Design and Performance of the CMS Pixel Detector Readout Chip

H.Chr. Kästli [a,*], M. Barbero [b,a], W. Erdmann [a], Ch. Hörmann [d,a], R. Horisberger [a], D. Kotlinski [a], B. Meier [c]

[a] *Paul Scherrer Institut, 5232 Villigen PSI, Switzerland*
[b] *Institut für Physik der Universität Basel, 4056 Basel, Switzerland*
[c] *Institut für Teilchenphysik, ETH Zürich, 8093 Zürich, Switzerland*
[d] *Physik-Institut der Univerität Zürich, 8057 Zürich, Switzerland*

Abstract

The readout chip for the CMS pixel detector has to deal with an enormous data rate. On-chip zero suppression is inevitable and hit data must be buffered locally during the latency of the first level trigger. Dead-time must be kept at a minimum. It is dominated by contributions coming from the readout. To keep it low an analog readout scheme has been adopted where pixel addresses are analog coded.

We present the architecture of the final CMS pixel detector readout chip with special emphasis on the analog readout chain. Measurements of its performance are discussed.

1. Introduction

CMS is a general purpose detector at the Large Hadron Collider (LHC) at CERN. Its innermost tracking device is the pixel detector, which consists of three barrel layers and two endcap disks both up and downstream[1]. Its basic building blocks are highly segmented silicon sensors with corresponding readout chips (ROCs). Each sensor segment (pixel cell) is connected to a charge sensitive amplifier in the corresponding pixel unit cell (PUC) of the ROC via an indium bump of about 20μm diameter. In the first barrel layer, the active area is 4.3cm from the interaction point. At this short distance the density of produced particles will be as high as $4 \cdot 10^7 \text{sec}^{-1}\text{cm}^{-2}$ at the full LHC design luminosity of $\mathcal{L} = 10^{34}\text{sec}^{-1}\text{cm}^{-2}$. Due to inclined tracks and Lorentz drift of the produced signal charge in the 4T magnetic field several pixels might be hit by a single track forming hit clusters. In order to improve position reconstruction for clusters through signal interpolation the amount of charge produced in each pixel is measured. The ROCs have to record position and charge for all hit pixels with a time resolution of 25ns, which is the time between two LHC bunch crossings. These information have to bo stored on-chip during the CMS first level trigger latency of 3.2μs after which they are either read out or discarded. The ROCs are read

---

* Corresponding author.
Email: hans-christian.kaestli@psi.ch

out serially via 40MHz analog links.

The first prototype of the readout chip has been developed in the radiation hard DMILL process [2,3]. In order to improve performance and yield and to reduce costs, the design has been migrated to a commercial process with much smaller feature size[4]. The second iteration of the translated chip, PSI46V2, is going to be the production version and is described in this article.

In section two the architecture of PSI46V2 is described. Section three focusses on the analog readout scheme and in section four measurements of the performance are presented.

2. Chip architecture

The readout chip PSI46V2 has been fabricated in a commercial 5-metal layer $0.25\mu m$ process available through CERN. The design has been made radiation tolerant by following special layout rules as proposed in [5]. The chip integrates 1.3 million transistors in an area of 7.9mm× 9.8mm. It can be divided into 3 functional blocks: a control and supply block in the chip periphery, an array of pixel unit cells organized in double columns, and the double column peripheries which control readout and trigger validation within double columns. The total number of pixels is 80×52 with a pixel size of $100\mu m$ ×$150\mu m$.

2.1. *Control and supply block*

The chip periphery houses various control and supply circuits. It contains

– A serial programming interface. This is an I^2C-like protocol, modified to run at a speed of 40 MHz. To accommodate this high speed the possibility to read back configuration data had to be given up. Nevertheless, there is a limited readback possibility through the analog data stream as will be shown below. Two low voltage differential signal (LVDS) pairs are needed for clock and data lines.

– A fast signal decoder. First level triggers and commands for reset and calibration signal injection are coded into a single LVDS signal. This signal has to be decoded and distributed over the chip.

– 21 8-bit digital-to-analog converters (DACs), five 4-bit DACs and one 3-bit DAC to adjust offsets, gains, thresholds, supply voltages, timings, etc.

– 2 control registers to set the trigger latency, readout speed (40 MHz/20MHz) and range for calibration pulses and to enable/disable the ROC.

– A bandgap voltage reference and 6 voltage regulators. Three of them can be wire-bonded to external filter capacitors. This leads to a good immunity against power ripple and reduces chip-to-chip cross talk. The ROC needs two external power supplies, 1.5V for the analog section and 2.5V for digital part. It consumes a total of about 120 mW which corresponds to only $29\mu W$ per pixel. The voltage regulators are programmable and hence the voltages can be set for each chip separately.

– An analog event generator. This is a circuit that collects the pixel hit information from the double columns and generates the output data stream as described in section 3.

– An on-chip temperature sensor for monitoring cooling performance.

– A cluster multiplicity counter. This is a fast trigger signal that could be used by the CMS first level trigger or for self-triggering in the lab when no external trigger is available. Two thresholds can be set for this mechanism: one sets the minimal number of hits within a double column below which hits are considered as background and are ignored, whereas the other one tunes the number of hit double columns above which a trigger signal is issued. Note, that even below this threshold a signal proportional to the number of clusters is available.

2.2. *Pixel unit cell*

A sketch of the PUC is shown in figure 1. It can be divided into an analog part (top) and the digital logic (bottom). The thick lines on the right are bus lines running along the double column.

The signal from the sensor enters a two stage charge sensitive pre-amplifier/shaper system. Al-
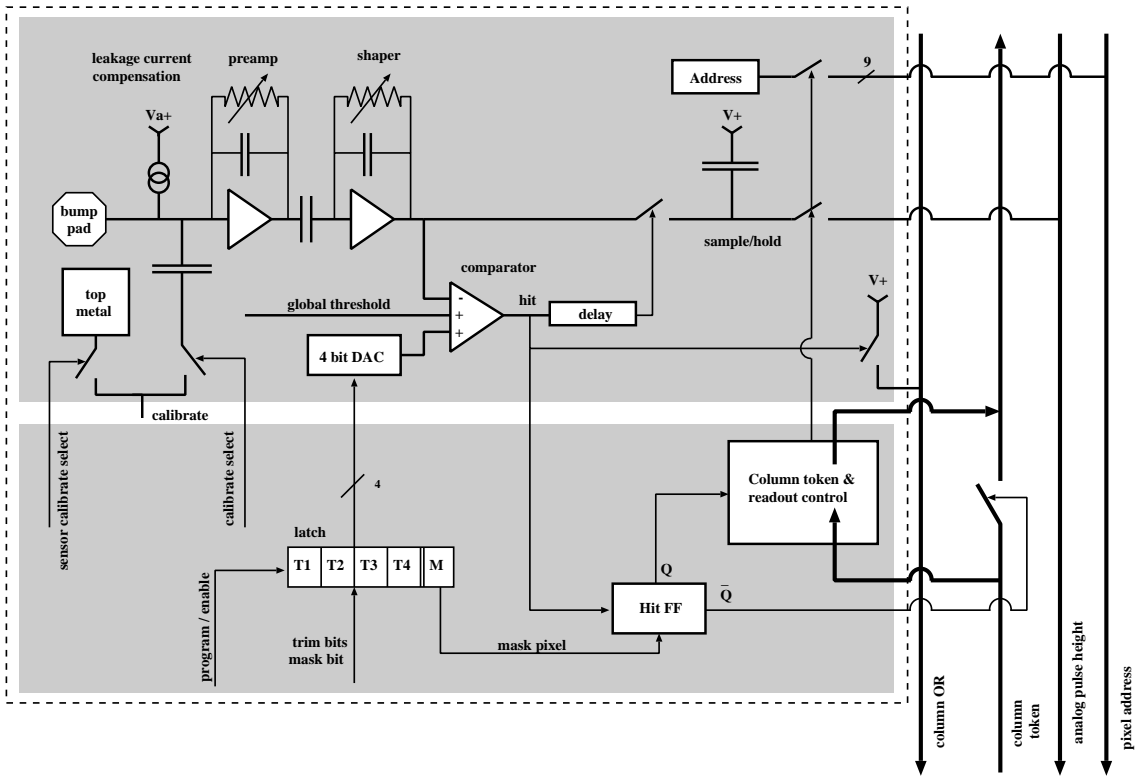
Fig. 1. Schematic view of a pixel unit cell

ternatively, calibration signals can be injected through either a 4.8fF injection capacitor connected directly to the amplifier input node, or via the sensor through the air gap between a top metal plate in the ROC and the sensor. The former is used for trimming the comparator threshold while the later can be used to test the bump bond connections[6]. A globally programmable current source at the input node compensates for sensor leakage current.

Zero suppression is performed with a comparator. A global threshold can be programmed for the whole chip. In order to compensate for local transistor mismatches each pixel has a 4-bit DAC to trim the threshold. Furthermore, a mask bit allows to disable noisy pixels. Once the comparator is above threshold the shaper output signal is stored in a sample-and-hold circuit (see below). The double column periphery is notified immediately through a fast hard-wired column OR. The pixel becomes insensitive and waits for a column readout token. When it arrives the analog signal is sent to the periphery together with the pixel row address. The token is then passed on and the pixel resumes data taking. Thus, dead-time is short but depends on the hit rate. The time needed for a column scan to finish is $\leq 50\text{ns} + (50\text{ns}\times$ number of hits).

The shaper output signal is sampled after a fixed time delay which is controlled by a global register (i.e. once per ROC). The signal within the sample-and-hold circuit follows the rising edge of the shaper output quite fast while it decays much slower. This makes the peak of the signal sufficiently flat and the circuit is not very sensitive to this time delay. Due to timewalk and device mismatch the exact peaking time depends somewhat on the pulse height and varies a bit from pixel to pixel. The relative error in the measured pulse height due to a non perfect sampling point has been measured and is shown in figure 2. If the delay is optimised for low signal charges (3500
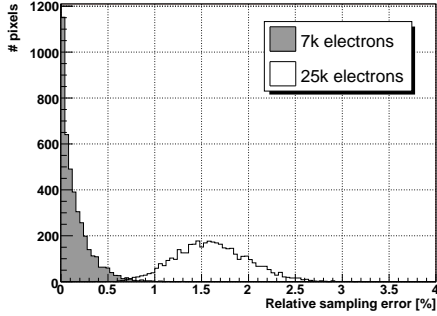
Fig. 2. Relative error of peak detection across a full chip for 2 different pulse heights

electrons in figure 2), the relative error for large signals is only a few per cent.

Downloading the full detector configuration at run start will take $\lesssim 0.5$ minute [7]. If needed, pixels can be reprogrammed during LHC machine orbits reserved for sub-detector calibration, however only at a limited rate. Thus single event upset (SEU) in the trim and mask bit storage cells has to be considered. A simple trick has been adopted to protect storage cells from SEU. This is illustrated in figure 3. If node (a) collects a large enough amount of charge, produced by a nuclear reaction, the stored logic level can be changed. The presence of the capacitor aided by Miller effect increases the critical charge and thus reduces the probability of an upset.

The SEU cross section for both protected and unprotected storage cells has been measured at PSI for 300 MeV/c pions. The result is $\sigma = 2.6 \cdot 10^{-16} \text{cm}^2$ per storage cell for protected and $\sigma = 2.4 \cdot 10^{-14} \text{cm}^2$ for unprotected cells. This is an improvement of 2 orders of magnitude. In
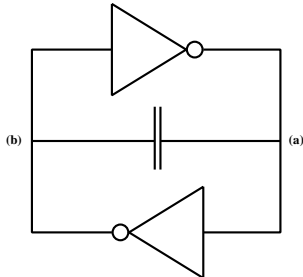

Fig. 3. SEU protected storage cell

PSI46V2 all storage cells in the PUCs are protected (4 trim bits and one mask bit). In case of an SEU the flipped bits have to be reprogrammed through control links. Each link serves several modules/pannels. In the worst case one link controls 4 modules plus 4 half-modules in each of the two innermost barrel layers, i.e. a total of 192 ROCs or 96 ROCs per layer. At high luminosity LHC operation the expected SEU rate for this control link is $< 3 \cdot 10^{-2}$Hz. This means that in about 8 hours 0.1% of the pixels changed their threshold. The occupancy of each pixel can be monitored online and pixels that show significant changes will be reprogrammed.

2.3. Double column periphery

The double column periphery controls the transfer of hit information from the pixels to the storage buffers (column drain mechanism) and performs trigger verification. A schematic view of the logic is shown in figure 4.

When the asynchronous fast column OR signal arrives at the periphery three actions are performed:

1. The present value of a bunch crossing counter (WBC) is latched into one of 12 time stamp buffer cells. This is needed later on for trigger validation.
2. The hits for this bunch crossing are acknowledged by the periphery. This tells the pixels to associate any later hits with another column drain and its corresponding bunch crossing. There are one active and up to two pending column drains allowed. Hits of any further bunch crossing are lost.
3. A column drain is initiated by sending out a readout token to the first pixel.

A pixel without a hit belonging to the active column drain just bypasses the token. This has been measured to happen at a rate of about 3.3 GHz, i.e. the token arrives at the first hit pixel within less than 50ns or 2 clock cycles. Hit information is transferred in parallel into the data buffer, one pixel per two clock cycles. The data buffer has a depth of 32 units. Each unit consists of a marker bit one analog and nine digital storage cells for pulse
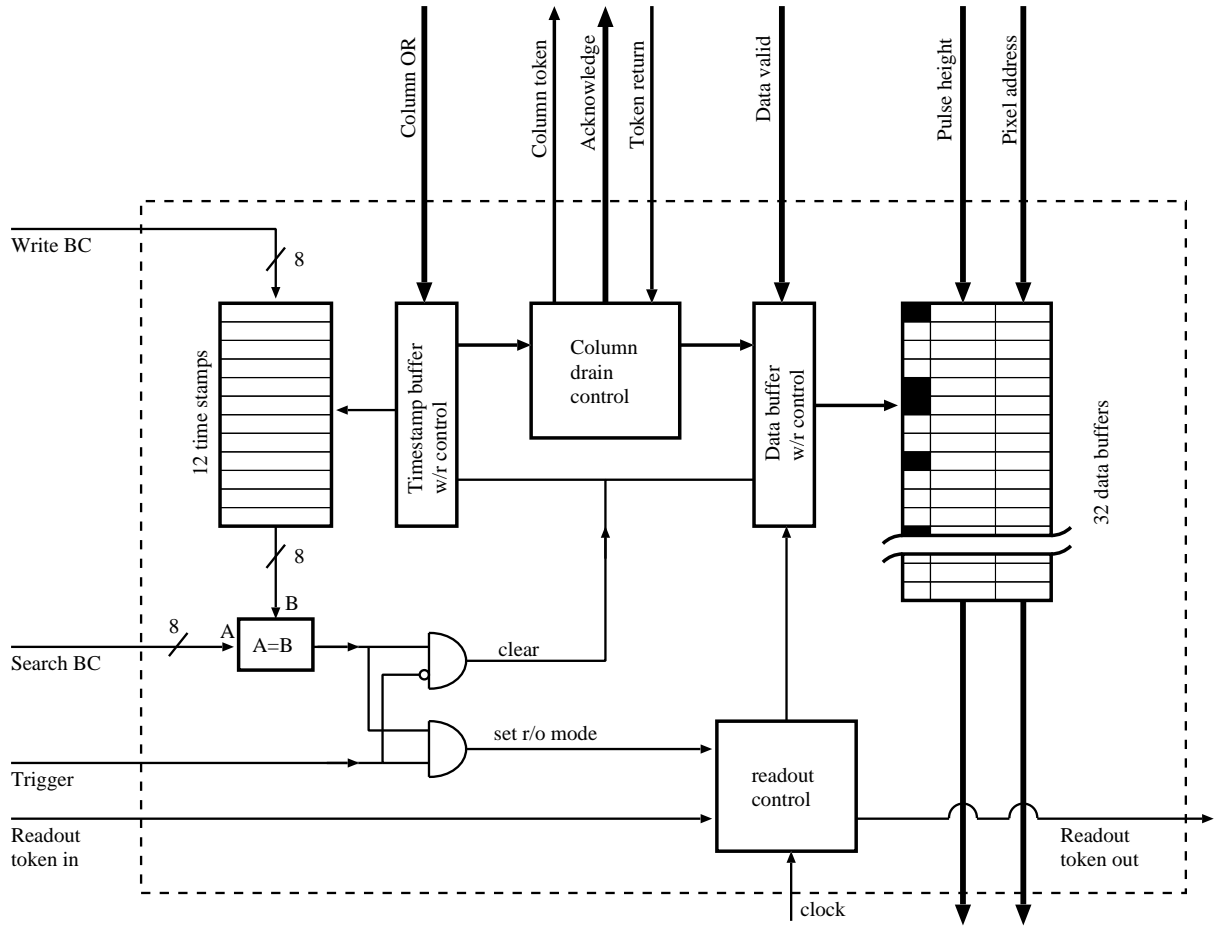
4

Fig. 4. Schematic view of the double column periphery

height and row address respectively. The marker bit indicates the beginning of a new event and thus is used to synchronize entries in the time stamp and the data buffers.

The oldest entry in the time stamp buffer is permanently compared to a second bunch crossing counter (SBC) which is delayed with respect to the WBC by a programmable amount which corresponds to the first level trigger latency. In case of agreement the time stamp is deleted and the presence of the CMS first level trigger signal is checked. Either the corresponding data is discarded or the column is set into readout mode. In this mode, the double column stops data acquisition in order to prevent overwriting of valid data. It waits for the readout token to arrive and sends data to the chip

periphery. Afterwards the double column resets itself and hence cannot have further valid hits for the duration of the trigger latency. Data loss also occurs when one of the buffers is completely filled up. In case of a full data buffer a double column gets reset. If the time stamp buffer is full, data acquisition is paused until the next buffer cell is freed.

3. Readout scheme and chip readout format

A couple of design inherent data loss mechanisms have been mentioned so far. They are summarized in table 3 for barrel modules, a luminosity of $10^{34} \cdot \text{sec}^{-1}\text{cm}^{-2}$ and 100 kHz first level trigger

rate. A large contribution, labeled 'readout', comes from the fact that a double column is insensitive between the time it validates a trigger and the time it finishes the readout. This stresses the importance of having short readout times. Therefore an analog readout scheme running at 40 MHz (alternatively it can be switched to 20 MHz) has been adopted where the pixel address is coded into 6 discrete levels. Pixel address and signal pulse height can be transferred in 6 clock cycles.

In the endcap part of the detector several sensor tiles of different size are organized in so called blades which contain 24 ROCs on the front side and 21 ROCs on the back side. In the barrel, sensors connected to 16 (8) ROCs are called modules (half modules). The ROCs of each side of a blade or of a (half) module are read out in a daisy chained way. A token bit, controlled by a special token bit manager chip (TBM,[8]), is sent to the first ROC. The ROC sends its data to the TBM and passes the token bit to the next ROC, until the token bit gets back to the TBM. The TBM amplifies the signals from the ROCs, adds a header and a trailer and drives the signals out to the detector supply tube. The readout of a single pixel hit is shown in figure 5. It starts with a header of 3 clock cycles. A large negative signal level well outside of the range of pixel data (ultra-black) followed by a zero differential level (black) separates the individual ROCs in the blade/module data stream. No other identification is sent. In order to get the ROC ID the DAQ
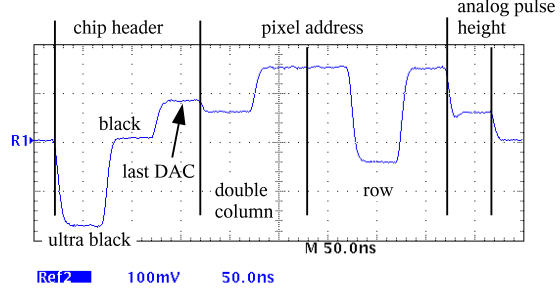

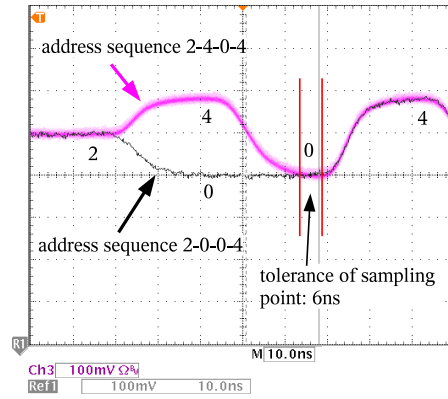Fig. 5. Readout sequence of a ROC with one pixel hit

Fig. 6. Readout signal output of a barrel module for two different address sequences

unit has to count the chip headers. In the third clock cycle of the chip header a signal inversely proportional to the value of the last addressed DAC is sent. This is the only way information about the configuration of the ROCs can be read back. It follows the double column address (2 cycles), the row address (3 cycles) and the analog pulse height for each hit pixel. Figure 6 shows the signal output of a barrel module as it arrives at the electro-optical converter. Shown are two different address level sequences. As can be seen, the risetime of the signals is short enough to allow for a tolerance of the sampling point in the DAQ unit of $\approx$ 6ns. The measured sampling distribution for a ROC is shown in figure 7. The address levels are clearly separated and allow a reliable reconstruction of the pixel addresses.

Radius [cm]	4	7	11
pixel busy	0.21%	0.078%	0.044%
column drain busy	0.25%	0.020%	0.004%
time stamp buffer overflow	0.17%	0.001%	0
data buffer overflow	0.17%	0.081%	0.065%
1-buffer	1.0%	0.40%	0.26%
readout	1.0%	0.20%	0.16%
double column reset	1.0%	0.44%	0.26%
total	3.8%	1.2%	0.79%

Table 1
Summary of simulated data loss for barrel modules at a luminosity of $10^{34}\mathrm{sec}^{-1}\mathrm{cm}^{-2}$ and 100kHz trigger rate. The upper part shows column drain losses, in the lower part readout related data losses are summarized
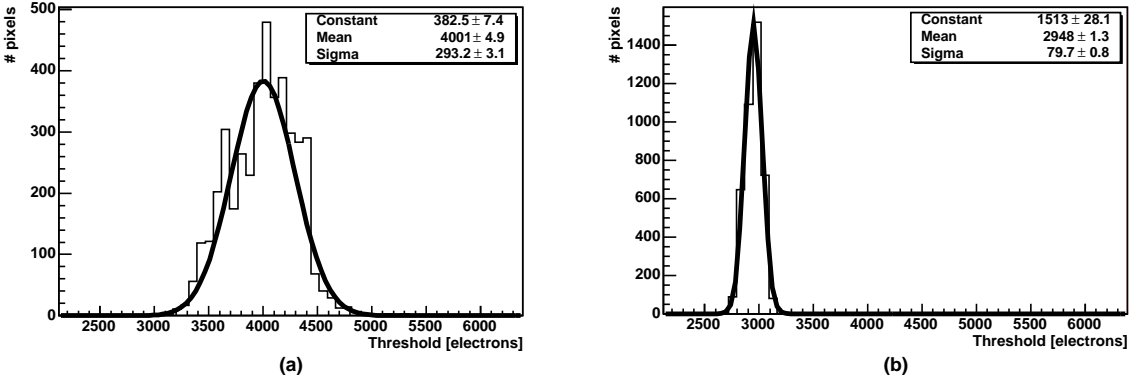
Fig. 9. Threshold distribution of a ROC before (a) and after (b) trimming
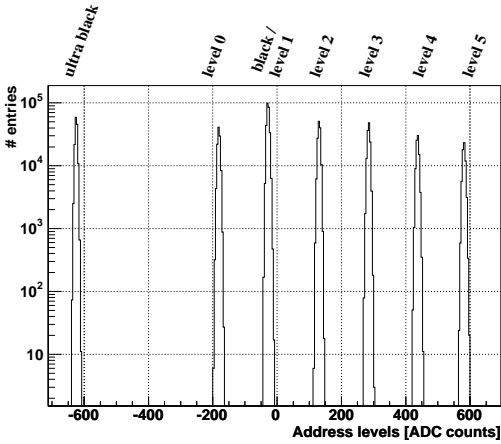

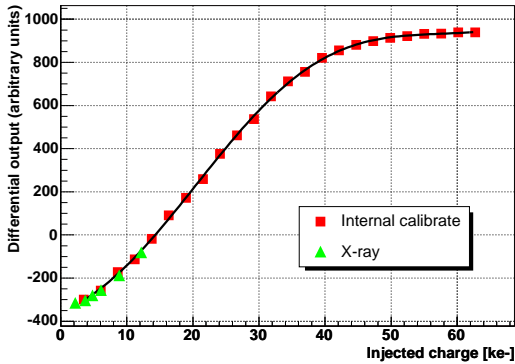Fig. 7. Histogram of the sampled address levels of a ROC. The levels are cleanly separated

Fig. 8. Analog signal transmission.

4. Measurements

In this section lab measurements of the performance of the ROC are shown. Figure 8 shows the pulse height response to calibration signal injection. It is slightly non-linear at low charges and saturates for high charges above about 2 MIPS. Injection capacitances have been calibrated with X-ray sources. K_α lines of 6 different isotopes between 8.04keV and 44.2keV have been used. These measurements are also shown in figure 8.

It is important to have a threshold distribution as uniform as possible. On one hand the threshold must be set well above the amplifier noise in order not to flood the system with random hits. On the other hand the position resolution depends almost linearly on the threshold (see [1]). The width of a gaussion fit to the untrimmed threshold distribution of a ROC is around 300 electrons as shown in figure 9 (a). Adjusting the trim bits is one of the most important steps in the module characterization process [6]. After trimming, the threshold dispersion can be improve considerably to typically 80 electrons as shown in figure 9 (b). The intrinsic noise of the preamp/shaper has been measured to be $\lesssim$ 190 electrons [6].

The threshold described so far is not directly relevant. The time between a particle crossing the sensor and the comparator going above threshold depends on the created ionization charge. Therefore it might happen that for low pulses the hit
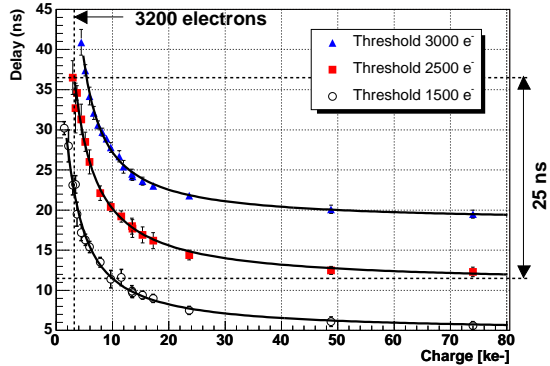
Fig. 10. Timewalk for 3 different thresholds. The offsets of the 3 curves are arbitrary for better readability

signal is delayed to the next bunch crossing and hence is lost during trigger validation. This time-walk can be measured by delaying the injected calibration pulses with respect to the chip clock. The result of such scans is shown in figure 10 for three different thresholds. For an absolute threshold of 2500 electrons it is shown that the effectively usable threshold (in-time threshold) is at about 3200 electrons, i.e. the time difference between a signal of 3200 electrons and very high charges above 100k electrons is below 25ns.

Test structures containing the analog section of the PUC have been irradiated with gamma rays up to a total dose of 13.2 Mrad. This corresponds to the absorbed dose in 2 years at high luminosity LHC operation for the 7cm barrel layer or 1 year for the innermost layer. Figure 11 shows the outputs of the shaper and the sample-and-hold circuit in transparent mode. Curves (a) show the response for unirradiated structures. After irradiation, the analog signals become somewhat slower (b), but can be brought back close to the original waveform by reprogramming the feedback settings (c).

5. Summary

The final version of the CMS pixel readout chip has been designed in a $0.25\mu m$ process. It is fully functional and measurements of the performance have been presented. The threshold can be tuned
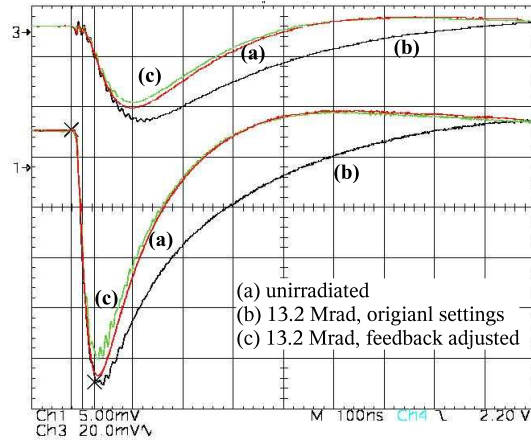

Fig. 11. Output of sample-and-hold circuit (top) and shaper output (bottom) before and after irradiation.

to values below 2500 electrons with a uniformity of 80 electrons. Due to timewalk the in-time threshold is somewhat higher at 3200 electrons. The readout uses an analog scheme with address levels coded into 6 levels. This has been proven to work reliably at 40 MHz as address decoding was 100% correct.

References

[1] The CMS collaboration, CMS Tracker Technical Design Report, CERN/LHCC 98-6

[2] M. Dentan et al., IEEE Trans. Nucl. Sci. 43 (1996) 1763

[3] M. Barbero et al., Design and test of the CMS pixel readout chip, Nucl. Instr. and Meth. A517 (2004) 349

[4] W. Erdmann, The $0.25\mu m$ front-end for the CMS pixel detector, Nucl. Instr. and Meth. A 549 (2005) 153

[5] W. Snoeys et al., Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip, Nucl Instr. and Meth. A439 (2000) 349

[6] A. Starodumov et al, Qualification Procedures of the CMS Pixel Barrel Modules, these proceedings

[7] D. Kotlinski, The Control and Readout System of the CMS Pixel Barrel Detector, these proceedings

[8] E.Bartz, The $0.25\mu m$ Token Bit Manager Chip for the CMS Pixel Readout, Proceedings of the 11th Workshop on Electronics for LHC and Future Experiments, Heidelberg, Germany