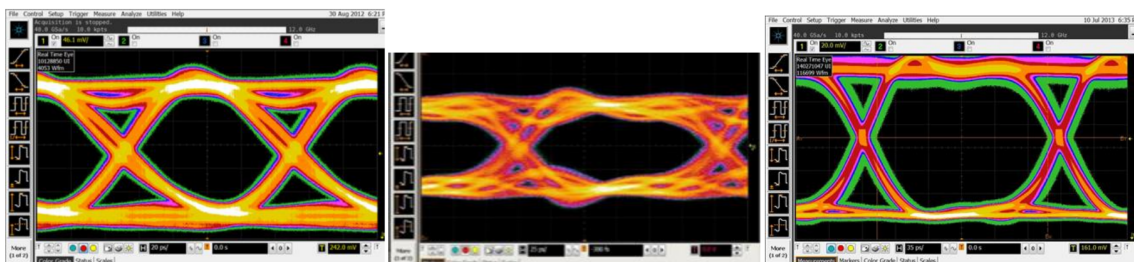
Figure 4. Eye-diagrams for the MP7 operating at 10 Gbps (left), the oRSC operating at 6.4 Gbps (centre) and the oSLB operating at 4.8 Gbps (right).

3 Testing

The most striking feature of the trigger upgrade proposal is the number of optical links and the speed at which those links are run; there are 864 optical links operating at around 10 Gbps between layer-1 and layer-2, and almost twice this number operating at around 4.8 Gbps or 6.4 Gbps, dependent on subsystem, from the TPGs to layer-1. Because of the dependence on these high-speed optical links, strong emphasis has been placed on their testing and qualification. In particular, it has been emphasized that stringent tests require the links to operate between independent boards, rather than in loopback; that scaling effects must be considered, both in terms of the number of links operating simultaneously and in the speed at which the links are run; and that the differences in the multi-gigabit transceivers between different generations of FPGA must be taken into account.

Testing of the serial links can be broadly divided into three categories: the quality of the eye-diagram on the optical link, the quality of the eye-diagram as seen by the transceiver and the bit-error rate on the line. For all boards in the system, the optical eye-diagram has been measured, a selection of which are shown in figure 4, and all boards meet industry standards.

A new feature in Xilinx 7-series FPGAs is the ability to generate eye-diagrams inside the transceivers themselves, based on a statistical analysis of data being sampled at different points across the UI. This feature has been used extensively in the testing of the MP7 board, where operating 72 links simultaneously makes the concurrent measurement of all optical eye-diagrams unfeasible. As stated previously, the performance at different line rates does not trivially scale and must be demonstrated: a selection of different LHC-asynchronous and LHC-synchronous line-rates have been tested in the MP7, with operation meeting telecoms-industry quality criteria at all speeds (see figure 5).

To validate the bit-error rate, all 72 outputs on one MP7 processor card were connected to the 72 inputs on a second card and the system left passing pseudo-random data for 16 days, transferring, in total, one exabit of data without bit-error. Spread over multiple tests, each link has now passed $O(10^{16})$ bits and performance is consistent over all the boards so far produced.

An integration facility has been set up at CERN for systematic testing of the interfaces between different subsystems. The facility furthermore allows the testing of the hardware using the same infrastructure as will be deployed in the final system, such as the Trigger, Timing and Control (TTC) system, computing and networking infrastructure, etc. Because of the tight schedule for the trigger upgrade, a set of tests were defined to prove all the technology required for the final system.
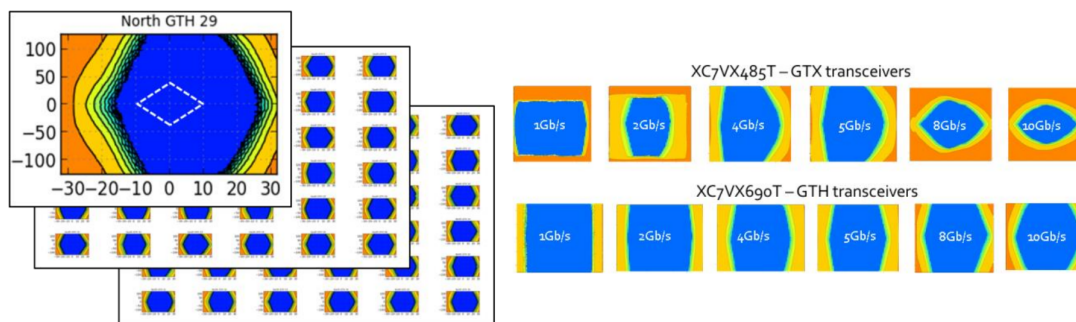
Figure 5. Eye diagrams generated in the Virtex-7 FPGA of the MP7 processor board generated using the Xilinx eye-scan tool. Left: 72 simultaneous eye-diagrams (created using PRBS-7 at 10 Gbps with transceivers in Low-Power Mode) and one diagram enlarged for clarity with an industry-standard eye mask shown. Right: eye-diagrams generated at different line-rates for the two variants of the Virtex-7 FPGA which are compatible with the MP7 board. The vertical scale of the eye-scan is not constant with line rate. All of the diagrams shown meet industry standards.

The first set of tests was set for July 2013 and was intended to demonstrate link quality between different board types, using the criteria of eye-diagram quality, bit-error rate using pseudo-random data, the transfer of data packets and the alignment of data with the LHC clock. Because the CTP-7 card is not yet available, the CTP-6 card was used in its place.

The number of tests scheduled was ambitious and around half of the goals were achieved in the time available. Eight board types and eleven different board-board links were tested:

- Eight of the eleven link types successfully demonstrated an industry compliant eye-diagram.

- Six, out of nine board-board links for which the test was meaningful, were able to generate an internal eye-diagram and evaluate the bit error-rates on pseudorandom data.

- Five of the eleven link types successfully demonstrated the transmission of data patterns.

- One test, MP7 to MP7, successfully demonstrated alignment of data with the LHC clock.

The integration program is continuing to address the full spectrum of tests.

4 System architecture and firmware

For the upgrade of the calorimeter trigger, CMS has chosen what is called a Spatially-Pipelined, Time-Multiplexed Trigger (TMT) architecture: this is, in essence, a fixed-latency event builder architecture. The TMT consists of two layers: the first layer (the "pre-processors") receives data from every bunch-crossing (BX) but for only a small region of the detector. The pre-processors buffer all the data from a bunch-crossing and retransmit it to a single board in layer two (the "main-processors"), over N bunch-crossings, where $N \leq 10$. The data from the next event is similarly buffered and transmitted, again over N bunch-crossings, but to a different main-processor card. This is repeated in a round-robin fashion over M main-processor cards, where $M \geq N$, ensuring that there is no dead time. Because both the volume of incoming data and the algorithm latency are

fixed, the position of all data within the system is fully deterministic and no complex scheduling mechanism is required. Such an architecture is new to CMS, and its choice was motivated by its implications for firmware design:

The calorimeter trigger receives data from the ECAL and HCAL in the form of towers, where each tower is approximately 0.087×0.087 in $\eta \times \varphi$. The geometry of the calorimeters, then, is such that the towers form a 2-dimensional (2D) array, consisting of 72 towers in φ by 56 in η in the barrel + endcap region, with additional towers from the HF, whose geometry will change with the upgrades to the on-detector electronics. The job of the calorimeter trigger is, in essence, to identify 2D features within this 2D dataset. The TMT architecture allows all data to be rearranged such that it arrives in geometric order (Spatially-Pipelined): that is, towers at given φ always occupy the same bits on the same optical links and all towers arrive in order of increasing $|\eta|$. By so doing, the 2D geometric problem is reduced into a 1D problem in space and a 1D problem in time, allowing all algorithms to be fully pipelined (i.e. at the full data-rate) rather than the LHC crossing rate as is done currently. This offers several key advantages: the processing within the FPGA is localised, reducing the size and number of fan-outs, minimising routing delays and eliminating register duplication, which is why engineers and computer-scientists recommend full-pipelining as the preferred design methodology for FPGAs [7].

Because the data is buffered and retransmitted over N bunch-crossings, a latency penalty of N bunch-crossings is incurred in changing from parallel to Time-Multiplexed format. This penalty is, however, a one-off: since the algorithms are fully-pipelined, processing begins as soon as the minimum quantity of data is received, rather than waiting for all data to arrive and processing it all simultaneously. Similarly, since all candidates along one dimension are formed simultaneously and sequentially in the other dimension, candidates in one dimension may be transmitted between systems as soon as they are available, rather than waiting for the formation of all candidates. Furthermore, because of the buffering and full-pipelining, algorithms may be run up to six-times faster than in the current trigger, compensating for both the Time-Multiplexing latency penalty and the latency incurred by using serial, rather than parallel, links.

A second integration test was carried out in September 2013, whose goal was to demonstrate a slice of the TMT architecture, including data-transfer, alignment, a set of baseline algorithms and integration with CERN infrastructure. The test architecture consisted of two MP7 cards acting as pre-processors, one MP7 card acting as main-processor and a fourth MP7 card demonstrating realignment of the data. The test successfully met, and in fact exceeded, all of the required criteria, giving us further confidence in both the MP7 card and in the Time-Multiplexing principle. A selection of "physics" validation plots are shown in figure 6.

As stated previously, the Spatially-Pipelined, Time-Multiplexed architecture was chosen since it allows full pipelining of the firmware, thereby localising the processing, reducing the size and number of fan-outs, minimising routing delays and eliminating register duplication. A major result of the integration test was also to demonstrate that full pipelining was, indeed, essential to the building of the firmware, with even small fan-outs being able to cause failures in the meeting of timing constraints.

Another lesson learned was the importance of "floor-planning" (constraining which areas of the chip are used for each task) on these exceptionally large FPGAs (see figure 7). This process was shown to have a very large impact on algorithm performance and offered significant improvements
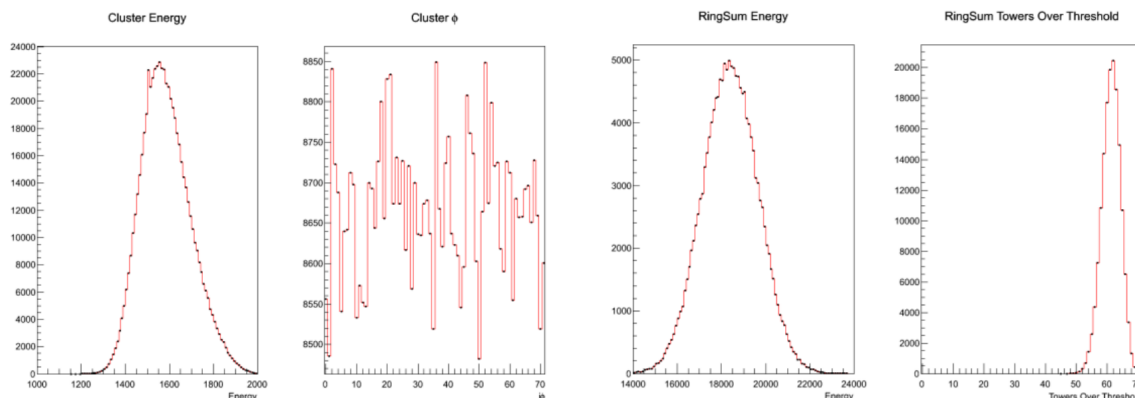
Figure 6. Example "physics" validation plots showing hardware output (red line) against software emulator (black dots). Precise agreement between hardware and software is demonstrated.
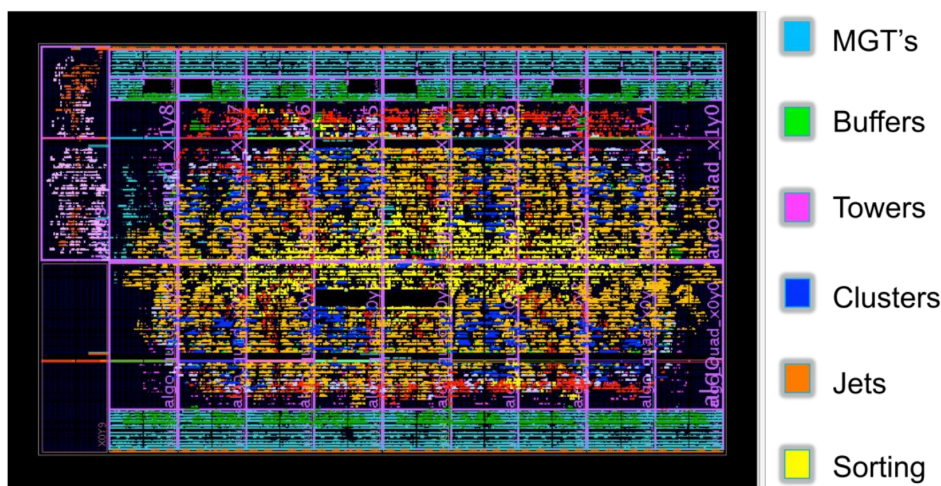

Figure 7. Floor-plan of the Xilinx XC7VX690T FPGA used in the September Integration test of the Time-Multiplexed Trigger architecture. By using the Xilinx floor-planning tools, the logic relating to the multi-gigabit transceivers (MGTs) and DAQ/Debug buffers (turquoise and green above) were restricted to the edges of the chip (near the MGTs) and logic utilization in these regions pushed to $\sim 90\%$. This process then left a large region in the centre of the chip for algorithms (red, pink, blue, orange and yellow above). The two regions at the left-hand end of the chip in the above diagram are reserved for communications, control and DAQ. The above design uses 35% of the available LUTs.

in meeting timing constraints. Experience showed this to be particularly important, since after a 6-hour build time, the build could fail because of a single net (of the ~ 6 million nets in the design) failing to meet timing constraints. Although the process of floor-planning is generally applicable, it is considerably easier if signals remain relatively local and, as such, was another vindication of the fully pipelined approach, and thus also of the choice of a Spatially-Pipelined, Time-Multiplexed Trigger architecture.

The integration test also measured the latency and it was shown that the latency at each stage matched that expected, in particular, demonstrating the one-off nature of the Time-Multiplexing penalty.

Throughout the integration test, the CMS-standard uHAL hardware access library and the IP-bus control bus were successfully used, making both debugging and multi-user crate access trivial. The AMC13 card [8], which is to be used for trigger and DAQ throughout CMS, was also found to work both well and reliably.

5 Conclusions

The upgrade of the CMS calorimeter trigger is making good progress, with most of the final hardware in hand, and the main processing card, the MP7, has been under continuous test for the past year. A baseline upgrade architecture has been specified in the TDR, and to optimise performance and build times for large FPGAs, that architecture is a Spatially-Pipelined, Time-Multiplexed Trigger. Focus has now moved to system-level integration, particularly on the interconnection of heterogeneous hardware and on demonstrating the capabilities of the Time-Multiplexed architecture.

Acknowledgments

References

[1] CMS collaboration, *CMS technical design report for the level-1 trigger upgrade*, CERN-LHCC-2013-011, CMS-TDR-12 (2013).

[2] Xilinx Incorporated, *7 Series FPGAs overview. Advance product specification* (2012), http://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf.

[3] A. Rose et al., *The MP7 and CTP-6: multi-hundred Gbps processing boards for calorimeter trigger upgrades at CMS*, 2012 *JINST* **7** C12024.

[4] Imperial College London, *MP7 homepage*, http://www.hep.ph.ic.ac.uk/mp7.

[5] Avago Technologies, *MiniPODTM AFBR-81uVxyZ 12-channel transmitter, AFBR-82uVxyZ 12-channel receiver high density, pluggable, embedded parallel-fiber-optics modules product brief* (2011), http://www.avagotech.com/pages/minipod_micropod.

[6] *CACTUS: Code Archive for CMS Trigger UpgradeS*, https://svnweb.cern.ch/trac/cactus/wiki.

[7] S.J.E. Wilton, S.S. Ang and W. Luk, *The impact of pipelining on energy per operation*, in Proceedings of *Field-Programmable Gate Arrays, Field Programmable Logic and Application: 14th International Conference (FPL 2004)*, Leuven Belgium, 30 Aug–1 Sep 2004, Springer [ISBN-10: 3540229892] [ISBN-13: 978-3540229896].

[8] Boston University, *The AMC13 project*, http://bucms.bu.edu/twiki/bin/view/BUCMSPublic/HcalDTC.