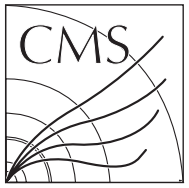
Available on CMS information server

CMS CR -2024/294



The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

29 October 2024 (v4, 01 November 2024)

Phase-2 CMS DAQ – Growing from prototype boards to demonstrator systems

Jeroen Hegeman for the CMS DAQ project

Abstract

For the Phase-2 upgrade of the CMS experiment, the CMS central DAQ group designed and developed two custom ATCA boards. These boards provide the interfaces between the sub-detector electronics and the central CMS systems. This paper describes our experience with the chosen prototyping strategy, with a focus on the design modification choices made along the way. It concludes with a brief overview of recent firmware developments, and a look at the transition towards the full board production.

Presented at *TWEPP2024 Topical Workshop on Electronics for Particle Physics*

Prepared for submission to JINST

Phase-2 CMS DAQ – Growing from prototype boards to demonstrator systems

Jaafar Alawieh,[a] **Miguel Durasov,**[a] **Ulf Behrens,**[b] **Andrea Bocci,**[a] **James Branson,**[c] **Philipp Brummer,**[a] **Jan Bugajski,**[a] **Sergio Cittolin,**[c] **Albert Corominas I Mariscot,**[a] **Georgiana-Lavinia Darlea,**[d] **Christian Deldicque,**[a] **Marc Dobson,**[a] **Antonin Dvorak,**[a] **Antra Gaile,**[a] **Dominique Gigi,**[a] **Frank Glege,**[a] **Guillelmo Gomez-Ceballos,**[d] **Patrycja Gorniak,**[a] **Magnus Hansen,**[a] **Jeroen Hegeman,**[a,1] **Thomas James,**[a] **Tejeswini Jayakumar,**[a] **Wassef Karimeh,**[a] **Dimitra Kostala,**[a] **Rafal Krawczyk,**[b] **Wei Li,**[b] **Kenneth Long,**[d] **Frans Meijers,**[a] **Emilio Meschi,**[a] **Srecko Morovic,**[c] **Babatunde Odetayo,**[a] **Luciano Orsini,**[a] **Christoph Paus,**[d] **Andrea Petrucci,**[c] **Marco Pieri,**[c] **Dinyar Rabady,**[a] **Attila Racz,**[a] **Theodoros Rizopoulos,**[a] **Hannes Sakulin,**[a] **Christoph Schwick,**[a] **Dainius Simelevicius,**[a] **Jan Troska,**[a] **Polyneikis Tzanis,**[a] **Cristina Vazquez Velez,**[a] **and Petr Zejdl**[a]

[a] *CERN, Switzerland*
[b] *Rice University, Houston, Texas, USA*
[c] *University of California, San Diego, San Diego, California, USA*
[d] *Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*

E-mail: jeroen.hegeman@cern.ch

Abstract: For the Phase-2 upgrade of the CMS experiment, the CMS central DAQ group designed and developed two custom ATCA boards. These boards provide the interfaces between the sub-detector electronics and the central CMS systems. This paper describes our experience with the chosen prototyping strategy, with a focus on the design modification choices made along the way. It concludes with a brief overview of recent firmware developments, and a look at the transition towards the full board production.

Keywords: Data acquisition concepts, Trigger concepts and systems

[1] Corresponding author.

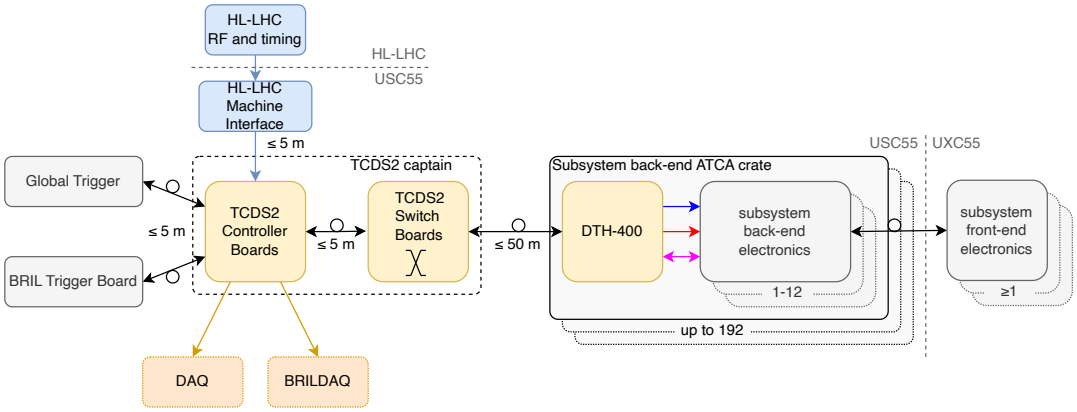


Figure 1. Architecture of the Phase-2 CMS Trigger and Timing Control and Distribution System (TCDS).

1 Introduction

The architecture of the central CMS [1, 2] trigger-DAQ system for the start of the HL-LHC era revolves around two full-custom ATCA boards [3]. These boards interface all sub-detector off-detector (or back-end) electronics to the central timing, trigger, and data acquisition (DAQ) systems.

The first board, the DAQ and Timing Hub, is a dual-FPGA ATCA hub board. One FPGA is dedicated to timing, fast control, and trigger distribution. The other FPGA provides data aggregation and protocol translation from a simple point-to-point 'SLinkRocket' protocol[1] to standard 100 Gigabit Ethernet. Based on its functionality, and on its maximum throughput of 400 Gbit/s, this board was dubbed the DTH400 (or DTH, for short). Given the wide spectrum of throughput requirements of subsystems [3], ranging from O(20 Gbit/s)/back-end crate for some of the muon detectors to O(2.2 Tbit/s)/back-end crate for the inner tracker, a second board was developed to serve as a DAQ 'throughput expander'. Implementing two DAQ units identical to the DAQ part of the DTH400, this 'DAQ800' board provides 800 Gbit/s of DAQ throughput. Each back-end crate contains one DTH400. In addition, one or more DAQ800s can be added to a crate.

2 The DTH400 and the DAQ800 as 'dual purpose' boards

In addition to the DAQ system itself, the CMS central DAQ project also encompasses the Trigger and Timing Control and Distribution System (TCDS). The DTH provides the interface between the back-end crates and the central part of the TCDS, the 'captain' (figure 1), which in turn interfaces to the CMS global trigger (GT) and beam monitoring and luminosity measurement (BRIL) systems, as well as to the HL-LHC RF. The captain will use DAQ800s to implement all control functions and to provide global connectivity, and DTH400s to interface to central CMS and LHC systems.

To reduce the number of specialised, single-purpose, boards to be developed, the DTH400 and DAQ800 were both designed to double as building blocks for the TCDS. To achieve this, the DTH400 is equipped with additional high-speed serial links connected to front-panel SFPs.

[1] The SLinkRocket is a 25 Gbit/s evolution of the S-LINK protocol [4] designed for the first generation LHC experiments.

In addition, the DTH provides several electrical clock and synchronisation inputs that can be adapted to accommodate the most common signal standards, once the HL-LHC RF distribution system is fully defined. Another, significant, implication of the 'dual purpose' nature of these designs is the presence of a dual clock tree to all FPGA transceivers. DAQ functionality operates (asynchronously) at standard Ethernet line rates, while all timing functionality uses bunch clock-synchronous transceiver reference clocks.

3 The DTH400 development and prototyping

The initial DTH design was conceived in early 2017. From its inception, the DTH prototyping strategy targeted a two-prong, two-step approach for the hardware development.

A first prototype (the P1) was to be developed to gain experience with the ATCA form factor and infrastructure, and to demonstrate the required baseline functionality. In parallel, a separate prototype aimed at studying the aspects related to the network connectivity required of the DTH. A managed Ethernet switch should provide a 1GbE connection between the experiment control network and each of the back-end boards. The choice for a managed switch was based on a desire to provide reliable access control and quality-of-service management, even in crates sharing boards of different designs and/or with different functionality.

As part of the separation of functionalities, and to ease the buffering requirements for back-end boards, the DTH provides all the buffering required to handle any fluctuations in the data-to-surface (D2S) and event builder networks. (I.e., the back-ends only buffer fluctuations in trigger rate and event size.) To simplify the DTH firmware, all data passes through the buffer memory (as opposed to buffering on-demand when a delay is detected). Based on throughput and D2S round-trip time, this requires a significant amount of buffer memory. This, in turn, requires a high memory throughput.

The DTH400-P1 was based on the Xilinx KU15P FPGA, with a Micron Hybrid Memory Cube (HMC) as DAQ data buffer, and using the Silicon Labs Si53xx family of jitter attenuators for clock generation and cleaning. Soon after the production of the first P1s, however, Micron retired its HMC. This triggered the design of a second version of the P1, the P1V2, which replaced the HMC with several banks of DDR4 RAM.

A batch of about 30 DTH400-P1V2s was produced and distributed among back-end developers. The P1V2 was also used to demonstrate that this design met all baseline timing and clock quality requirements for the CMS upgrade [5]. The DAQ part of the DTH was verified by aggregating event fragments from multiple back-end links into per-orbit 'super fragments', and transferring those over a 100GbE link to a commercial switch. The reliance on multiple, external, RAM banks to achieve the required throughput, however, was considered a potential weak point in the design. Further study led to a redesign of the DAQ unit based on the Xilinx VU35P FPGA, with built-in high-bandwidth memory (HBM). For simplicity (both in design and in price negotiations), the TCDS FPGA was changed to the same type. The result was the DTH400-P2.

Like most of the ATCA boards designed for the Phase-2 CMS upgrade, the DTH contains an on-board computer module for control and monitoring. Apart from the FPGA change, the P2 design introduced a Rear Transition Module (RTM), housing this on-board controller. Placing the controller on the RTM eases the placement and routing constraints on the front board, at the same time providing a potential controller upgrade path. The design of the RTM to hold an independent

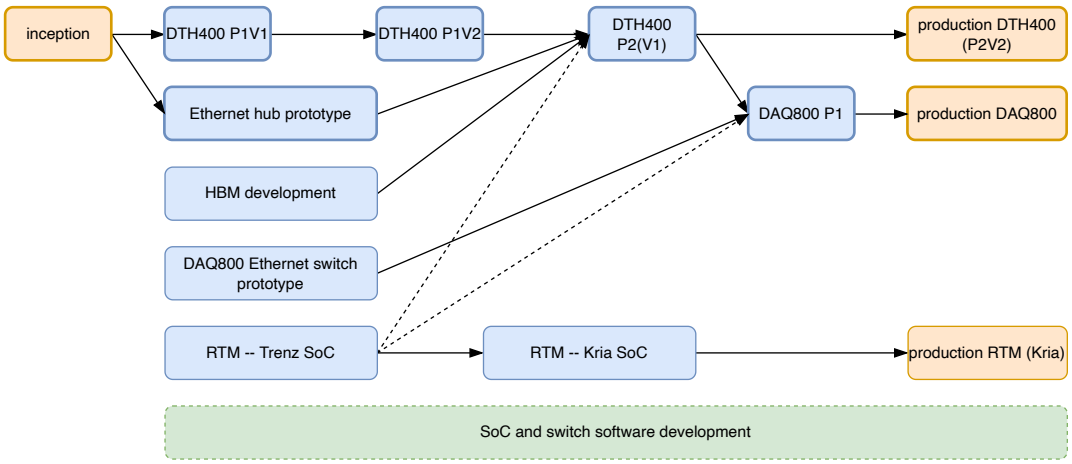


Figure 2. The 'branch and merge' prototyping strategy chosen for the DTH400 and the DAQ800.

functional part of the DTH made it possible to treat its development as a separate branch of the overall prototyping scheme.

The first version of the RTM was based on a Trenz MPSoC module, and was shown to meet all technical requirements. For reasons of availability, pricing, and commonality with other back-end boards, we decided to switch to the Xilinx Kria K26 System-on-Module (SOM). The similarity of the core components used on these two modules made this a minor hardware modification. Migrating the software stack from one board and vendor environment to another turned out to be more work this time.

Figure 2 depicts the overall development process followed for the DTH400 and the DAQ800. Our experience shows that the 'branch and merge' approach is a powerful tool in completing a complex design like an ATCA hub board.

4 Design convergence

After successful completion of all prototyping branches, the development lines were carefully recombined into the designs of the DTH400-P2 and the production RTM. The design of the DAQ800 is based on that of the DTH DAQ unit, implemented twice. Ethernet connectivity on the DAQ800 (to the controller, the IPMC, an RJ45 debug port, and the backplane) was developed in a separate branch demonstrating correct integration between all networking components, as well as the switch hardware configuration. (See also figure 2.)

Slightly reminiscent of the HMC, Xilinx recently announced the end-of-life of their UltraScale and UltraScale+ HBM FPGAs. This again shows the relative volatility of memory technologies compared to, for example, programmable logic components. For designs with longer development times, and for projects targeting multiple production batches separated in time, this emphasises the need to closely watch the component markets and be reactive, either by stocking components for future hardware production, or by adapting the design to follow the market.

Whereas the DAQ800 schematics to a large extent represent a copy of those of the DTH, this is not the case for the PCB layout. The bottom half of the DTH is clearly dominated by the Ethernet

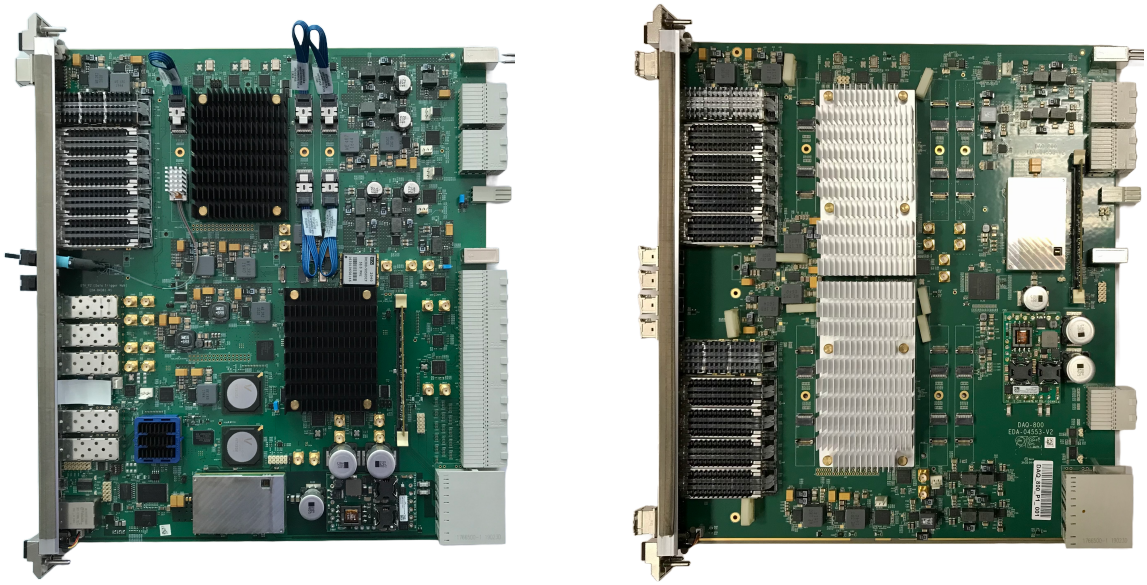


Figure 3. The DTH400-P2 (left) and the DAQ800 (right), without their RTMs.

and TCDS backplane connectivity. The DAQ800, on the other hand, distinctly shows the presence of the two identical DAQ units, located close to the front panel, with all board infrastructure grouped at the rear of the board. (See also figure 3.)

5 From prototype to pre-production, to full production

A few DTH400-P2 and DAQ800 prototype boards have been produced. Just enough to validate their designs. Based on their evaluation, a few final modifications were made to the placement of some parts, minimising any interference with heat sinks and other mechanical parts. A pre-production series of 40 P2s is under way at the time of writing, for final validation together with back-end developers. The plan is to follow this by the full DTH400 and DAQ800 production series next year.

6 Recent progress in DAQ and TCDS firmware

The validation batches of the DTH400-P2 and the DAQ800 have allowed us to expand our demonstrator systems, albeit not yet to truly representative system scales. It has been shown that both the DTH and the DAQ800 can operate multiple 100GbE links simultaneously at speeds close to the line rate. The limiting factor in those cases was always the receiving computer. Firmware monitoring indicates that sufficient idle clock cycles remain to push the data aggregation and framing logic to well beyond the output line rate.

Based on our own testing experience, the original semi-fixed grouping of back-end data links to Ethernet links [6] was redesigned to be more flexible, and partially reconfigurable without firmware modifications. Figure 4 shows the resulting mapping matrix. The marked and coloured cells indicate possible assignments of back-end link data to Ethernet links.

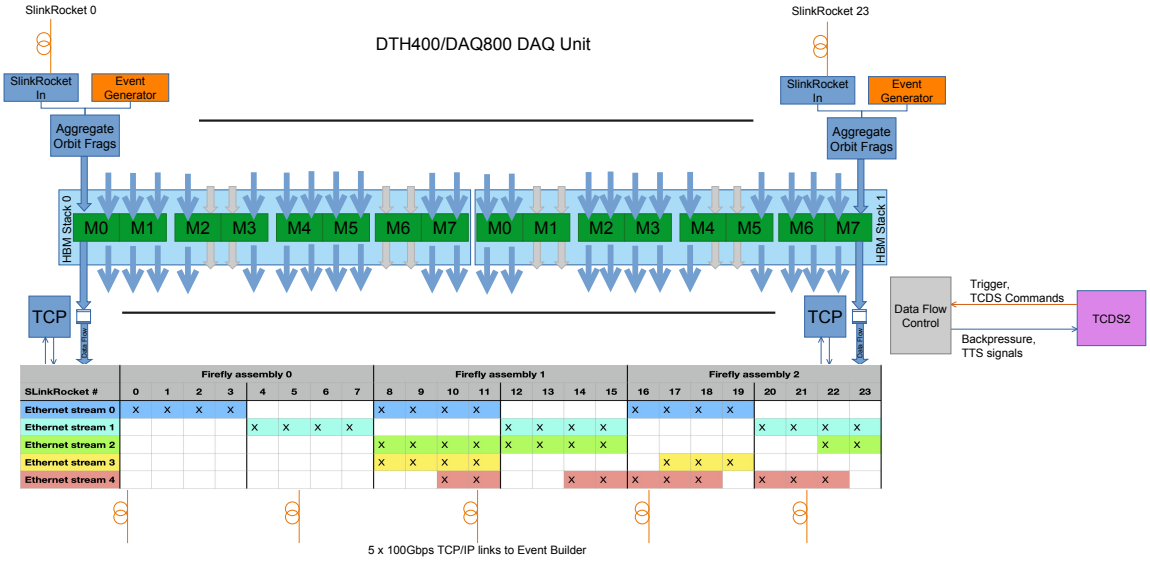


Figure 4. New, configurable, routing between SLinkRocket back-end links and 100GbE output links provides enhanced flexibility for sub-detectors presenting disparate throughput requirements.

On the TCDS side, all DAQ800 links have been validated for bunch clock-synchronous operation, and work has started on the implementation of the central timing and trigger control functionality on the validation hardware. The aim is to demonstrate the full-chain timing system before launching the full DAQ800 production.

7 Conclusion

After several years of development, final production of both the DTH400 and the DAQ800 is now imminent. The chosen prototyping strategy, branching developments into separate hardware lines and merging the results into the final design, has proven very effective.

The development of these boards, as well as that of other CMS back-end boards, is clearly subject to trends in technology and in the component markets. Flexibility in design, component purchasing, and production planning is therefore more than ever a requirement for successful completion of these complex hardware development projects.

Even though not all Phase-2 functionality has been demonstrated, with all features and at full scale, yet, we are confident these designs are mature, and the hardware well-tested and production-ready.

References

[1] CMS collaboration, *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, *JINST* **3** (2008) S08004.

[2] CMS collaboration, *Development of the CMS detector for the CERN LHC Run 3. Development of the CMS detector for the CERN LHC Run 3*, *JINST* **19** (2024) P05064 [2309.05466].

[3] CMS collaboration, *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger*, Tech. Rep. CERN-LHCC-2021-007, CMS-TDR-022, CERN, Geneva (2021).

[4] H. van der Bij, R. McLaren, O. Boyle and G. Rubin, *S-link, a data link interface specification for the lhc era*, *IEEE Transactions on Nuclear Science* **44** (1997) 398.

[5] J. Hegeman, R. Blažek, U. Behrens, J. Branson, P. Brummer, S. Cittolin et al., *First measurements with the CMS DAQ and Timing Hub prototype-1*, *PoS* **TWEPP2019** (2020) 111.

[6] V. Amoiridis, U. Behrens, A. Bocci, J. Branson, P.M. Brummer, S. Cittolin et al., *CMS Phase-2 DAQ and Timing Hub – Prototyping results and perspectives*, Tech. Rep. 05, CERN, Geneva (2022), DOI.