Development of the monolithic "MALTA" CMOS sensor for the ATLAS ITk outer pixel layer


a University of Glasgow, United Kingdom
b CERN, Switzerland
c University of Bonn, Germany
d Jozef Stefan Institute, Slovenia
e University of Oxford, United Kingdom
E-mail: bojan.hiti@ijs.si

Depleted Monolithic Active Pixel Sensors (DMAPS) are an option for the outermost layer of the upgraded ATLAS ITk Pixel Detector at the CERN LHC. Two large size DMAPS named TJ MALTA and TJ Monopix were produced in a 180nm CMOS imaging process in a small collection electrode design. The TJ MALTA chip combines a low power front end with a novel matrix readout design to achieve a low power consumption of < 80mW/cm². Threshold values of ≈ 250e⁻ with a dispersion of ≈ 30e⁻ and an ENC of < 10e⁻ can be achieved before irradiation which is consistent with the results from TJ Monopix. Test beam measurements indicate an average efficiency of 96% before irradiation, with the inefficiency mainly due to a non fully functional masking scheme, forcing operation at relatively high thresholds. After neutron irradiation to $1 \times 10^{15} \text{neq/cm}^2$ the efficiency in pixel centres is retained, but it is reduced in pixel corners. A proposal to improve charge collection in the corners is backed up by TCAD simulations and promises an improved performance with small modifications.
1. Introduction

The ATLAS experiment [1] at the Large Hadron Collider (LHC) at CERN actively investigates Depleted Monolithic Active Pixel Sensors (DMAPS) produced in CMOS technology as an option for the outermost pixel layer of the Phase-II upgrade Inner Tracker detector (ITk) [2] and develops a CMOS sensor conform to specifications, including radiation tolerance to non-ionising energy doses of $10^{15} \text{neq/cm}^2$ and a total ionising dose of 80MRad, a timing resolution of 25ns, and hit rates up to 10MHz/cm$^2$.

After promising results on prototypes [3], two large scale DMAPS with almost identical pixel geometry and analogue front end derived from the ALICE ITS chip ALPIDE [4, 5] were produced in a 180nm imaging CMOS process modified to fully deplete the sensitive layer [6]: TJ MALTA uses a novel asynchronous pixel matrix readout and TJ Monopix a more common column-drain architecture. This work presents first results with the MALTA chip.

2. Sensor and front end design

The MALTA contains a $512 \times 512$ pixel matrix with a pixel size of $36.4 \mu\text{m} \times 36.4 \mu\text{m}$, and a small collection electrode to achieve a low pixel capacitance, low noise and fast timing at minimal power consumption. No clock is distributed over the pixel matrix to further reduce digital power consumption. The sensitive volume is a $25 \mu\text{m}$ thick, highly resistive ($>1\,k\Omega\text{cm}$) p-type epitaxial layer with a n-type collection electrode of $\approx2\,\mu\text{m}$ in diameter and a capacitance of $<5\,\text{fF}$, readout circuit included (Figure 1). A deep p-well contains the in-pixel circuitry and shields the n-wells with PMOS transistors from the epitaxial layer preventing them from collecting charge parasitically, and also shields the sensor from digital signals. The sensor bias is limited to $-20\,\text{V}$ on the substrate and $-6\,\text{V}$ on the deep p-well. The fabrication process was modified using an additional low dose n-implant to ensure sensor depletion over the entire pixel area at low bias voltages [6].

![Figure 1: MALTA pixel cross section [7]: (a) Standard process. (b) Modified process.](image-url)

The MALTA front end with masking and analogue test pulse injection is derived from the ALPIDE front end adapted to the 25ns timing requirement of ATLAS. Simulations predict a threshold level of $300\,\text{e}^{-}$ with a dispersion of $10\,\text{e}^{-}$ and an ENC of $10\,\text{e}^{-}$, sufficiently low to avoid the need for an in-pixel threshold adjustment circuit [7]. The pulse height information can be obtained from the time walk between the bunch crossing time and the leading edge of the discriminator signal with a simulated 25ns in-time threshold of $300\,\text{e}^{-}$. With an analogue bias current of 500nA per pixel, the analogue and digital power consumption are $70\,\text{mW/cm}^2$ and $10\,\text{mW/cm}^2$ respectively.
The asynchronous matrix readout is based on double columns organised into groups of $2 \times 8$ pixels sharing the same address. A hit group promptly transmits the hit position to the periphery via a parallel data bus. Alternating groups of pixels in the same double column share two separate data buses. The signals have an adjustable duration of $0.5\text{ns}$–$2\text{ns}$ and a maximal latency of $8\text{ns}$ for propagation along the double column. The end-of-column circuit merges the signals onto a common data bus. An arbitration circuit, called hit merger, is implemented to time sort multiple simultaneous signals to avoid data collisions and data loss. The final $40$ bit word contains hit position, column address, time stamp and hit merger delay bits and is asynchronously transmitted off-chip on a parallel, $40$ bit wide bus, using LVDS drivers. The LVDS driver with programmable output current and pre-emphasis has been proven in another test circuit to work up to $5\text{Gb/s}$.

The first MALTA chips correctly propagate the signals to the output, but need to be operated at a higher charge threshold than desired because at present a non fully functional slow control prevents us from masking pixels individually. The presence of a few (several tens out of $250000$ before irradiation) very noisy pixels creates too high an activity level in the chip and in the merger at lower thresholds. The best way to operate the chip in this situation was to disable the merger, accepting data collisions on the bus with the threshold sufficiently high to not completely saturate the output bandwidth. Such data collisions, or two simultaneous hits on the bus, will be interpreted as a hit in another part of the matrix, and as an inefficiency. This will be fixed in a next submission to enable operation at lower thresholds. We are also considering to include a per pixel threshold adjust in future prototypes.

3. Measurements

Threshold measurements were carried out with unirradiated chips and samples irradiated with neutrons at Ljubljana TRIGA reactor [8]. Pixel threshold and noise were extracted using electrical charge injection. Due to the inability to mask individual noisy pixels the measurement in MALTA is challenging, as the threshold cannot be tuned to a specific noise occupancy. Different arbitrarily selected threshold settings were used in the measurements instead. Before irradiation the lowest threshold setting has a value of $230\text{e}^{-}$ with a dispersion of $36\text{e}^{-}$ and an ENC of $7\text{e}^{-}$. In a $5 \times 10^{14} \text{n}_{\text{eq}}/\text{cm}^{2}$ neutron irradiated sample the threshold increases to $500\text{e}^{-} \pm 70\text{e}^{-}$ with an ENC of $12\text{e}^{-}$. This is comparable to the values reported for TJ Monopix [9]: $271\text{e}^{-} \pm 31\text{e}^{-}$ (ENC $11\text{e}^{-}$) before irradiation and $470\text{e}^{-} \pm 50\text{e}^{-}$ (ENC $20\text{e}^{-}$) after neutron irradiation. Both chips also show a non-Gaussian noise spectrum with entries at higher noise values, caused by random telegraph noise in the transistors. The efficiency of MALTA was measured in the test beam at CERN SPS with $120\text{GeV}$ pions. Tracks were reconstructed with a reference telescope with a resolution of $3\text{\mu m}$ [10]. The highest efficiency was obtained when the substrate was biased to $-15\text{V}$ and the deep p-well to $-6\text{V}$. Figure 2 shows in-pixel efficiency maps with the pixel matrix projected into a $2 \times 2$ pixel area. Before irradiation (2a, threshold $250\text{e}^{-}$) an average efficiency of $96\%$ with a slightly reduced efficiency in pixel corners was measured. In a sample irradiated to $1 \times 10^{15} \text{n}_{\text{eq}}/\text{cm}^{2}$ the average efficiency reduced to $74\%$ (2b, threshold $350\text{e}^{-}$), with the inefficient area in the corners increasing in size. The central area closer to the collecting electrode maintained a high efficiency.

The reduced efficiency in the pixel centre is caused by data loss due to the deactivated merger circuit: Figures 2b and 2c show a higher efficiency in the centres at a higher threshold, where the
data loss is lower. The inefficiency in the pixel corners is due to charge sharing and degraded lateral charge collection. The asymmetric efficiency distribution in the pixel corners correlates with the shape of the deep p-well and motivated potential improvements discussed in Section 4.

4. TCAD simulation and further improvements

The process modified to achieve full depletion of the epitaxial layer (Figure 1) yielded quite promising results after irradiation on 25 and 30 µm pitch prototypes [3, 6]. The larger pixel pitch in MALTA, almost 40 µm, causes severe post-irradiation efficiency loss at the pixel edges. Figures 3b and 3c show for a 2D TCAD simulation the signal induced on the collection electrode when a minimum ionizing particle traverses the sensor perpendicularly at the pixel corner, the worst position for timing and post-irradiation charge loss. Figure 3b clearly indicates a slower signal collection for larger pixel pitch causing more charge loss after irradiation. Two possible approaches are being considered to increase the lateral field and speed up the charge drift towards the collection electrode: an additional p-type implant at the pixel edges (Figure 3a), or creating a gap in the n-layer at the pixel edge. Figure 3c shows the speed improvement with the additional p-type implant for the 40 µm pixel pitch. The gap in the n-layer gives similar results. 3D TCAD simulations are presently ongoing for a detailed quantitative optimisation before fabricating new prototypes.

5. Conclusions

Two large size MAPS were produced with small, low capacitance collection electrodes. The TJ MALTA chip features a novel asynchronous matrix readout which in combination with low sensor capacitance results in a significantly reduced power consumption at a high sensor granularity. Measured threshold dispersion and noise are higher than predicted from simulations with further increase after irradiation. To compensate these effects an in-pixel threshold tuning may be required. This and a problem with the slow control forced us to operate the chip at higher charge thresholds, degrading detection efficiency. The degradation is especially significant in pixel corners after irradiation. Two pixel layout modifications were simulated and show a large improvement in charge collection, and will be implemented in further submissions along with the fix of slow control.
Figure 3: (a) Proposed pixel modification to improve charge collection with an additional p-type implant. (b) Simulated transient pulses in standard pixels for pixel sizes of 30 µm and 40 µm. (c) Simulated pulses for the standard pixel and a pixel with the p-type implant. Significant improvement in collection time is observed after modification, as well as in pixels of smaller size.

Acknowledgements

This project has received funding from the EU’s Horizon 2020 Research and Innovation programme under Grant Agreement no. 654168. We acknowledge the financial support from the Slovenian Research Agency (research core funding No. P1-0135 and project ID PR-06802).

References