

## 40 MHz Level-1 Trigger Scouting for CMS

*Gilbert Badaro*<sup>8</sup>, *Ulf Behrens*<sup>6</sup>, *James Branson*<sup>2</sup>, *Philipp Brummer*<sup>1,c</sup>, *Sergio Cittolin*<sup>2</sup>, *Diego Da Silva-Gomes*<sup>3,b</sup>, *Georgiana-Lavinia Darlea*<sup>4</sup>, *Christian Deldicque*<sup>1</sup>, *Marc Dobson*<sup>1</sup>, *Nicolas Doualot*<sup>3,b</sup>, *Jonathan Richard Fulcher*<sup>1</sup>, *Dominique Gigi*<sup>1</sup>, *Maciej Gladki*<sup>1</sup>, *Frank Glege*<sup>1</sup>, *Dejan Golubovic*<sup>1</sup>, *Guillermo Gomez-Ceballos*<sup>4</sup>, *Jeroen Hegeman*<sup>1</sup>, *Thomas Owen James*<sup>1</sup>, *Wei Li*<sup>6</sup>, *Audrius Mecionis*<sup>3,a</sup>, *Frans Meijers*<sup>1</sup>, *Emilio Meschi*<sup>1</sup>, *Remigius K. Mommsen*<sup>3</sup>, *Keyshav Mor*<sup>1</sup>, *Srecko Morovic*<sup>2</sup>, *Luciano Orsini*<sup>1</sup>, *Ioannis Papakrivopoulos*<sup>5,b</sup>, *Christoph Paus*<sup>4</sup>, *Andrea Petrucci*<sup>2</sup>, *Marco Pieri*<sup>2</sup>, *Dinyar Rabady*<sup>1</sup>, *Kolyo Raychinov*<sup>1</sup>, *Attila Racz*<sup>1</sup>, *Alvaro Rodriguez-Garcia*<sup>1</sup>, *Hannes Sakulin*<sup>1</sup>, *Christoph Schwick*<sup>1</sup>, *Dainius Simelevicius*<sup>7,b</sup>, *Panagiotis Soursois*<sup>1</sup>, *Andre Stahl*<sup>6</sup>, *Mantas Stankevicius*<sup>3,a</sup>, *Uthayanath Suthakar*<sup>1</sup>, *Cristina Vazquez-Velez*<sup>1</sup>, *Awais Zahid*<sup>1</sup>, *Petr Zejdl*<sup>3,b</sup>

<sup>1</sup>CERN, Geneva, Switzerland

<sup>2</sup>University of California, San Diego, San Diego, California, USA

<sup>3</sup>FNAL, Chicago, Illinois, USA

<sup>4</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>5</sup>Technical University of Athens, Athens, Greece

<sup>6</sup>Rice University, Houston, Texas, USA

<sup>7</sup>Vilnius University, Vilnius, Lithuania

<sup>8</sup>American University of Beirut, Beirut, Lebanon

<sup>a</sup>Also at Vilnius University, Vilnius, Lithuania

<sup>b</sup>Also at CERN, Geneva, Switzerland

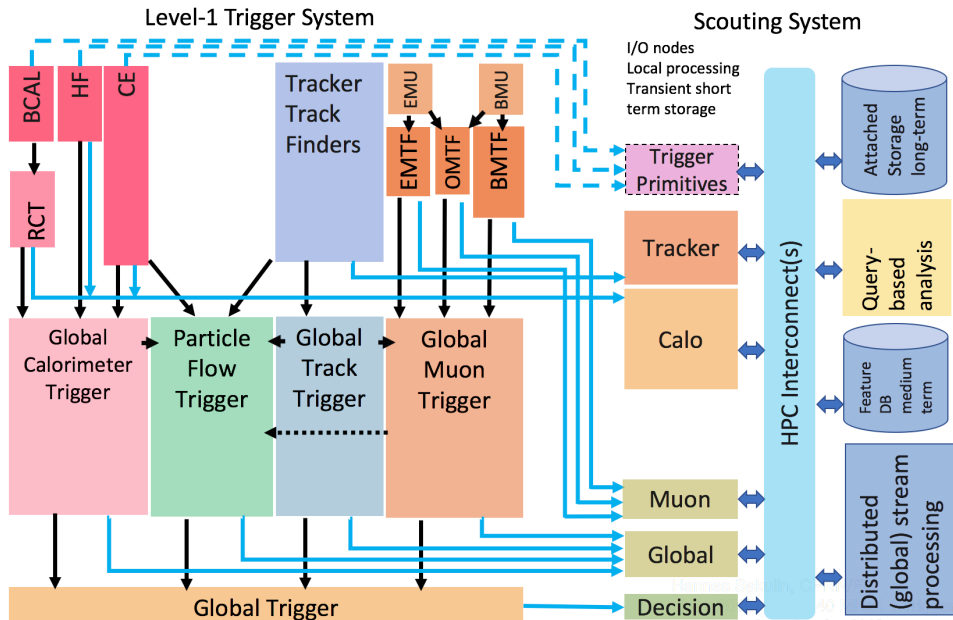
<sup>c</sup>Also at Karlsruhe Institute of Technology, Karlsruhe, Germany

**Abstract.** The CMS experiment will be upgraded for operation at the High-Luminosity LHC to maintain and extend its physics performance under extreme pileup conditions. Upgrades will include an entirely new tracking system, supplemented by a track finder processor providing tracks at Level-1, as well as a high-granularity calorimeter in the endcap region. New front-end and back-end electronics will also provide the Level-1 trigger with high-resolution information from the barrel calorimeter and the muon systems. The upgraded Level-1 processors, based on powerful FPGAs, will be able to carry out sophisticated feature searches with resolutions often similar to the offline ones, while keeping pileup effects under control. In this paper, we discuss the feasibility of a system capturing Level-1 intermediate data at the beam-crossing rate of 40 MHz and carrying out online analyzes based on these limited-resolution data. This 40 MHz scouting system would provide fast and virtually unlimited statistics for detector diagnostics, alternative luminosity measurements and, in some cases, calibrations. It has the potential to enable the study of otherwise inaccessible signatures, either too common to fit in the Level-1 accept budget, or with requirements which are orthogonal to “mainstream” physics, such as long-lived particles. We discuss the requirements and possible architecture of a 40 MHz scouting system, as well as some of the physics potential, and results from a demonstrator operated at the end of Run-2 using the Global Muon Trigger data from CMS. Plans for further demonstrators envisaged for Run-3 are also discussed.

# 1 Introduction

For its High-Luminosity phase of operation, currently planned to start in 2027, the Large Hadron Collider at CERN will be upgraded to provide an instantaneous luminosity of up to  $7.5 \times 10^{34} / \text{cm}^2 / \text{s}$  resulting in a pile-up of up to 200 inelastic collisions per event. In order to cope with these conditions, the Compact Muon Solenoid (CMS) experiment [1] will undergo a major upgrade including an entirely new tracking system with a track finder providing tracks at Level-1 [2], a new high-granularity calorimeter in the endcap region [3], and new readout electronics for the barrel calorimeter and muon systems providing finer granularity. With a total event size of 7.5 MB, reading out the entire detector at the 40 MHz bunch crossing rate would not be possible due to power density related limitations in some of the frontends, nor would it be economically feasible, even in 2027. CMS will therefore continue to employ two trigger levels: a Level-1 trigger based on state-of-the-art FPGAs, selecting events at 750 kHz and a high-level trigger running on a farm of heterogeneous compute nodes performing the second level of selection in software. The upgraded Level-1 trigger [4], shown in the left-hand part of Figure 1, will be able to evaluate sophisticated algorithms previously impossible at Level-1, such as vertex finding, particle flow with pile-up per particle identification (PUPPI), and a Kalman Filter resulting in object resolutions similar to offline.

While the two-stage trigger of CMS is designed to provide excellent physics performance, the study of some physics processes can potentially benefit from an analysis of the full available statistics at 40 MHz, albeit with the resolution of the Level-1 trigger. We present our ideas for a scouting system that would receive data from the Level-1 trigger over spare output links, and process them quasi-online in a heterogenous compute farm. This scouting system would run largely independently of the standard CMS Trigger/DAQ chain. Results found with the scouting system may in some cases be sufficient to gain new physics insights



**Fig. 1.** The CMS Level-1 trigger system for Phase-2 (left) and the proposed Scouting System (right). BCAL: Barrel Calorimeter; HF: Hadronic Calorimeter Forward; CE: Calorimeter Endcap (high granularity); RCT: Regional Calorimeter Trigger; EMU/BMU: Endcap/Barrel Muon System; EMTF/OMTF/BMTF: Endcap / Overlap / Barrel Muon Track Finder.

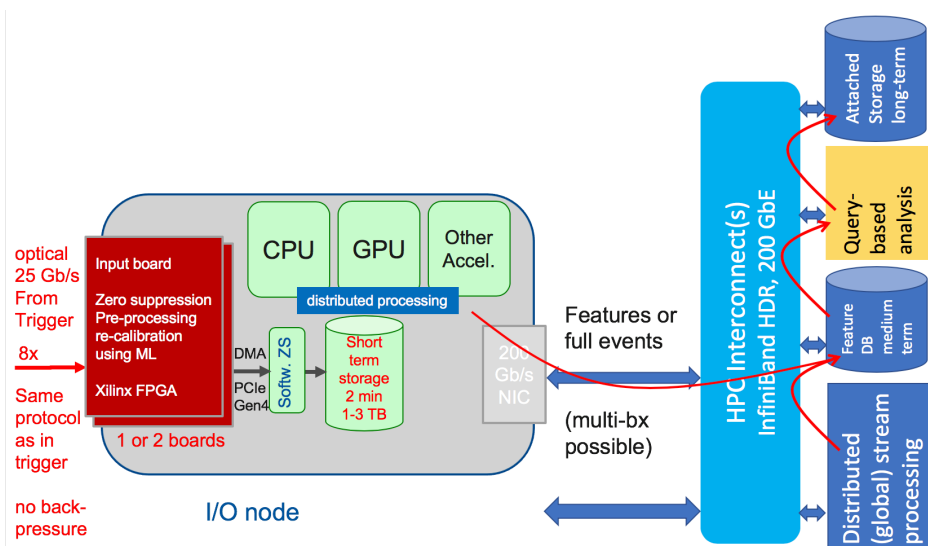
while in other cases they could be used to more precisely point the standard trigger chain to a certain signature. For example, in dark sector searches, where the models predict a wide range of signatures with low signal rates, requiring the best trigger efficiency, the scouting, applying no other thresholds than the implicit ones, can be used for early identification of promising potential signals. Once a potential signal is indicated by the scouting system, it may be possible to develop a dedicated Level-1 trigger algorithm for the corresponding signature.

Physics processes that may benefit from such a scouting-based analysis include: Higgs rare decays such as  $H \rightarrow J/\psi \gamma$ ,  $H \rightarrow \phi \gamma$ ,  $H \rightarrow \rho \gamma$ , where for example for the latter scouting could provide an alternative in identifying low momentum track pairs of a  $\rho$ . For displaced muons, scouting could provide alternative methods of matching muons with tracks or calorimeter deposits. Other examples include long lived particles with lifetimes covering multiple bunch crossings that are only partly (up to  $\pm 3$  bunch crossings) accessible at the standard Level-1 trigger; flavor anomalies,  $\tau$ -physics, requiring extremely efficient  $\tau$  identification;  $B_s \rightarrow \tau \tau$ , again requiring highly efficient  $\tau$  identification; hadronic physics in the hidden sector such as soft bombs; and QCD measurements with high statistics in regions of phase space inaccessible to the standard Level-1 trigger.

Moreover, the scouting system could serve as an invaluable diagnostics tool for the Level-1 trigger with unprecedented statistical reach and could be used for instant online measurements with excellent statistical precision, for example to use certain physics processes to measure instantaneous luminosity.

## 2 Planned Architecture

I/O nodes in the scouting system are expected to house one or two input boards, each receiving eight 25 Gb/s links (or more links at 16 Gb/s) using the same protocol as the Level-1 trigger. Data are transmitted unidirectionally with no back-pressure to the trigger. The scouting system therefore does not in any way interfere with the standard trigger system. Input boards with eight or more links based on a modern Xilinx FPGA are expected to be commercially available. The FPGA firmware will perform zero suppression and pre-processing of the data such as re-formatting or re-calibration, possibly using neural nets



**Fig. 2.** Data flow in the scouting system I/O node and scouting system. ZS: Zero suppression;

implemented with FPGA resources. From the input boards, data will be transferred by DMA via a Gen-3 or Gen-4 PCIe bus into the memory of the I/O nodes. Even after fine-grained zero-suppression, the long-term storage of the huge amount of raw data produced by the trigger processors, in view of a subsequent "classic" multi-tiered offline analysis and reduction, does not represent a viable approach.

Fast, local short-term storage (realized with a suitable technology guaranteeing large volumes and low latency) in the I/O node will collect pre-processed input data. Asynchronous distributed algorithms running on the I/O nodes and using a low-latency interconnect will combine multiple inputs and identify relevant physics objects or features. The short-term storage will be sized to provide enough buffer space to match the overall latency of the distributed processes.

Features deemed interesting will be handed, over the interconnect, to a distributed stream-processing system, to be further selected and combined into higher-level data, organized and stored in medium-term storage. For efficient operation of the subsequent analysis steps, the medium-term storage needs to be organized around some form of database technology allowing random access to, and reorganization of, features using dynamic criteria for analysis. For example, search-engine technologies could be adapted to work on mainly numerical data, organized and indexed using columnar formats or a key-value store. The medium-term storage decouples high-level analysis from the feature search, in much the same way as reconstruction and analysis phases in classic multi-tiered offline processing, while allowing convenient reorganization of the data. The distributed algorithms will be easily re-configurable without affecting the performance of the experiment, allowing new ideas to be tested, which can then possibly be imported into the standard Level-1 trigger.

Multiple parallel analyses/searches, possibly query-based, will be run off the medium-term storage content to produce a synthesis of relevant information, in the form e.g. of high-level distributions, "tuples" or collections of candidate "events", which will be used for analysis, diagnostics and monitoring, and stored long-term.

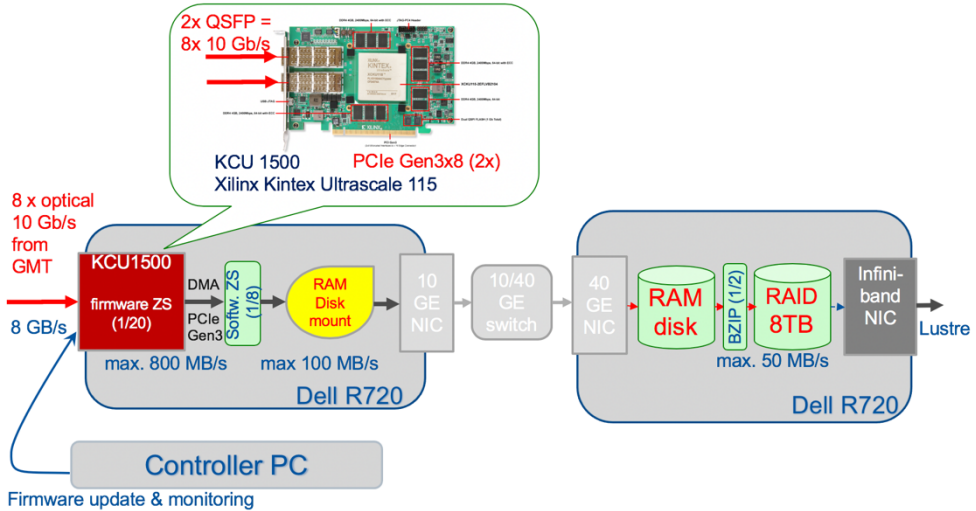
As shown in Figure 1, data from various stages of the trigger may be received by the scouting system, allowing for a staged deployment. An initial scouting system receiving only the Level-1 trigger decision and the inputs to the Global Trigger may be realized with relatively modest resources. Even this relatively small system could be used for feature searches and would serve as a diagnostics tool for the Global Trigger. The possible evolution of the scouting system and required resources are shown in Table 1. Outputs of the local muon trigger, local calorimeter trigger or the track finder processors may be added as needed, requiring significant resources. Adding full low-level primitives from the calorimeters would increase data volume and resource needs by an order of magnitude and is currently not considered.

**Tab. 1.** Stages for the evolution of the scouting system and resource requirements.

System	#input links (speed Gb/s)	#I/O nodes	Total short term storage / TB	#200Gb/s links to interconnect
Initial: Decision + GT inputs	68 (25)	9	6	9
Muons	36 (16)	3	33	3
Calo (Barrel+endcap+forward)	612 (16)	76	93	49
Track finder processors	324 (25)	41	120	21
Low-level calo primitives	thousands	hundreds	2-5k	hundreds

### 3 Prototypes in Run-2 and Run-3

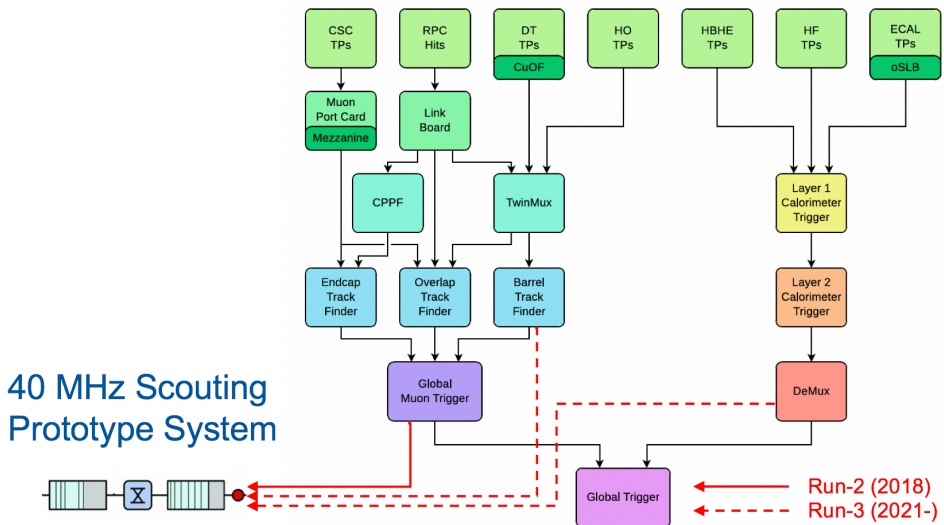
A prototype scouting system, illustrated in Figure 3, was developed towards the end of Run-2 and tested during the last weeks of the proton-proton ( $p-p$ ) run in October 2018 and



**Fig. 3.** Prototype scouting system deployed at the end of Run-2. ZS: Zero Suppression

during the heavy-ion (HI) run in November 2018. A commercial Kintex Ultrascale development kit (Xilinx KCU 1500 [5]) was used to receive 8 links at 10 Gb/s from the Global Muon Trigger  $\mu$ TCA board, containing the muon candidates sent to the Global trigger and additional muon candidates from the barrel region (see Figure 4). The Global Muon Trigger was modified to send an additional copy of its output links to spare output ports and to add the orbit and bunch counter on these links so that the scouting system prototype, which was not connected to the Trigger Control and Distribution System, could mark the data for later correlation with data from the regular Trigger/DAQ system.

Firmware for the scouting board was developed to decode the link protocol of the trigger, to align the 8 links with respect to each other and to zero-suppress bunch crossings not



**Fig. 4.** The Phase-1 Level-1 trigger of CMS and connection of scouting system prototypes. CSC: Cathode Strip Chambers; RPC: Resistive Plate Chambers; DT: Drift Tubes; HO: Hadronic Calorimeter – Outer; HF: Hadronic Calorimeter - Forward; ECAL: Electromagnetic Calorimeter; TPs: Trigger Primitives; oSLB: optical Synchronization and Link Board; CuOF: Copper to Optical Fibre; CPPF: Concentration, Pre-processing and Fan-out system

containing a muon candidate. This first zero-suppression step allowed the throughput to memory to be reduced by a factor of 20 during  $p$ - $p$  runs and more during HI runs. Data were buffered in FIFOs before being transferred to host memory over 8-lane PCIe Gen-3 using the Xilinx DMA engine. Software in the host PC performed more fine-grained zero-suppression gaining another factor of 8 before writing data to a RAM disk on another node mounted over 40 Gb/s Ethernet. On the second node, data were compressed with bzip2 [6], gaining another factor of 2, and stored to a RAID array, with the possibility to transfer them to the Lustre global file system. About 1 TB/day of compressed data were recorded during the last days of  $p$ - $p$  data taking, less during the HI run.

Valuable experience was gained in the few weeks of operation. Recovery from synchronization loss with the trigger system was largely automated, but in some cases still needed manual intervention. Some stability issues with the DMA driver still remain to be addressed.

The recorded data were analyzed offline for a number of studies. One analysis of the data showed for example that the count of muons recorded with the scouting system can be used as a per-bunch luminometer with a resolution comparable to other luminosity sources in CMS.

The distributed processing aspect of the proposed scouting system has also been prototyped at the test stand for upgraded electronics for drift-tube chambers, in Padova [7]. Here Apache Spark [8] and Apache Kafka [9] were demonstrated, which are likely to be used in a future scouting system.

For Run-3, starting in 2021, an upgraded demonstrator system is planned, capturing data also from the Layer-2 Calorimeter Trigger and from the Kalman Barrel Muon Track Finder. This extension will use two additional scouting boards and input nodes. The setup with three input nodes will enable the test of distributed processing prototypes.

## 4 Summary

We have presented a 40 MHz Level-1 trigger Scouting System for the Phase-2 upgrade of CMS. This system would analyze data from the Level-1 trigger in a largely separate data path from the standard Trigger/DAQ. Using accelerators and distributed processing it would extract features in near real-time and perform searches on the full statistics using the resolution of objects at the Level-1 trigger. Several physics channels are expected to benefit from such an analysis. Moreover, the scouting system would act as a sophisticated diagnostics tool for the Level-1 trigger. Prototypes have been developed and tested at the end of Run-2 and are going to be expanded for Run-3, starting in 2021.

## References

- [1] CMS Collaboration, JINST **3** S08004 (2008)
- [2] CMS Collaboration, CERN-LHCC-2017-009 (2017)
- [3] CMS Collaboration, CERN-LHCC-2017-023 (2017)
- [4] CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger”, CERN/LHCC (to be published)
- [5] Xilinx Kintex UltraScale FPGA KCU1500 Acceleration Development Kit, <https://www.xilinx.com/products/boards-and-kits/dk-ul1-kcu1500-g.html> [Accessed 2020-02-10]
- [6] J. Seward, “bzip2” [software] Available at <https://www.sourceware.org/bzip2/> [Accessed 2020-02-10]

- [7] M. Zanetti *et al.*, “Big Data solutions for the online processing of trigger-less detectors data”, in these proceedings
- [8] The Apache Software Foundation, “Spark” [software] Available at <https://spark.apache.org> [Accessed 2020-02-10]
- [9] The Apache Software Foundation, “Kafka” [software] Available at <https://kafka.apache.org> [Accessed 2020-02-10]