IL NUOVO CIMENTO **47 C** (2024) 122
DOI 10.1393/ncc/i2024-24122-0

COLLOQUIA: IFAE 2023

40 MHz Scouting for the CMS Level-1 Trigger(*)

S. GIORGETTI([1])([2]) on behalf of the CMS COLLABORATION

([1]) *INFN, Sezione di Padova - Padova, Italy*
([2]) *Dipartimento di Fisica, Università degli Studi di Padova - Padova, Italy*

received 13 February 2024

Summary. — The Level-1 trigger Data Scouting (L1DS) is a novel data acquisition system under development for the Phase-2 CMS detector at the High-Luminosity LHC (HL-LHC). Its purpose is to capture and process Level-1 trigger (L1T) information at the bunch crossing collision frequency of the LHC preceding the standard L1T selections. Referred to as 40 MHz Scouting, this system has the potential for filterless diagnostics for the detector, luminosity studies and investigations into signatures and processes that would be otherwise inaccessible or constrained due to the bias introduced by the trigger. An outline of Phase-2 L1DS is provided alongside a description of the Run-3 demonstrator's architecture and the preliminary results obtained.

1. – Introduction

The proton-proton collisions that the Large Hadron Collider (LHC) delivers to the CMS experiment generate a raw data stream of the order of TB/s. The CMS experiment designed a two-level trigger system that ensures excellent sensitivity to most physics signatures while reducing the event rate from 40 MHz to around 100 kHz at the Level-1 trigger (L1T) and then to around 1 kHz through the High Level Trigger (HLT) [1].

Introduced in 2011 at the HLT [2], the concept of scouting implies the acquisition of reduced event-content data at considerably higher rates than the standard accept rate. The 40 MHz Scouting system records the L1T primitives at the LHC bunch crossing frequency to carry out quasi-online studies on these data with limited resolution. The Level-1 trigger Data Scouting (L1DS) acquisition system will be fully integrated into the CMS Phase-2 L1T for the High-Luminosity phase of the LHC (HL-LHC); a demonstrator for the LHC Run-3 is currently being developed as a proof of concept [3]. The L1DS can provide a high-statistics dataset for detector and trigger diagnostics or luminosity

(*) IFAE 2023 - "Poster" session

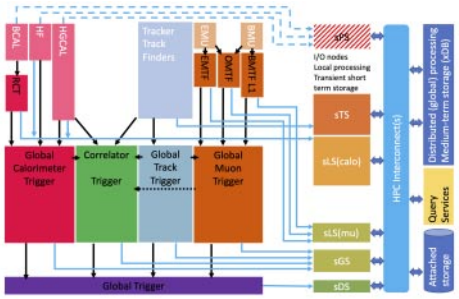


Fig. 1. – Design of CMS Phase-2 L1T (left) and the L1DS (right) [3].

measurements. Furthermore, it enables the study of events in unexplored phase space for which the trigger may be inefficient due to bandwidth and/or latency limitations. Several physics processes can be enhanced, for instance, particles spanning multiple bunch crossing, multiple soft jets or rare decay processes like $W \to \pi\pi\pi$.

2. – The Level-1 trigger Data Scouting for CMS Phase-2

For the high-luminosity phase, the HL-LHC will deliver an instantaneous luminosity of up to $7.5 \cdot 10^{34}$ cm^{-2}s^{-1}, nearly a fourfold increase to the current condition; additionally, the average number of proton-proton collisions per bunch crossing (pileup) is expected to reach approximately 200. The upgraded Phase-2 CMS detector will continue to use a two-level trigger system, with the Phase-2 L1T expected to achieve a 750 kHz rate and access, for the first time, the tracker and high-granularity calorimeter data [3]. The L1DS will benefit from the upgraded design of the Phase-2 L1T, the architecture is presented in fig. 1. The L1DS will read several L1 subsystems via spare output links and process them online with a heterogeneous computing farm adopting a scalable architecture based on two stages. Stage 1, which includes the scouting Decision System (sDS) and the scouting Global System (sGS), will mainly provide trigger diagnostic functionality for the Global Trigger (GT). The sDS will capture the output of the GT Final-OR and information on prescaling. The sGS will receive the output from the Global Calorimeter Trigger, the Global Track Trigger, the Global Muon Trigger, and the Correlator Trigger. Stage 2 can be additionally included to incorporate the endcap calorimeter primitives and local muon and regional barrel calorimeter trigger via a scouting Local System (sLS). In terms of hardware, Phase-2 L1DS will employ DAQ-800, the readout board of the Phase-2 CMS central data acquisition (DAQ). It will host two Xilinx VU35P FPGAs, each with 24x25 Gb/s of input bandwidth and 5 QSFP connections that provide 5x100 Gb/s of output bandwidth [3].

3. – The 40 MHz Scouting demonstrator for LHC Run-3

The Run-3 demonstrator aims to demonstrate the ability to operate on distributed data and includes processing units based on FPGA hardware and software [3]. The technical design is displayed in fig. 2. The data recorded from the L1T boards are transmitted at 10 Gb/s to a patch panel, from where they are distributed to the three classes of receiving boards. Table I outlines the four L1T sources and the characteristics of each receiving board. The VCU128 cards are linked to the host PC via 100 Gb/s

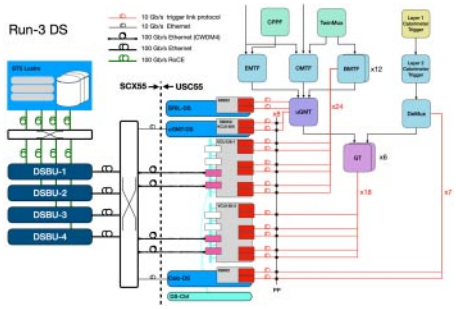


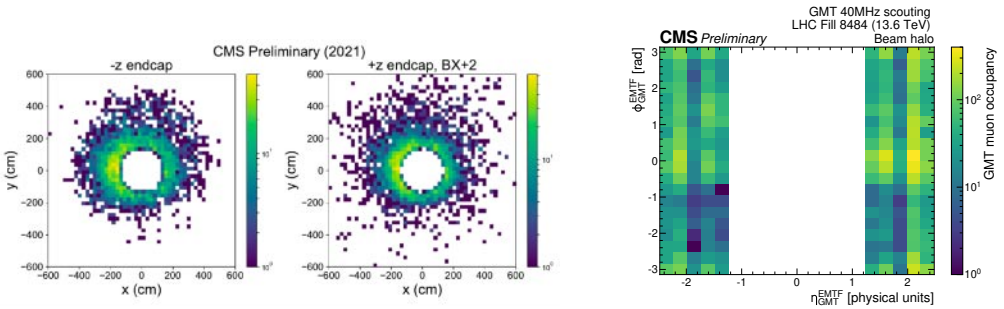
Fig. 2. – Architecture of the 40 MHz Scouting demonstrator for Run-3 [4].

Ethernet links, employing TCP/IP for transfer. The remaining boards utilize a direct memory access (DMA) engine through 10 Gb/s Ethernet links. Initially in firmware, then in software, a granular zero suppression is performed, reducing the data by a factor of about 10 [8]. The raw data are then written to RAM disks (DSBU) and further compressed on a second server before being stored in the CMS Lustre global file system.

3˙1. *Preliminary test with the demonstrator and first results at LHC Run-3* . – The demonstrator gathered data from the muon and calorimeter detectors for the first time during the LHC test beam in October 2021. Regular data collection from the μGMT and DeMux boards began in July 2022, with the start of LHC Run-3. To observe the effects of the beam halo, we can select μGMT muons reconstructed in the endcap regions by the Endcap Muon Track Finder (EMTF) in non-colliding bunch crossings (BX). The beam halo muons are a consequence of the interaction of beam particles with the pipe material or residual gas molecules in the vacuum chamber. Findings from the first test are displayed in fig. 3(a): larger beam halo in this fill can be seen for beam 2, which travels from the negative z side to the positive z side in CMS coordinates [5]. Figure 3(b) shows the occupancy in the azimuthal angle ϕ and pseudorapidity η for non-colliding bunch crossing in which the beam halo is visible. Beam halo muons are predominantly produced on the beam pipe plane, and a higher occupancy is observed in the detector

TABLE I. – *Description of the L1DS demonstrator receiving FPGA-based boards and of the inputs collected from the Global Muon Trigger (μGMT), the Barrel Muon Track Finder (BMTF), the Calorimeter Trigger (DeMux) and the Global Trigger (GT).*

Source	Objects per BX	N. links	Receiving boards
μGMT	Up to 8 μGMT final muons and BMTF intermediate muons	8	Xilinx KCU1500 (KU115 FPGA) and Micron SB852 (VU9P FPGA)
BMTF	Input muon stubs	12x2 (24)	Xilinx VCU128 (VU37P FPGA)
DeMux	Up to 12 e/γ, τ, jets. Energy sums	7+1 spare	Xilinx KCU1500 (KU115 FPGA)
GT	Decision algorithm bits	18	Xilinx VCU128 (FPGA VU37P)



(a) (b)

Fig. 3. – Observation of beam halo in L1DS data. (a) Occupancy in the negative (-z) and positive (+z) muon endcap of opposite charge muons selected requiring a 50 ns delay from the negative side to the positive, corresponding to ΔBX = 2 [5]. (b) Occupancy of endcap muons in ϕ and η coordinates for 1 hour of 40MHz Scouting data taking in 2022 [6].

region corresponding to the accelerator plane ($\phi \simeq 0$ and $|\phi| \simeq \pi$).

3˙2. *Boosting muons recalibration with ML algorithms* . – Machine learning (ML) models have been developed in the L1DS chain to perform online recalibration of the L1T muons and improve the resolution for a physics analysis. The resolution is limited since the L1 reconstruction does not have access to the full-granularity data and information from all sub-detectors, as in the offline case. The FPGA-based accelerator employed to run the inference is the Micron Deep Learning Accelerator (MDLA), a proprietary inference engine that translates ML models into instructions for Micron Technology's SB852 FPGA-based processing board [7]. The recalibration model takes as input ϕ, η, the transverse momentum p_{T}, the charge c and the muons reconstruction quality Q; the targets are the parameters of muons reconstructed offline matched to the L1 muons with $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ <0.1 . The neural network (NN) architecture consists of four layers, each with 128 hidden neurons. The training is performed using the Zero Bias 2022 dataset, which was collected by CMS without any trigger selection bias, and is therefore closely comparable to the scouting data. The difference between the prediction from the NN inference and the results of the offline reconstruction is illustrated in fig. 4. The NN algorithm proves to be a valid approach to boost the calibration of muon's parameters. An improvement in resolution is evident for ϕ, but a smaller gain is expected for η since

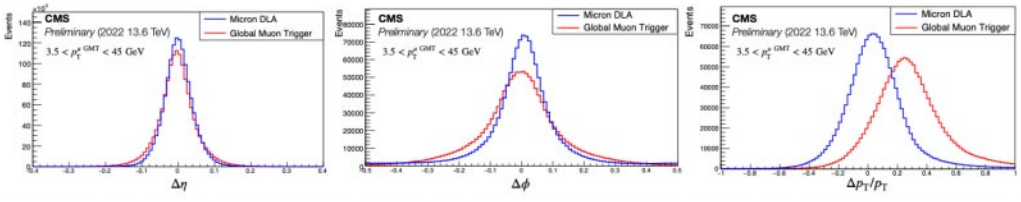


Fig. 4. – The distribution of differences for ϕ, η and p_{T} between the Micron DLA prediction or μGMT values, and the offline reconstructed matched muon. A significant improvement in track parameter resolution is observed in the MDLA result (red) when compared to the μGMT output (blue) [8].

particles do not bend with respect to this angle. Concerning transverse momentum, the observed offset in the μGMT case is also corrected. This offset arises from the calibration of L1T objects to ensure a specific efficiency at different p_{T} thresholds. With an expected throughput in Run-3 of 2M muons per second, the performance in terms of latency and throughput is satisfied with around 2.7M inferences per second [8].

REFERENCES

[1] CMS Collaboration, *JINST*, **3** (2008) S08004.
[2] CMS Collaboration, *Phys. Rev. Lett.*, **117** (2016) 031802.
[3] CMS Collaboration, *The Phase-2 Upgrade of the CMS Level-1 Trigger*, Technical Design Report CERN-LHCC-2020-004, CMS-TDR-021, `https://cds.cern.ch/record/2714892`.
[4] CMS Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, arXiv:2309.05466, submitted to *JINST* (2023).
[5] Ardino R. *et al.*, *Nucl. Instum. Methods Phys. Res. A*, **1047** (2023) 167805.
[6] CMS Collaboration, *40 MHz Scouting with Deep Learning in CMS*, `https://cds.cern.ch/record/2843741` (2022).
[7] Micron Technology, *Micron DLA Software Development Kit - SDK*, `https://github.com/FWDNXT/SDK`.
[8] James T. O., *The Level 1 Scouting system of the CMS experiment*, CMS-CR-2023-024 (2023).