PROCEEDINGS OF SCIENCE

Readout and testing procedures to characterize the CMS inner tracker pixel detector for HL-LHC

Mauro E. Dinardo[a,*] **on behalf of the CMS Collaboration**

[a] *Università degli Studi di Milano - Bicocca and INFN,*
Piazza dell'Ateneo Nuovo n. 1, 20126 Milano, Italy

E-mail: mauro.dinardo@cern.ch

The LHC will be upgraded to the High Luminosity LHC in the coming years, aiming to reach an instantaneous luminosity of up to $5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The CMS Tracker detectors will be replaced and significantly upgraded to cope with the increased radiation fluence while ensuring excellent performance. In particular, a new hybrid pixel detector chip was developed for the Inner Tracker detector (ROC). The chip is capable of coping with extreme hit rates of up to $3\,\mathrm{GHz/cm}^2$ ($\sim$12 GHz per chip), together with a trigger rate of $\sim$ 1 MHz, and features an efficient readout rate of up to 5.12 Gbits/s. The chip exhibits radiation tolerance of up to 1 Grad and an induced single-event upset rate of up to 100 upsets per second. The new Inner Tracker will have six times smaller area pixels covering a surface close to $5\,\mathrm{m}^2$, thus resulting in approximately two billion pixels over about 3900 modules. The individual detector modules will need to be characterized and calibrated before being mounted on the final detector structure. To this extent, a dedicated data acquisition system (DAQ), based on minimal hardware featuring a custom FPGA board, was developed. A description of the DAQ, the testing procedures, and experience with the ROC is presented in this document.

33rd International Workshop on Vertex Detectors (VERTEX2025)
25-29 August 2025
University of Tennessee, Knoxville, USA

*Speaker

 ISSN 1824-8039 . Published by SISSA Medialab.

PoS(VERTEX2025)030

1. Introduction

The High-Luminosity LHC (HL-LHC) [2] is an upgrade of the CERN LHC expected to operate for at least 10 years starting in 2030. It will run at a nominal center-of-mass energy of 14 TeV with a 25 ns bunch spacing. The peak luminosity will reach $5 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ with an average of up to 140 overlapping proton-proton collisions per bunch crossing. To preserve physics object performance, despite the high average number of interactions per crossing, LHC experiments will require significant upgrades (Phase-2 upgrade). In particular, the CMS [1] Inner Tracker detector (IT), shown in Fig. 1, will have to withstand a trigger rate of ~ 1 MHz, a hit rate of $\sim 3\,\mathrm{GHz/cm^2}$ for the innermost layer, with a trigger latency of 12.5 μs. Over the 10 years of the foreseen experiment lifetime the detector innermost layer will be exposed to an unprecedented radiation level of $\sim 2 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$, and 1 Grad of dose [3].

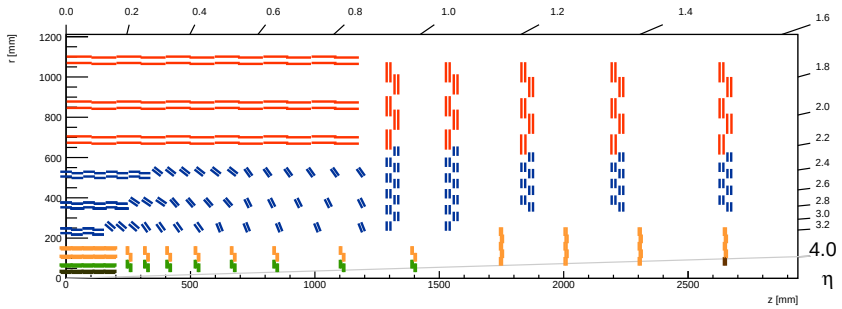

Figure 1: One-quarter of the CMS Phase-2 tracker in the $r-z$ plane. IT modules with planar silicon sensors are shown in green for modules with 1×2 and orange for modules with 2×2 readout chips. Modules with 1×2 readout chips and silicon sensors in 3D technology are shown in brown. OT modules with Pixel-Strip (PS) and Strip-Strip (2S) sensors are shown in blue and red, respectively [3].

1.1 The sensors

The Inner Tracker sensors have a $25 \times 100\,\mu\mathrm{m^2}$ pixel size (i.e., a factor of 6 reduced with repsect to the current CMS detector) for a total of ~ 2 billion pixels, with an active thickness of 150 μm. All layers but the innermost one will be instrumented with planar pixels. The innermost barrel layer will be instrumented with 3D pixel sensors for their higher radiation tolerance and for their better thermal performance. In fact, the temperature needed to prevent thermal runaway in planar pixel sensors is significantly below the minimum achievable temperature, while for 3D pixel sensors it features a margin of more than 4^o C.

1.2 The readout chip

The Inner Tracker ReadOut Chip (ROC) [4] features a matrix of $336 \times 432 = 145\,152$ pixels, $50 \times 50\,\mu\mathrm{m^2}$ wide, with zero suppression, and a Time-over-Threshold (ToT) hit charge measurement with a 4-bits ADC. The global threshold can be trimmed at pixel level with 5-bits precision, and the gain features a dual-slope to improve low charge sensitivity while avoiding saturation at high charges. The ROC features an internal injection circuit that can initiate the signal acquisition as if a particle crossed the sensor.

The ROC receives clock, trigger, and commands at 160 Mbps, while it can transmit data on multiple lanes, up to four, at a speed of 1.28 Gbps per lane, for a total of 5.12 Gbps. ROCs are organized in two size modules, with 1×2 and 2×2 ROCs, for a total of $\sim$ 3900 modules, that need to be thoroughly tested before installing them on the detector. On the modules the ROC features data merging capabilities as shown in Fig. 2, where two possible configurations are presented.

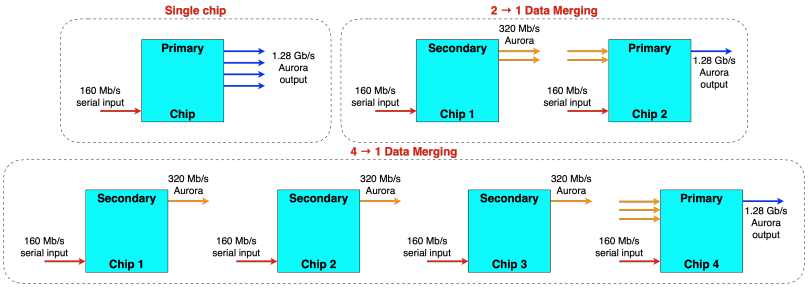

Figure 2: Sketch showing the data merging capability of the readout chip. In particular, are represented two possible data merging configurations: one where the secondary chip sends the data through two lanes, at 320 Mbps each, to the primary chip, which sends the data to the central data acquisition using one lane at 1.28 Gbps; the second where three secondary chips send the data through one lane, at 320 Mbps, to the primary chip, which sends the data to the central data acquisition using one lane at 1.28 Gbps.

Data are sent (received) to (from) a custom System-on-Chip board (SoC), which is called Data, Trigger and Control - DTC board, via optical fiber at 10 Gbps (2.5 Gbps). Between the frontend modules and the DTC there is an ad-hoc board, portcard, that houses the Low-power Gigabit Transceiver (LpGBT) chip. The communication between the module and the LpGBT is through twisted pairs, while from the LpGBT to the DTC is through optical fibers. Figure 3 shows the connection configuration in case of high and low occupancy links.

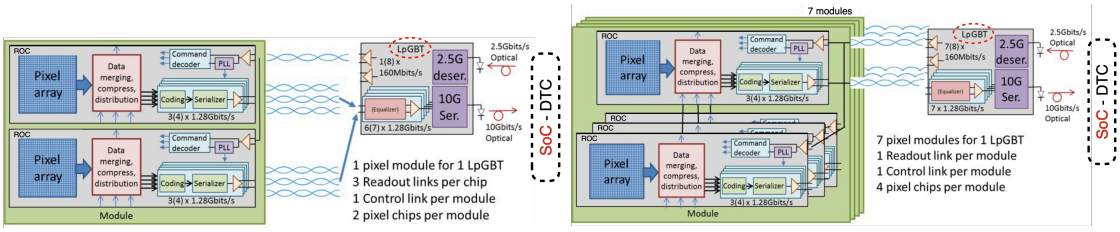

Figure 3: Schematic of the connection between the frontend module and the DTC in case of high (left) and low (right) occupancy links, respectively. The frontend module is connected through twisted pairs to the portcard housing the Low-power Gigabit Transceiver (LpGBT) chip, while the connection to the DTC is through optical fibers [3].

1.3 Data acquisition

A dedicated data acquisition system (DAQ) was developed to assist the readout chip and sensor development, both on the test bench and at beam test experiments. The DAQ also features the capability of managing a large number of modules. It is currently being used to perform thorough tests, and to grade the modules, before installing them on the detector. The software, called the Phase-2 Acquisition and Control Framework - Ph2-ACF [6], is written in C++ and is designed to

be user-friendly, based on a minimal hardware, and capable to conduct stress tests on modules to ensure they meet all requirements. An example is the burn-in test, where modules undergo multiple thermal cycles to guarantee the required production throughput. The DAQ consists of three main components: software (Ph2-ACF), hardware (called μDTC and shown in Fig. 4), and firmware.

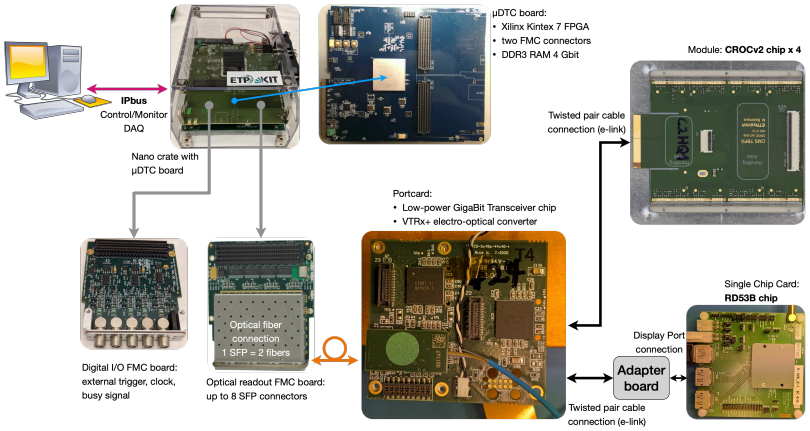

Figure 4: Different components and their interconnection of the current IT module testing hardware with an optical readout chain.

The software is capable of performing several scans and optimizations. In particular, it can perform a `PixelAlive`, injecting n-times the pixels, and recording how many times they respond, hence computing the efficiency. Analogously, the software can also scan the injection delay time (`InjectionDelay` scan), and record the average efficiency as a function of the injection delay. The software is also capable of measuring the threshold by scanning the injected charge and measuring the efficiency (`SCurve` scan). Finally, the software can adjust the per-pixel threshold to reduce its spread across the pixel matrix (`ThrEqu` calibration).

Together with the scan and tuning capabilities, the software allows the monitoring of the status of the frontend electronics by measuring voltages, currents, and temperatures as a function of time. Data are stored in `ROOT`[5] files, asynchronously with respect to other tasks, such as the calibration scans, as they are performed using multiple threads.

Module production validation is performed using two frameworks, a Graphical User Interface (GUI) and a Text-based User Interface (Dirigent), which are reliable, consistent (based on `JSON` files), and easy-to-use. They orchestrate power supplies and interlocks, Ph2-ACF (its configuration files, and its output files), the grading software, and the upload to the database (DB). In particular, they can run multiple instances of Ph2-ACF in parallel, and disable any module that is problematic during testing.

The next sections show some particular measurements performed with this DAQ. In particular, Sec. 2 shows the evaluation of the pixel cross-talk. Section 3 presents the evaluation of the analog and digital efficiency of the ROC. Sections 4 and 5 show how to properly run the ROC at low temperature and after irradiation, respectively. Finally, the material is summarized in Sec. 6.

2. Cross-talk

Pixel cells are capacitively coupled and therefore they exhibit cross-talk, which differs between planar and 3D sensors. Figure 5 shows four adjacent pixel cells and the meaning of "coupled" and "uncoupled" pixels. However, 3D sensors show that both coupled and uncoupled pixels have the same coupling capacitance, hence the same cross-talk with respect to the injected cell, which is ~10 fF. Planar sensors, instead, show a different coupling capacitance for coupled pixels (~14 fF), and uncoupled pixels (~6.5 fF).

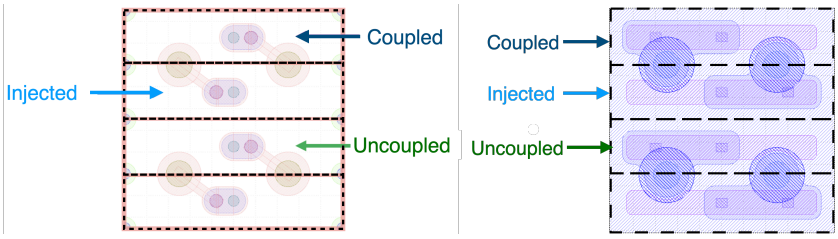

Figure 5: Layout of four adjacent pixels: 3D on the left and planar on the right. In particular, the bump pads between pixel cells are shown, which determine whether adjacent cells are "coupled" or "uncoupled".

The cross-talk (XT) is evaluated by means of `SCurves` measured for both injected and adjacent pixels, coupled and uncoupled. If we call Q_1 the measured threshold of the injected pixel, then the actual threshold should be $Q_1 \cdot (1 - \mathrm{XT})$. If we call Q_2 the measured threshold of the adjacent pixel, then the actual threshold should be $Q_2 \cdot \mathrm{XT}$. At first order all pixels have same threshold, therefore $Q_1 \cdot (1 - \mathrm{XT}) = Q_2 \cdot \mathrm{XT}$. We can isolate $\mathrm{XT} = \frac{r}{1-r}$, where $r = Q_1/Q_2$. Hit data are recorded only if there is a clock transition while the comparator signal is "high", and the comparator signal stays "high" as long as the input charge is above threshold. This implies that the actual threshold is dependent on the charge injection time. Therefore, XT is evaluated as a function of injection time.

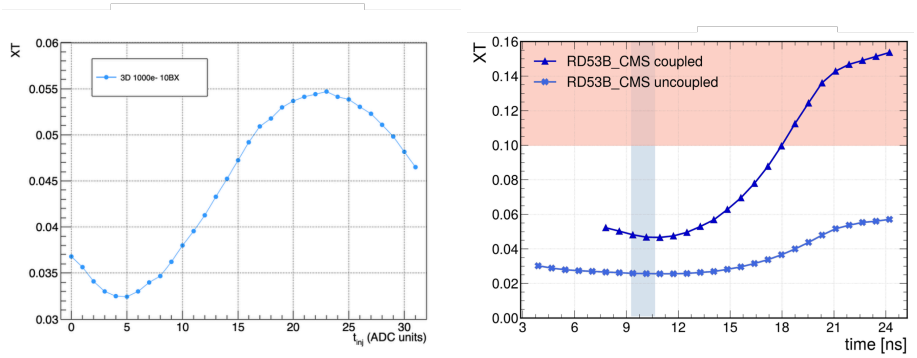

Figure 6: Cross-talk as a function of the charge injection time measured by means of `SCurve` scans (3D sensors on the left, and planar sensors on the right).

Figure 6 shows the cross-talk as a function of the charge injection time both for 3D and planar sensors. Cross-talk turned out to be useful to detect open bumps because neighboring pixels "sense" the induced charge only if the bump is properly connected, and a fast procedure based on `PixelAlive` was put in place to detect the open bumps.

3. Hit detection efficiency

Two main mechanisms are responsible for the degradation of the ROC hit detection efficiency: "analog" and "buffering" inefficiencies. Both are evaluated using high hit-rate generated by an X-ray source, and superimposing hits from the internal injection circuit, whose timing and number are very well controlled. The efficiency is then evaluated as the fraction of injections successfully detected by the chip. The two types of efficiency are:

Analog efficiency: to register two separate hits within the same pixel the first hit must return below threshold before the second one comes in, if this doesn't happen then the ROC is experiencing "analog inefficiency".

Buffering inefficiency: the trigger for a particular bunch crossing is delayed by some fixed latency, therefore hit data from the pixels must be temporarily stored inside buffers for an equally long amount of time, after which they can be safely discarded. Any hit detected while their buffer is full will be lost, hence the ROC will be experiencing "buffering inefficiency".

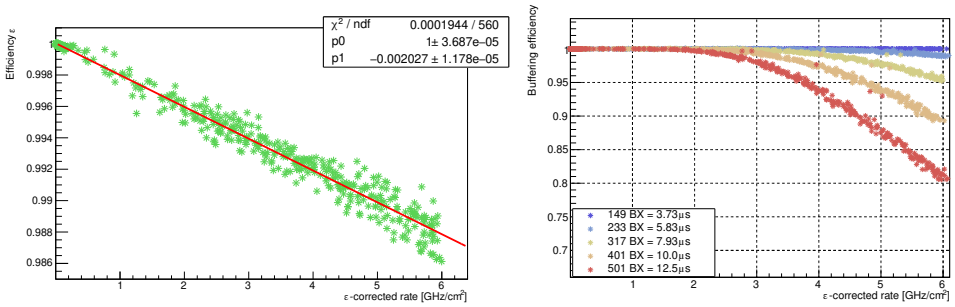

Figure 7: The plots illustrate the degradation of hit detection efficiency as a function of particle rate. The left plot presents the measurement of "analog efficiency", while the right plot displays "buffering efficiency". Analog efficiency exceeds 99% at a hit rate of 3 GHz/cm^2, whereas buffering efficiency falls below 99% at this rate. In practical experimental conditions, particles generate clusters of adjacent pixels; therefore, efficiency should be assessed at a particle rate of 1.78 GHz/cm^2, where buffering efficiency surpasses 99%. The right plot also includes superimposed curves corresponding to various latencies.

4. How to run cold

As mentioned in Sec. 1.2, the threshold dispersion can be reduced thanks to a per-pixel threshold trimming (`TDAC`) by means of the `ThrEqu` calibration. At room temperature the typical threshold dispersion is ~8 (in arb. units proportional to the number of electrons), while at -14^o C the threshold dispersion doubles. To reduce the dispersion at low temperature one can act on the TDAC gain, called `LDAC`, as shown in Fig. 8.

5. Irradiated module operation

The IT ROC is made in 65 nm CMOS technology, which is radiation tolerant up to 100 Mrad. Further precautions were adopted to make it more radiation resistant up to 1 Grad, i.e., wider and

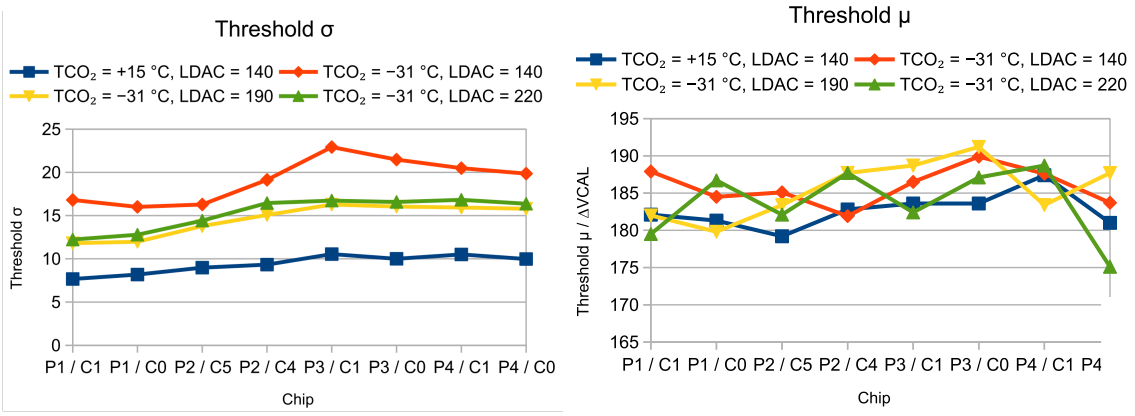

Figure 8: Threshold dispersion (left) and average threshold (right) measured for different ROCs (x axis). The various curves correspond to different coolant temperatures and different `LDAC` values (a coolant temperature of -31^o C corresponds to $\sim -14^o$ C on the ROC).

longer transistors, triple modular redundancy, limit annealing periods (i.e., run below -10^o C), homogenous Copper vias thickness for the current mirrors. However, to operate the ROC after heavy irradiation ($> 10^{16}\,\mathrm{n}_{eq}/\mathrm{cm}^2$) one needs to tune the pre-amplifier and comparator currents. An ad-hoc optimization procedure was adopted for irradiated chips (cfr. Fig. 9): (1) set the threshold to a high value (e.g. $4500\,\mathrm{e}^-$) and all TDACs to 0; (2) set `BIAS_COMP` to 0 and measure the analog current, then tune `BIAS_COMP` to increase the current consumption by $+2.2\,\mu$A per pixel; (3) set `IPA_IN_BIAS` to 0 and measure the analog current, then tune `IPA_IN_BIAS` to increase the current consumption by $+3\,\mu$A per pixel; (4) set all `TDACs` to 16 and tune `LDAC` to increase the current consumption by +400 nA per pixel; (5) rescale `IFC_BIAS` and `COMP_STARV` by the same scale factor of `BIAS_COMP`.

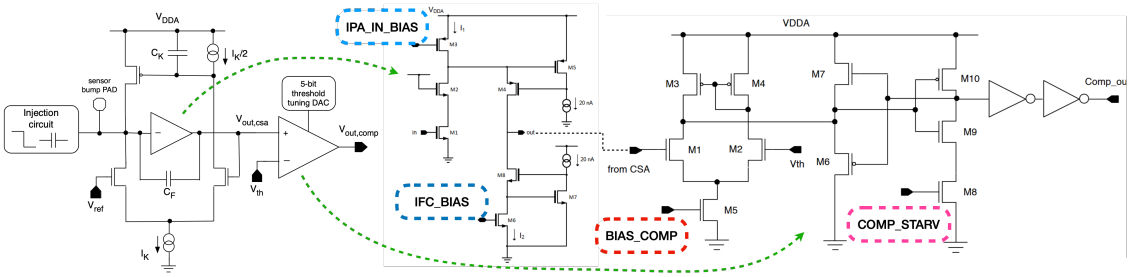

Figure 9: Schematic view of the pre-amplifier and comparator building blocks (left), together with their implementation at transistor level (right) [4]. The currents mentioned in the text are highlighted in the figure.

6. Summary

The CMS Inner Tracker (IT) detector will be upgraded to address the operational challenges foreseen for the High-Luminosity Large Hadron Collider (HL-LHC). Key requirements include radiation hardness up to $2 \times 10^{16}\,\mathrm{n}_{eq}/\mathrm{cm}^2$ and 1 Grad of dose, the ability to withstand a hit rate

of 3 GHz/cm^2, and a trigger rate of approximately 1 MHz with a latency of 12.5 μs, all while maintaining a power budget of approximately 1 W/cm^2. A specialized readout chip was designed based on these specifications. As the CMS IT approaches the production phase, all aspects of the pixel detector will be systematically evaluated, optimized, and assessed using a dedicated readout system. This document discusses several critical features of the IT, including cross-talk, hit detection efficiency, and operational performance under low-temperature conditions and after significant irradiation.

References

[1] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004 [DOI:10.1088/1748-0221/3/08/S08004]

[2] G. Apollinari et al., *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*, *DOI:10.5170/CERN-2015-005* (2015).

[3] CMS Collaboration, *The Phase-2 Upgrade of the CMS Tracker*, *DOI:10.17181/CERN.QZ28.FLHW* (2017).

[4] G. Alimonti et al., *RD53 pixel readout integrated circuits for ATLAS and CMS HL-LHC upgrades*, *JINST* **20** (2025) P03024 [DOI:10.1088/1748-0221/20/03/P03024].

[5] R. Brun and F. Rademakers, *ROOT – An object oriented data analysis framework*, *NIM-A* **389** (1997) 81-86 [DOI:10.1016/S0168-9002(97)00048-X].

[6] M. E. Dinardo et al., *The CMS Inner Tracker DAQ system for the High Luminosity upgrade of LHC: from single-chip testing, to large-scale assembly qualification*, *EPJ Web of Conferences* **295** (2024) 02028 [DOI:10.1051/epjconf/202429502028].