HGCROC2: the front-end readout ASICs for the CMS High Granularity Calorimeter

[2]F. Bouyjou, [4]G. Bombardi, [1]S. Callier, [1]P. Dinaucourt, [1]F. Dulucq, [1]M. El Berni, [2]F. Guilloux, [3]M. Idzik, [1]C. de La Taille, [4]A. Marchioro, [3]J. Moron, [1]L. Raux, [1]D. Thienpont, [4]T. Vergine

[1]OMEGA/CNRS/Ecole Polytechnique, bât. 404, route de Saclay, 91128 Palaiseau, France

[2]CEA-IRFU, Site CEA de Saclay, 91191 Gif-sur-Yvette, France

[3]AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland

[4]CERN , Esplanade des Particules 1, 1217 Meyrin, Switzerland

E-mails: [1]damien.thienpont@in2p3.fr

Abstract. The CMS High-Granularity Calorimeter (HGCAL) imposes extremely challenging specifications for the front-end electronics: high dynamic range, low noise, high-precision time information and low power consumption, as well as the need to select and transmit trigger information with a high transverse and longitudinal granularity. HGCROC2 is the second prototype of the readout chip embedding almost all the final functionalities. It has 72 channels of the full analog chain: low noise and high gain preamplifier and shapers, a 10-bit 40 MHz SAR-ADC which provides the charge measurement over the linear range of the preamplifier, after the preamplifier saturation a discriminator and TDC provide the charge information from ToT (200 ns dynamic range and 50 ps binning), and a fast discriminator and TDC provide timing information to 25 ps accuracy. This paper reports on the performance in terms of noise, charge and timing, the DAQ and Trigger paths, as well as results from radiation qualification with total ionizing dose (TID) and heavy ions for single-event effects (SEE).

1. Introduction

The High Granularity Calorimeter (HGCAL) [1], presently being designed by the Compact Muon Solenoid collaboration (CMS) to replace the existing endcap calorimeters for the High Luminosity phase of the LHC (HL-LHC), will feature unprecedented transverse and longitudinal readout and triggering segmentation for both electromagnetic and hadronic sections. The full system will be operated at -30°C. As described in figure 1, two types of active elements will be implemented: 640 m² of silicon sensors with Cu/CuW/Pb absorbers within the full electromagnetic calorimeter (CE-E) and within the higher intensity region of the hadronic calorimeter (CE-H), and 370 m² of SiPM-in-scintillators tiles within the lower intensity region of the hadronic calorimeter.

Silicon sensors will have three different active thicknesses (300, 200 and 120 µm) in order to optimize the charge collection and operation conditions over the full lifetime of the HGCAL. Each sensor has either 192 or 432 individual diodes, which act as sensor cells.

The hadronic calorimeter will use scintillator as the active material in regions where the integrated dose is low-enough for the scintillator, and the fluence is limited to retain good overall performance over the whole life of the HL-LHC.

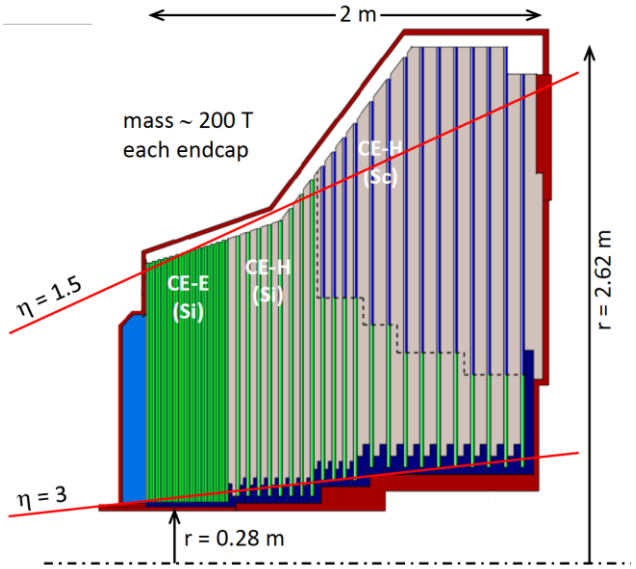

Figure 1:cross-section view of HGCAL. In green, the region with silicon sensors and in blue with SiPM-in-scintillators tiles.

As a result, two versions of the chip are required to read out the two types of sensors, silicon and SiPM. They have to provide a charge measurement up to 10 pC for the silicon-version and 300 pC for the SiPM-version, with a noise level below 0.4 fC (2500 electrons). In order to mitigate the out-of-time pile-up, the chip has to achieve a fast shaping time (peak < 25 ns), and having less than 20 % of the signal in the next bunch crossing. Furthermore, it has to provide a precise timing capability with 25 ps binning (~100 ps resolution around 10 fC, < 25 ps resolution after 100 fC).

2. HGCROC2 overview

The front-end electronics measures and digitizes the charge deposited in the silicon sensors pads or generated in the SiPMs, provides a high precision measurement of the time of arrival of the pulses, and transmits the digitized data to a concentrator chip. It also computes, at every bunch crossing, digital sums of neighbouring cells that are transmitted to the trigger concentrator chip and then to the back-end electronics to build trigger primitives.

2.1. HGCROC2 architecture

HGCROC2 embeds 72 regular channels of the full analog chain achieving charge and timing information, 2 calibration channels which will be connected on smaller cells for calibrating on the MIP throughout the entire detector life and 4 common mode channels for possibly subtracting the common noise.

Each channel is made of a low-noise preamplifier. Its gain can be adjusted so that MIP gives 10 ADC counts. In the preamplifier's linear regime, the signal goes to the 10-bit SAR-ADC [2] through the shaper which optimizes the signal to noise ratio. After the preamplifier saturation, the charge measurement is provided by a discriminator associated to a 12-bit TDC which measures the Time-over-Threshold (ToT) with 50 ps binning and so up to 200 ns. Another fast discriminator associated to a 10-bit TDC provide the Time-of-Arrival (ToA) with 25 ps binning. The chip provides ADC, ToT and ToA data of each bunch crossing at 40 MHz. Those data are stored in a 512-depth circular memory for 12.8µs [3]. Only those selected by a L1A trigger request, after 12.5 µs, will be sent out through two 1.28 Gbps links.

HGCAL contributes to the L1A trigger primitives, the first trigger level of CMS, and therefore the chip has to provide a compressed view of all the events at 40 MHz. This is achieved by summing and compressing over 7 bits the charge from 4 or 9 channels; four 1.28 Gbps links are devoted to send out the trigger data.

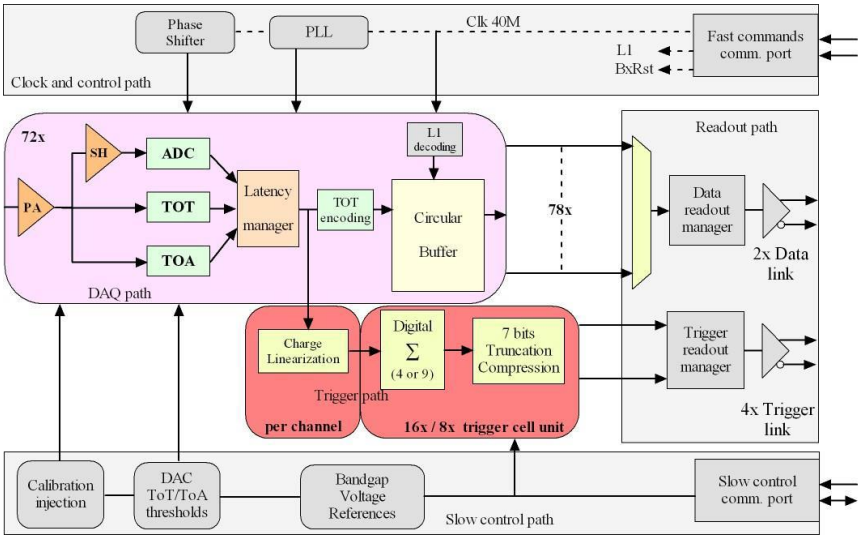

Figure 2: HGCROC2 architectural overview.

The I2C protocol is used to set or read the more than 7900 parameters of the chip. This part is triplicated to resist to Single Event Effects (SEE). The chip is controlled by the Fast Command block which receives a clock and a command link at 320 MHz. This allows to configure the operating mode of the system: links synchronization, reset, calibration, L1A request, etc. The 40 MHz clock, in phase with the LHC clock, is extracted from the 320 MHz fast command link and provides the clock to the digital part of the ASIC (digital processing, I2C) and to the PLL which generates the others clocks needed to operate the chip: the 640 MHz clock for the 1.28 Gbps links, the adjustable phase 40 MHz clock for the ADCs, the adjustable phase 160 MHz clock for the TDCs.

The chip is made fully symmetric. Each half hosts 8 blocks of 4 channels, 2 common mode blocks, the bias, a bandgap and all the reference voltages, the digital processing for the trigger sums and the data storage. Common to the chip, there are only the I2C, the fast command block, the PLL and the 1.28 Gbps links.

2.2. Packaging and characterization boards

The chip does not use wire bonds but C4 bumps which allow a much better power distribution and led to a large improvement in terms of noise and crosstalk compared to the first prototype.

The chip will be packaged in two different BGA:

- Low density package, 0.8 mm pitch for Hexaboard with ~200 channels (200 µm and 300 µm thick silicon sensors) and Tileboards with ~70 channels
- High Density package (0.6 mm pitch) for Hexaboard with ~400 channels (120 µm thick silicon sensors)

A first characterization board has been produced with a bare chip directly flipped on a mezzanine (FlipChip board) where all power supplies are kept separated, and a very low ground impedance has been achieved by relaxing the design constraints. That makes possible the chip characterization in an ideal environment. This board demonstrates the chip reaches very good performance in terms of noise, crosstalk and digital coupling.

Another board has been developed to validate the operation of the circuit in a BGA and in a more realistic environment (Single BGA board). A coupling of the digital clock appears on the waveforms

and comes from digital current spikes on preamp ground node. Further details are given in section 3.2 of this paper.

3. Charge measurement

3.1. Linearity performance

In the preamplifier linear range, the charge is provided by the 10-bit ADC. Figure 3 shows the linearity of the ADC measurement as a function of the injected charge. The Integral Non Linearity (INL) curves show the linearity remains within 1% of the full dynamic range with 2 ADC counts (~0.3 fC) resolution. The uniformity over the full chip is below 1 %: the small difference between the two halves is due the fact that the ADCs are not identical and do not have exactly the same LSB.

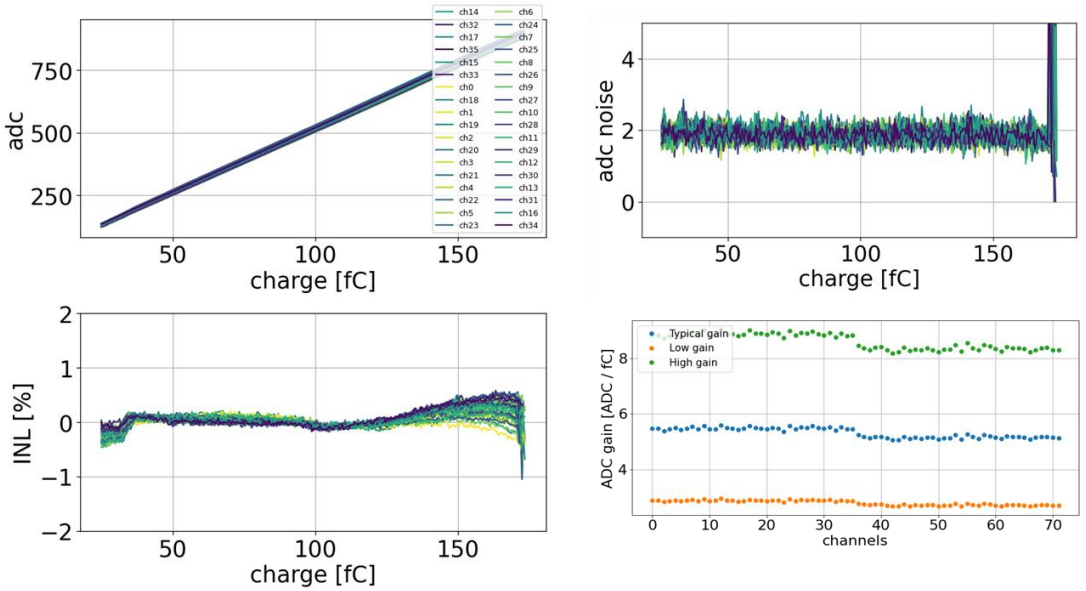

Figure 3: ADC measurement. Top left: ADC vs. Charge for all channels from 1 half; Bottom left: INL; Top right: ADC noise vs. charge; Bottom right: ADC gain uniformity for the lower (orange), typical (blue) and higher (green) preamplifier gains.

After the preamplifier saturation, the charge is obtained by measuring the width of the preamplifier output which gives the ToT. Figure 4 shows the ToT linearity as a function of the injected charge, the INL which remains within 2 % of the full dynamic range with around 50 ps (1 LSB) resolution. Some residual wiggles are visible on the INL which are due to digital clock coupling on the channel inputs passing through the PCB ground. Also some peaks appear in the resolutions which come from outliers. Indeed, the TDC goes sometimes wrong when converting the time, this effect has been understood, reproduced in simulation and corrected in the next version of the chip.

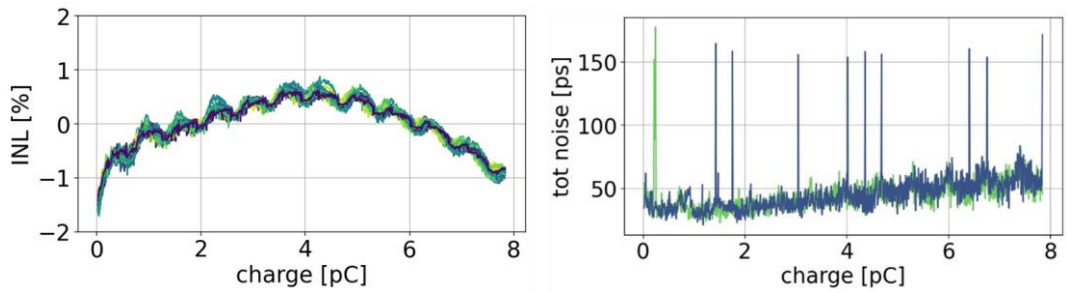

Figure 4: ToT measurement for all channels from 1 half. Left: INL; Right: ToT jitter

3.2. Noise performance

The noise performance is one of the most important parameters for a chip dedicated to calorimetry. The series noise whose main contributor is the preamplifier input transistor depends on the detector capacitance value. From the measurements shown in Figure 5, we calculate a series noise of 0.7 nV/√Hz with a preamplifier input capacitance of 6 pF in agreement with the simulations and the specifications.

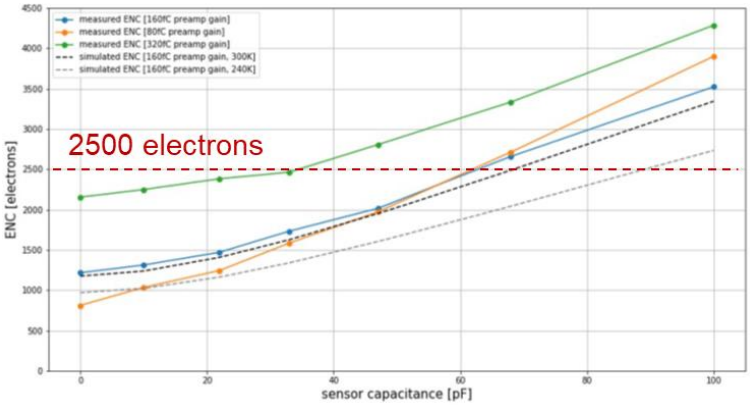

Figure 5: Equivalent Noise Charge (ENC) in electrons with respect to the input capacitance for the lower (orange), typical (blue) and higher (green) preamplifier gains. The dashed black curve is a simulation corresponding to the typical gain configuration at room temperature and is in accordance to the corresponding measurement of the blue curve.

In the final system, the high voltage, needed to bias the silicon sensor, as well as both power supplies and ground ripple could be important sources of coherent noise. It is therefore mandatory to check the noise performance at chip level and to identify the critical constraints for the module design. A first way to check the presence of coherent noise is to look at the correlation matrix which shows how much the noise from each channel is correlated to any others channels. The matrix in Figure 6 shows that there is small correlation coefficient, below 0.1, between channels from the same half and no correlation between channels from different halves. This result is achieved with both FlipChip board and low-density BGA package. With high-density BGA package the correlation coefficient increases up to 0.15. In order to furthermore quantify the coherent noise contribution, we can extract it by comparing direct ($DS = \Sigma\, ch_i$) and alternate ($AS = \Sigma\,(-1)^i ch_i$) sums of the $n = 72$ channels, the incoherent noise is then expressed as $IN = rms(AS)/\sqrt{n}$ and the coherent noise as $CN = \frac{\sqrt{rms(DS)^2 - rms(AS)^2}}{n}$. We find for low-density BGA package 1.6 ADC units incoherent noise and 0.3 ADC units coherent noise while for the high-density BGA package, coherent noise is 0.42 ADC units. On FlipChip board and on low-density BGA package, the noise performance are as expected, close to the expected performance; on high-density BGA package, the degraded performance due to higher coherent noise demonstrates the need to optimize the PCB grounding and the power supply decoupling.

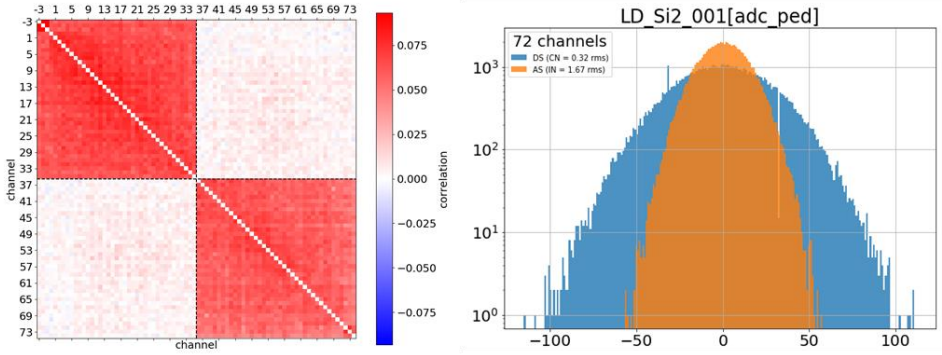

Figure 6: Noise analysis. Top left: correlation matrix; Top right: Direct and Alternate sums distribution.

The previous noise discussion focused on the ADC noise but at a constant phase with respect to the 40 MHz digital clock. By scanning the ADC sampling phase, by step of 1.5625 ns, the waveform can be reconstructed and the digital-induced effects are then appearing. A 40 MHz modulation can be seen on the pedestal with the Single high-density BGA board: 30 ADC counts amplitude which corresponds to 3 MIPs; but nothing appears on the FlipChip board. After having optimized the grounding and the digital power supply decoupling, the modulation has been reduced by a factor 10. This modulation comes from digital current spikes on preamplifier ground node. Due to the values of the input capacitor (50 pF) and the preamplifier gain, there is a gain of 100 on the ground node, therefore 10 µV on ground gives 1 ADC count. This is the reason why the PCB must have a very low ground impedance. Furthermore, the decoupling capacitance on the digital power supply is critical. Detailed studies have been carried out on this digital coupling and have provided recommendations for hexaboard designs.

4. Timing measurement

4.1. Time walk profile

The minimum providing a ToA measurement is around 20 fC which is limited by the digital coupling. A Time Walk of 5 ns is achieved with 50 pF input capacitance.

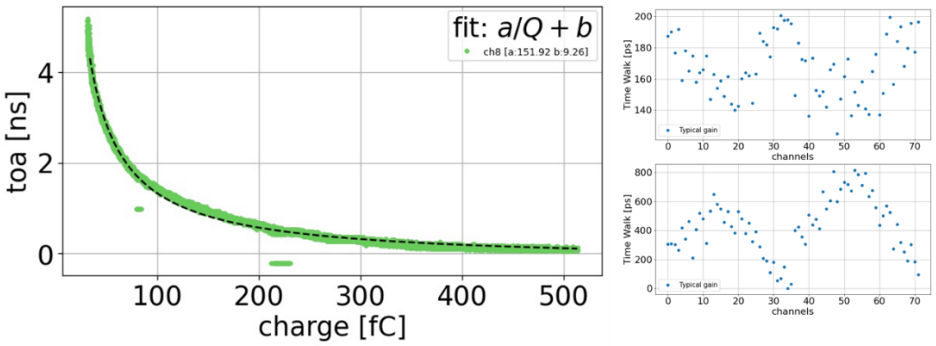

Figure 7: Timing analysis. Left: time walk profile. Top right: 'a' parameter of the fit. Bottom right: 'b' paramater of the fit.

The Time Walk profile can be fitted in order to find the parameters needed to make the timing reconstruction. Figure 7 shows the distribution of the fit parameters and the pattern is due to the fact that the TDC clocks are coming from the middle of each half.

4.2. Timing resolution

Regarding the ToA jitter, and so the timing resolution, between 90 ps and 150 ps resolution is achieved for the lower charge around 20 fC. Above 100 fC, the resolution goes down around 22 ps.

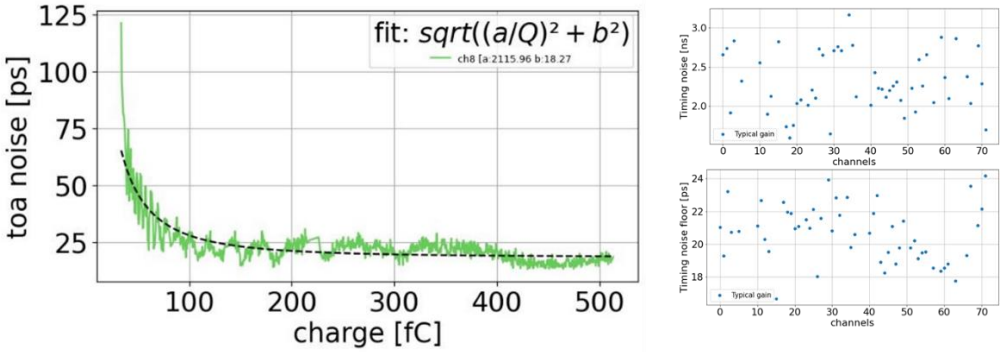

Figure 8: ToA jitter. Left: ToA jitter vs. charge. Top right: 'a' parameter of the fit. Bottom right: 'b' parameter of the fit.

5. SiPM version

The SiPM-version is roughly the silicon-version with an input current conveyor placed in front of the preamplifier. This conveyor is based on a low impedance common gate structure, an input DAC for the overvoltage adjustment to tune the SiPM gain, and a stage of current mirror to adapt the conveyor gain. Some others minor changes have been done on the preamplifier to make it compatible with the 2.5 power supply voltage of the conveyor.

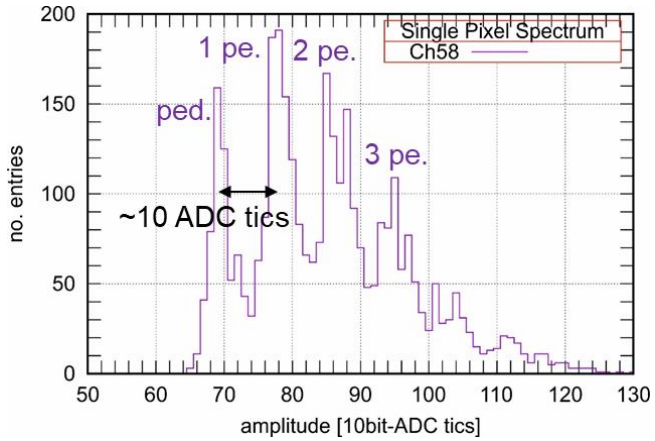

Figure 9: Single-Photon Spectrum (SPS)

This chip has been tested on the tileboard and in beam test. The Single-Photon Spectra has been measured by injecting light from a LED implemented on the tileboard, a gain around 10 ADC counts per fired pixels has been obtained as shown in Figure 9.

6. Trigger path

An excellent agreement between charge from the DAQ path and compressed charge from the trigger path has been found. This is requested to find all the parameters needed for the trigger path processing: the adc pedestal value for each channel, the tot pedestal, the tot threshold, the multiplication factor, etc.

7. Temperature mitigation, TID and SEE

Since the full analog chain is DC coupled from the preamplifier input to the shaper output, a temperature sensitivity mitigation is mandatory. The bandgap value moves over 2 mV between -40°C and +40°C, but with a lower shift around -20°C. The pedestals move around 1 ADC count per degree at room temperature, but around 0.4 ADC count per degree around -30°C. It would be around 5 ADC count per degree without mitigation. The gain does not change with temperature.

Both Silicon version and SiPM version have been tested in TID campaigns: one at room temperature up to 300Mrad on the silicon chip, another one at cold, around 0°C. And one TID campaign has been performed on the SiPM chip up to 5 Mrad at room temperature. In Figure 10, measurements from the second silicon-version campaign at cold are shown: the PLL frequency as a function of the VCO input voltage is shown on the left, there is a higher shift after 1 Mrad but that still works up to 300 Mrad. The ADC as a function of the charge after 300 Mrad is shown on the right: no visible difference with the pre-rad behavior is visible. The chip still works after irradiation, and for some blocks which have shown some weakness after 100 Mrad at room temperature, they recover good behavior after annealing.

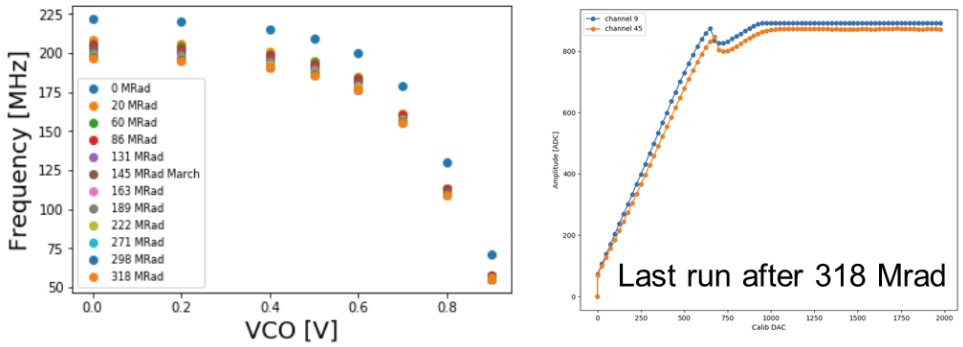

Figure 10: TID results at cold on the Si-version. Left: PLL freq. vs. VCO voltage. Right: ADC vs. charge after 300 Mrad.

The chip has been tested twice against Single Event Effect (SEE) at Louvain with heavy ions. The I2C registers react as expected with a flip rate below 2E-7 Hz/chip. Some bits shifts have been observed in both memory pointers and DAQ links, but it is not surprising because these parts have not been triplicated in this version of the chip.

8. Conclusion

With HGCROC2, a big step has been taken to reach the HGCAL's requirements. It is probably one of the most complex chips ever designed for imaging calorimetry: high dynamic charge, precision timing measurements, high speed links, a lot of digital, harsh radiation environment.

Measurements are now well advanced on both Silicon and SiPM versions. Charge performance reaches the specification: 1-2 % linearity, for both ADC and TOT, noise below specifications. Timing performance are at the state of the art: less than 25 ps resolution, time walk correction is feasible. Some changes need to be implemented on the next version to the TDC to remove outliers which have been reproduced in simulation. A digital coupling appears on the analog waveforms when the chip is packaged in high-density BGA and mounted on a PCB. By improving the ground impedance and the digital decoupling capacitors, this modulation has been reduced from 30 ADC counts down to 3 ADC counts, which is acceptable. TID and SEE campaigns show very good performance: SEE appear only in the non-triplicated parts of the chip, and there is no SEU errors on the slow-control parameters.

The final version of the chip has been submitted on December 2020. Minor changes have been done in analog part, the largest changes are in digital, where all the crucial parts have been triplicated, both DFFs and clocks. Another memory has been added to handle the triggered data from the DAQ. Hamming encoding has been implemented for all data stored in the two memories. CRC-32 checksum encoding has also been implemented and sent in the DAQ frame.

References

[1] CMS Collaboration, The Phase-2 Upgrade of the CMS Endcap Calorimeter, CERN-LHCC-2017-023, 2017

[2] M. Firlej *et al.*, *A fast, ultra-low and frequency-scalable power consumption, 10-bit SAR ADC for particle physics detectors*, JINST 10 P11012, 2015

[3] A. Marchioro et al., Design of power-efficient RAM memories for LHC Front-End chips, Contribution to the Front-End Electronics Conference, 2018.