A 1-μW Radiation-Hard Front-End in a 0.18-μm CMOS Process for the MALTA2 Monolithic Sensor

F. Piro, *Graduate Student Member, IEEE*, P. Allport, I. Asensi, I. Berdalovic, *Member, IEEE*, D. Bortoletto, *Member, IEEE*, C. Buttar, R. Cardella, E. Charbon, *Fellow, IEEE*, F. Dachs, V. Dao, D. Dobrijevic, M. Dyndal, L. Flores, P. Freeman, A. Gabrielli, L. Gonella, T. Kugathasan, M. LeBlanc, K. Oyulmaz, H. Pernegger, P. Riedler, M. van Rijnbach, H. Sandaker, A. Sharma, C. Solans, W. Snoeys, *Senior Member, IEEE*, T. Suligoj, *Member, IEEE*, J. Torres, *Senior Member, IEEE*, and S. Worm

***Abstract*— In this article, a low-power, radiation-hard front-end circuit for monolithic pixel sensors, designed to meet the requirements of low noise and low pixel-to-pixel variability, the key features to achieve high detection efficiencies, is presented. The sensor features a small collection electrode to achieve a small capacitance (<5 fF) and allows full CMOS in-pixel circuitry. The circuit is implemented in the 180-nm CMOS imaging technology from the TowerJazz foundry and integrated into the MALTA2 chip, which is part of a development that targets the specifications of the outer pixel layer of the ATLAS Inner Tracker upgrade at the LHC. One of the main challenges for monolithic sensors is a radiation hardness up to 10^{15} 1-MeV n_{eq}/cm^2 non-ionizing energy loss (NIEL) and 80 Mrad total ionizing dose (TID) required for this application. Tests up to $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 100 Mrad were performed on the MALTA2 sensor and front-end circuit, which still show good performance even after these levels of irradiation, promising for even more demanding applications such as the future experiments at the high-luminosity large hadron collider (HL-LHC).**

***Index Terms*— Front-end circuits, monolithic active pixel sensors (MAPSs), pixel detectors, radiation hardness.**

Manuscript received April 2, 2022; accepted April 25, 2022. Date of publication April 27, 2022; date of current version June 16, 2022. This work was supported by the European Unions Horizon 2020 Research and Innovation Program under Grant 675587 and Grant 101004761.

F. Piro is with the Department of Experimental Physics, CERN, 1211 Geneva, Switzerland, and also with the Advanced Quantum Architecture Laboratory (AQUA), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: francesco.piro@cern.ch).

P. Allport and L. Gonella are with the Department of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: allport@cern.ch; laura.gonella@cern.ch).

I. Asensi, F. Dachs, V. Dao, D. Dobrijevic, M. Dyndal, L. Flores, P. Freeman, A. Gabrielli, T. Kugathasan, M. LeBlanc, H. Pernegger, P. Riedler, M. van Rijnbach, A. Sharma, C. Solans, and W. Snoeys are with the Department of Experimental Physics, CERN, 1211 Geneva, Switzerland (e-mail: ignacio.asensi@cern.ch; florian.dachs@cern.ch; valerio.dao@cern.ch; dominik.dobrijevic@cern.ch; mateusz.dyndal@cern.ch; leyre.flores.sanz.de.acedo@cern.ch; moriishifreeman@gmail.com; andrea.gabrielli@cern.ch; thanushan.kugathasan@cern.ch; matt.leblanc@cern.ch; heinz.pernegger@cern.ch; petra.riedler@cern.ch; milou.van.rijnbach@cern.ch; abhishek.sharma@cern.ch; carlos.solans@cern.ch; walter.snoeys@cern.ch).

I. Berdalovic was with the Department of Experimental Physics, CERN, 1211 Geneva, Switzerland. He is now with the Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia (e-mail: ivan.berdalovic@fer.hr).

D. Bortoletto is with the Department of Particle Physics, University of Oxford, Oxford OX1 2JD, U.K. (e-mail: daniela.bortoletto@physics.ox.ac.uk).

C. Buttar is with the Faculty of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: craig.buttar@glasgow.ac.uk).

R. Cardella was with the Department of Experimental Physics, CERN, 1211 Geneva, Switzerland. He is now with the Department of Nuclear and Particle Physics, University of Geneva, 1205 Geneva, Switzerland (e-mail: roberto.cardella@unige.ch).

E. Charbon is with the Advanced Quantum Architecture Laboratory (AQUA), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: edoardo.charbon@epfl.ch).

K. Oyulmaz is with the Department of Physics, Bolu Abant Izzet Baysal University, 14030 Bolu, Turkey (e-mail: kaan.yuksel.oyulmaz@cern.ch).

H. Sandaker is with the Department of Physics, University of Oslo, 0315 Oslo, Norway (e-mail: heidi.sandaker@fys.uio.no).

T. Suligoj is with the Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia (e-mail: tomislav.suligoj@fer.hr).

J. Torres is with the Department of Electronics Engineering, University of Valencia, 46010 Valencia, Spain (e-mail: jose.torres@uv.es).

S. Worm is with the Deutsches Elektronen-Synchrotron (DESY), 15738 Zeuthen, Germany (e-mail: steven.worm@cern.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNS.2022.3170729.

Digital Object Identifier 10.1109/TNS.2022.3170729

I. Introduction

MONOLITHIC active pixel sensors (MAPSs) constitute an attractive alternative to the more largely used hybrid pixel sensors for high-energy physics experiments. Their main advantage stems from the integration of the readout electronics and sensors in the same silicon die, avoiding the expensive fine-pitch bump bonding. MAPS, therefore, facilitate significantly the detector assembly and reduce the production cost. Without bump bonding, they also tend to offer a smaller pixel pitch and thus a better spatial resolution. Furthermore, the sensor capacitance can be made so small (<5 fF) that it can offer a higher voltage signal even with a reduced sensor thickness, that is, a lower generated charge. This leads to a better power-performance ratio, allowing a significant reduction of the material related to the powering and cooling of the detectors. A lower material budget reduces the probability of multiple scattering of the particles emerging from the interaction point, improving the impact parameter resolution and momentum resolution on the reconstructed tracks and the overall detection efficiency of a tracker. Encouraging results from CMOS sensor prototypes [1]–[3] prompted the development of large-scale CMOS sensors [4], [5] which would meet the requirements of the outer pixel layer of the ATLAS inner tracker (ITk) upgrade [6]. This entails a non-ionizing energy loss (NIEL) tolerance up to 10^{15} 1-MeV n_{eq}/cm^2,

a total ionizing dose (TID) tolerance up to 80 Mrad, and a timing response within 25 ns with a power density below 500 mW/cm^2. To this end, the 20.6 mm $\times$ 20.2 mm MALTA sensor was designed in the TowerJazz 180-nm imaging technology. The MALTA pixel matrix features a sensor with a small collection electrode, an open-loop charge-sensitive front-end, and a fast, low-power, asynchronous digital readout architecture [7]. Measurements on the chip showed a timing response within the specifications [8]; however, the efficiency was degraded in the pixel corners already after 10^{14} 1-MeV n_{eq}/cm^2 [9]. This was addressed in a subsequent small-scale prototype called Mini-MALTA [10] by modifying the sensor to enhance the lateral electric field and improve charge collection. The sensor modifications were then implemented in another large-scale prototype, MALTA2, along with improvements on the front-end circuit. This article presents the design of the MALTA2 front-end, able to cope with the ATLAS ITk outer pixel layer requirements. For this circuit, the main challenge is to amplify the generated charge with high pixel-to-pixel uniformity and low noise, which are the key features to set low charge thresholds to determine particle hits and obtain good detection efficiencies. Extensive characterization of the prototype is currently in progress and the first measurements, performed on samples irradiated up to $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 100 Mrad, are shown here.

II. Sensor

The cross section of the standard sensor in the TowerJazz 180-nm CMOS imaging technology is shown in Fig. 1(a). It implements a small collection electrode, defined by an n-well implant, which is located inside the sensing volume, typically a high-resistivity p-type epitaxial layer. The in-pixel circuitry is placed outside the collection electrode and it is shielded from it by a deep p-well, which avoids collection of the signal charge by parts of the circuit other than the designated collection electrode. The key features of the technology are therefore the deep p-well, which allows full CMOS in-pixel circuitry and the possibility to use different starting materials compatible with particle detection. The depletion volume within the sensor can be increased by increasing the reverse bias between the collection electrode and the surrounding p-well and p-type substrates. For visible light, the signal is generated within a depth of a few microns, whereas high-energy particles generate charge over the full thickness of the silicon ($\sim$60 electron/hole pairs per micrometer traversed [11]) which needs to be collected within the time resolution of the event reconstruction. For the TowerJazz standard sensor of Fig. 1(a), the epitaxial layer is only partially depleted and the signal charge generated outside the depletion area is collected by diffusion (with a collection time of $\sim$100 ns). The ATLAS experiment has more stringent timing requirements since the particle hits have to be associated with different bunch crossings, 25 ns apart from each other. To obtain a faster response, the depletion zone needs to be extended over the whole sensitive layer. The objective is to push the charge carriers toward the collection electrode by drift and thus reduce

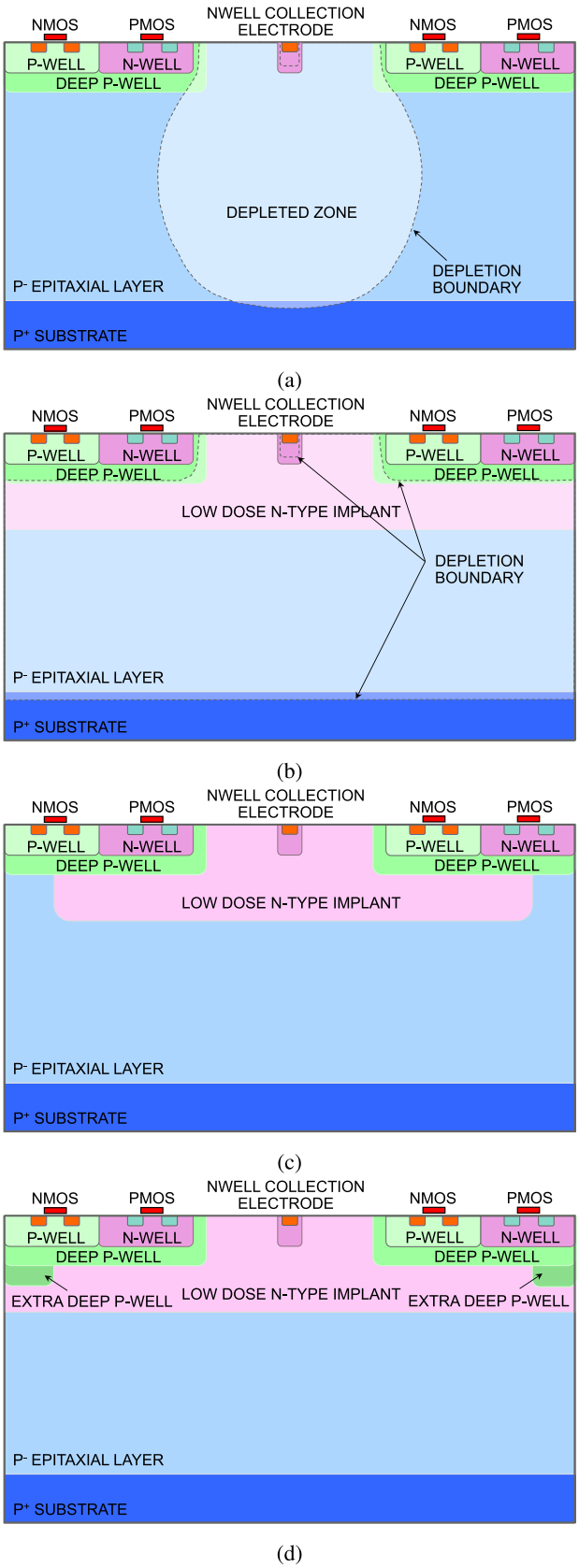

Fig. 1. Cross section of the TowerJazz 180-nm CMOS imaging sensor: (a) standard process; (b) modified process with low-dose n$^-$ implant; (c) with gap in the low-dose n$^-$ implant; (d) with extra deep p-well. Sensors in (c) and (d) also fully deplete with a sufficiently large reverse bias.

the collection time. Faster collection times also reduce the probability for charge carriers to get captured by the radiation-induced defects, improving the sensor tolerance to NIEL. To achieve full depletion of the sensor volume, the process has been modified [12] by adding a uniform ion-implanted low-dose n^- layer under the deep p-well covering the entire matrix/pixel area, as shown in Fig. 1(b). Due to its low doping, the n^- layer is fully depleted in typical biasing conditions, and the electrode is effectively still defined by the small n-well. This modification creates a planar junction deep in the epitaxial layer and the depletion extends immediately over the full pixel area. However, despite the full depletion of the epitaxial layer, the lateral electric field in the pixel corners and along the pixel edges is still quite low, resulting in a relatively long collection time and hence a high probability for the charge to get captured by the radiation-induced traps, which leads to efficiency loss in these regions [9]. The strength of the lateral electric field in the corner regions can be increased by introducing a lateral gradient in the doping profile [13]. This can be achieved by patterning the n^- layer, that is, removing it at the pixel edges, as shown in Fig. 1(c), or with the introduction of an extra deep p-well implant, as shown in Fig. 1(d) [10].

The advantage of the small collection electrode is its low sensor capacitance, which is key to achieve a lower analog power consumption for a given charge, bandwidth, and a signal-to-noise ratio (SNR) [14]. The SNR of the analog front-end can be calculated by comparing the input signal created by the ionization of a high-energy particle to the input-referred rms noise. The latter is typically dominated by the input transistor thermal noise, inversely proportional to the square root of its transconductance g_m, and can be expressed as an equivalent series voltage. The input signal can be calculated as the ratio of the generated charge divided by the sensor capacitance, that is, the Q/C ratio. Therefore, assuming the front-end power consumption is dominated by the input transistor current, one can write

$$\frac{S}{N} \propto \frac{Q}{C}\sqrt{g_m} \propto \frac{Q}{C}\sqrt[m]{P} \tag{1}$$

with $m = 2$ for the input transistor in weak inversion, where g_m is proportional to the biasing current, or $m = 4$ in strong inversion, where g_m is proportional to the square root of the biasing current. Rearranging, for a fixed SNR and bandwidth

$$P \propto \left(\frac{Q}{C}\right)^{-m} \tag{2}$$

with $2 \leq m \leq 4$, depending on the operating point of the input transistor. The expression states that the required power consumption to reach a given SNR is heavily dependent on the Q/C ratio. It is therefore important to mention that the process modification does not introduce any penalty on the sensor capacitance, thanks to the full depletion of the additional n^- layer [12]. This article shows results for pixels that implement the sensor from Fig. 1(d). The epitaxial layer is 30-μm thick and has a resistivity larger than 1 k$\Omega \cdot$ cm. The collection electrode is an octagonal-shaped n-well with a diameter of 2 μm, distanced 4 μm in all the directions from the surrounding p-well containing the circuitry. This geometry is the result of a trade-off between a small sensor capacitance (<5 fF) and a large lateral electric field [15], with collection times in the nanosecond range [13]. The sensor in this modified process was tested up to 10^{15} 1-MeV n_{eq}/cm^2 [3], still showing good tolerance to NIEL. The transistors also have a good tolerance to TID due to the thin oxide thickness provided by the technology [16].

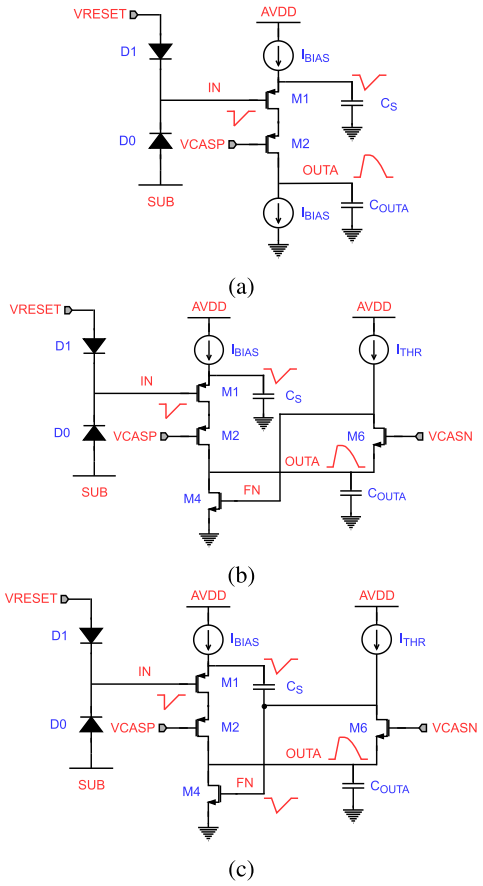

Fig. 2. Front-end principle: (a) simplified schematic; (b) with feedback to adjust the operating point; (c) with additional gain mechanism.

III. Front-End

The front-end is a continuously active circuit that performs the reset of the collection electrode, the amplification of the generated charge, the shaping of the analog signal, and the digitization of this signal through a discrimination stage. The basic principle of the amplification stage is illustrated in Fig. 2(a). The continuous reset mechanism uses the diode $D1$ to hold the collection electrode voltage. The latter is set to $\sim$1 V for a proper front-end operation by tuning the V_{RESET} bias depending on the sensor leakage current. The p-type substrate and p-well containing the circuitry can be biased down to -6 V, allowing to achieve a sensor reverse bias of ~ -7 V. When no charge is collected by the collection electrode, $D1$ is biased by the leakage current of the sensor diode $D0$. Upon a particle crossing, the electrode collects the generated electrons and a negative voltage step with an amplitude of $\Delta V = Q/C$ is generated on it. This causes

the reset diode to conduct more current and to slowly charge the input node back up to its original value, which can take several hundreds of μs. The reset diode is implemented with a small p^- implant in the n-well of the collection electrode, adding only a small capacitance to the sensor. A closed-loop charge-sensitive amplifier that integrates the charge on a feedback capacitor [17] is typically used in these applications. In such architecture, for a very small sensor capacitance, the feedback capacitance might be non-negligible with respect to the sensor capacitance and create a noise penalty for the same power consumption. Lowering the feedback capacitance to a small fraction of the low sensor capacitance will make it also typically lower than the output capacitance, degrading the speed of the circuit. The proposed architecture implements an open-loop amplifier that integrates the charge on the sensor itself and avoids the aforementioned limitations, resulting in a simpler and more power-efficient solution. The input node (gate of the transistor $M1$) is connected directly to the collection electrode. The input transistor $M1$ acts as a source follower and, when the input voltage drops because of the collected charge, forces its source to follow transferring charge from the capacitance C_S to the output node capacitance C_{OUTA}. Ideally, for the voltage on OUTA, one can write

$$\Delta V_{\text{OUTA}} = \frac{Q_S}{C_{\text{OUTA}}} = \frac{C_S \cdot \Delta V_{\text{IN}}}{C_{\text{OUTA}}} = \frac{C_S}{C_{\text{OUTA}}} \frac{Q_{\text{IN}}}{C_{\text{IN}}}. \quad (3)$$

Therefore, a large gain is obtained for $C_S \gg C_{\text{OUTA}}$. The overall effective sensor capacitance is the sum of the sensor junction capacitance, the reset diode parasitic capacitance, the input line, and the input transistor gate capacitance. After settling, the following action of the input transistor reduces the contribution of its gate–source capacitance to the total capacitive load on the electrode. Furthermore, the cascode transistor $M2$ mitigates the Miller effect on the gate–drain capacitance of the input transistor. A more practical implementation of the circuit is shown in Fig. 2(b). Since the two I_{BIAS} current sources are difficult to match, low-frequency feedback that sets the operating point of the transistor $M4$ is introduced: its gate voltage is now adjusted for it to sink $I_{\text{BIAS}} + I_{\text{THR}}$, where I_{THR} is a small fraction of the main biasing current I_{BIAS}. This branch defines the dc voltage of the amplifier output node and its return to baseline. Upon a particle hit, when the voltage on OUTA rises, the gate–source voltage of the transistor $M6$ reduces, forcing I_{THR} to charge up the gate of the transistor $M4$, discharging OUTA and bringing it back to its baseline value. An additional gain mechanism is introduced by connecting the capacitance C_S to the gate of the transistor $M4$, as done in Fig. 2(c). A part of the signal on the input transistor source is now transferred to the gate of the transistor $M4$, which behaves as a common-source device. In this scheme, the capacitance C_S plays an important role not only in the gain of the amplifier, but also in determining its return to baseline, since it is connected to the feedback node (FN). A larger I_{THR} increases the speed of the feedback loop, resulting in a faster return to baseline, but could also provide excessive filtering at low frequencies on the gate of the transistor $M4$, reducing the amplifier gain. Indeed, due to this low-frequency internal feedback and the gain mechanism

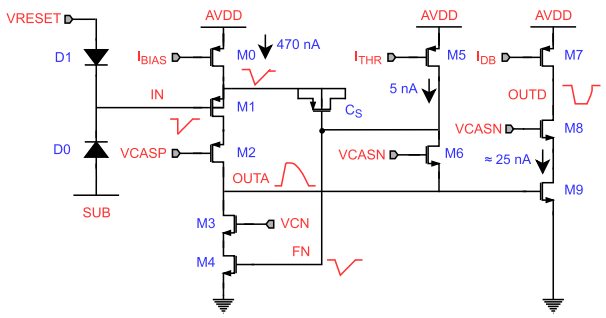

Fig. 3. Complete front-end schematic with discriminator.

introduced by the capacitance C_S, the front-end response behaves at low frequencies as a high-pass filter. The gain, however, drops at high frequencies due to the poles on the output node (OUTA) and FN which exhibit a high impedance. Overall, the circuit is characterized by a bandpass response and no additional shaping is required after the amplification stage. The bandwidth of the amplifier can be optimized for the signal bandwidth to improve the SNR or, in other words, reduce the equivalent noise charge (ENC).

The complete front-end circuit which includes the amplifier and discriminator is shown in Fig. 3. The capacitance C_S is implemented with a pMOS device whose source, bulk, and drain are connected together to exploit the capacitance of the MOS structure in inversion. The capacitor C_{OUTA} includes only the parasitic contributions of transistors connecting to it since it needs to be as low as possible. The input transistor $M1$ is placed together with the capacitor C_S in a separate n-well connected to its source to eliminate the body effect and achieve a gain close to unity for the input source follower. An improvement to the circuit from Fig. 2(c) is provided by cascoding the transistor $M4$. For good timing performance, a large transconductance is required for this transistor. However, a too large aspect ratio would increase the output parasitic capacitance, detrimental both for gain and speed itself. The cascode decouples the transistor $M4$ from the output node, giving more freedom in its sizing and is optimized for a reduced output capacitance. Additionally, it increases the output impedance of the amplifier, which is thus dominated by the transconductance of the transistor $M6$ that works in weak inversion, leading to a higher gain. The discriminator consists of a common-source amplification stage, the transistors $M7$–$M9$, which can be better seen as a current comparator. In a steady state, the output baseline of the amplifier sets the standby current of the transistor $M9$, while the transistor $M7$ is biased to provide a current I_{DB} higher than the dc current forced by the transistor $M9$, charging the node OUTD to the supply voltage. As the signal on OUTA rises upon a particle hit, the current drawn by the transistor $M9$ increases, eventually exceeding I_{DB} and discharging the output node to the ground. The threshold of the discriminator is therefore controlled by the I_{DB} current setting and the amplifier output baseline (through V_{CASN}). The cascode transistor $M8$ is again used to reduce the large capacitance penalty on OUTA due to the Miller effect on the transistor $M9$ and the coupling between this node and the rail-to-rail OUTD signal.

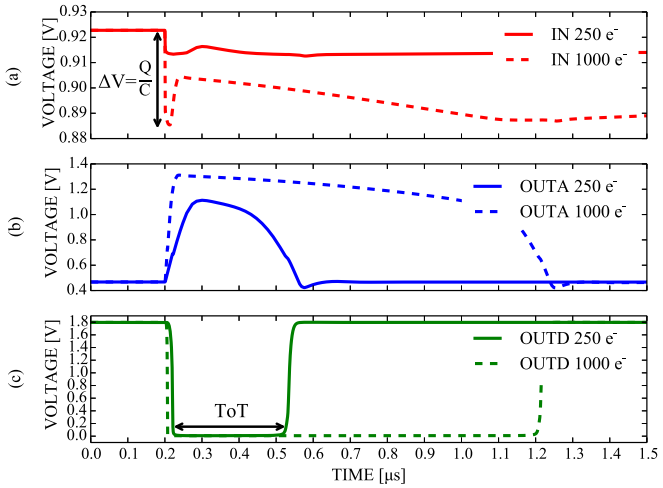

Fig. 4. Front-end simulated transient response with a 100 e$^-$ threshold: signals at the (a) sensing node, (b) output of the amplifier, and (c) output of the discriminator.

The circuit is designed to have peaking times in the order of tens of ns with low power consumption. For the input follower action, the input transistor more quickly discharges its load capacitance with a larger transconductance g_m. For the gain contribution provided by the coupling with the transistor $M4$, its transconductance g_m and the output capacitance C_{OUTA} define the gain–bandwidth product. Therefore, the peaking time decreases with a higher transconductance g_m of the amplifying devices and a lower output capacitance C_{OUTA}. The transistors' dimensions and the layout are therefore optimized to reduce C_{OUTA} to less than $\sim$5 fF. To reach the target timing response, the main biasing current I_{BIAS} needs to be $\sim$470 nA for a sufficiently large g_m of the amplifying devices. The I_{THR} current, typically a few nA, and the discriminator off current, typically a few tens of nA, need to be added to the I_{BIAS} current to obtain the total current consumption which is $\sim$500 nA. With a supply voltage of 1.8 V, the total power consumption is less than 1 μW per pixel, which is used efficiently, thanks to the current reuse between the input follower $M1$ and the common-source device $M4$. A parasitic-extracted simulation of the transient waveforms at the input IN, analog output OUTA, and discriminator output OUTD of the front-end with the charge threshold set to 100 e$^-$ are shown in Fig. 4. The solid lines show the response for a collected charge of 250 e$^-$, whereas the dashed lines for a charge of 1000 e$^-$. The simulation was performed using a current pulse at the input, that is, by injecting the input charge uniformly in a collection time of 1 ns. The sensor is modeled as a capacitance of 2.5 fF, which is a value previously measured on prototype chips [18], in parallel with a leakage current source of 10 pA. The red curves represent the input signals and show that the voltage step on the electrode is proportional to the collected charge. The blue curves represent the amplified signals on OUTA. The front-end gain is nonlinear since the transistor M6 dynamically turns off as the output voltage rises, offering a larger impedance on the output node. At threshold, the gain is $\approx$1.9 mV/e$^-$, whereas for a charge of 250 e$^-$, as seen in Fig. 4, it is $\approx$2.5 mV/e$^-$. For larger charges, the analog output signal on OUTA is sufficiently large to push the cascode transistor $M2$ out of saturation, and the front-end gain drops. This makes the cascode ineffective, so the equivalent input capacitance increases due to the Miller effect. The input signal is, therefore, lower during this transition time, as shown from the dashed red curve of Fig. 4, and saturation of the analog output signal is reached. However, its time over threshold (ToT), that is, the duration of the discriminator output pulse shown in green in Fig. 4, has a linear dependence on the input charge. Indeed, the ToT depends on the time required for the feedback circuit to charge up the capacitor C_S through the current I_{THR}.

TABLE I
OUTPUT NOISE CONTRIBUTIONS

Device	Noise source	Percentage
M1	Thermal noise	53%
M4	Thermal noise	36%
D1	Shot noise	2.5%
M6	Thermal noise	2%
M5	Thermal noise	1.75%

For a reliable operation of the sensor, the front-end ENC has to be considerably lower than the charge threshold. Even though the circuit is highly nonlinear, a small-signal ac noise analysis helps to gain insights into the main noise sources. The results of this type of simulation are shown in Table I that reports the highest contributions with the corresponding percentage of the output noise, integrated from 1 Hz to 10 GHz, that is, in a frequency range much broader than the amplifier passband, which extends from $\sim$15 kHz to $\sim$10 MHz.

The thermal noise of the amplifying devices $M1$ and $M4$ is the dominating noise contributor, as assumed in the previous section for the derivation of the power consumption as a function of the sensor capacitance. The next main contributor is the shot noise from the reset diode. The simulation was performed with a leakage current of 10 pA which is an overestimate for an unirradiated sensor. A small percentage of noise comes from the I_{THR} current source and the device $M6$ that define the baseline of the amplifier output. Other contributions to the output noise are much less significant. An important noise source not accounted for in the simulation is the random telegraph signal (RTS) noise. Indeed, the simulation models do not include RTS noise which is therefore difficult to estimate during design. The most critical devices for RTS noise are again the transistors $M1$ and $M4$ whose sizing required an iterative process [10]. Increasing the gate area of the transistor $M1$ to combat RTS noise results in larger effective sensor capacitance. A gate area of 0.18 μm^2 has been chosen since it is a good compromise between capacitance penalty and noise. A gate area of $\sim$2.4 μm^2 has instead been chosen for the transistor $M4$ since it exhibits a larger noise transfer function to the output node and the RTS noise is typically larger in nMOS transistors. One of the main parameters for good noise performance is the size of the capacitance C_S: a larger capacitance provides dynamically more charge to

the output node for the same input signal and improves the coupling between the input transistor source and the gate of the transistor $M4$. From a frequency standpoint, it widens the amplifier passband toward lower frequencies where the input signal has a large frequency content. The output signal, therefore, increases more than the noise level and a larger SNR is obtained. The pMOS transistor implementing this capacitor has a gate area of $\sim$14.24 μm^2, providing a capacitance of $\sim$114 fF, and it is one of the largest components of the circuit.

Apart from the noise, another limit to the minimum operating threshold is the pixel-to-pixel variation of the transistor parameters, which causes the threshold to vary over the matrix. It is well known that the transistors' mismatch scales down with the square root of their area [19]. The pixel size is, however, limited and often dictated by the target sensor's spatial resolution. To optimize the space, it is, therefore, necessary to identify the devices with the largest impact on the threshold dispersion and increase their area. In the amplification stage, the most critical devices are the transistors $M5$ and $M6$. As previously said, the current I_{THR} of the transistor $M5$ defines the speed of the feedback loop and significantly influences the amplifier gain. Regarding the transistor $M6$, its gate–source voltage directly defines the amplifier output baseline, setting the discriminator dc current and hence its switching threshold. The transistor $M5$ is biased with a current of only a few nA and therefore operates in weak inversion, which makes the impact of its mismatch even more prominent. For these reasons, it is designed with a low aspect ratio and a large area (20 μm^2). The size of the transistor $M6$, however, cannot be increased to the same extent because of the capacitance penalty on the output node. In the discriminator stage, the input transistor $M9$ is the main critical device: a variation of its threshold voltage directly shifts the switching point of the discriminator, appearing effectively as an offset. As for the transistor $M6$, it has to be kept small to prevent increasing the amplifier output capacitance and it represents the largest contribution (nearly 50%) to the overall threshold dispersion. Fig. 5 shows the front-end probability to generate a hit as a function of the input charge, obtained with transient-noise simulations [see Fig. 5(a)] and Monte Carlo simulations for transistors' mismatch [see Fig. 5(b)]. The mean value of the Gaussian error function fit gives the nominal threshold, which is $\sim$100 e$^-$, whereas its standard deviation gives the ENC and the pixel-to-pixel threshold variation for the two simulations, respectively, which are $\sim$6.4 and $\sim$2.5 e$^-$.

The radiation effects that influence the front-end operation and performance include the increase of the sensor leakage current due to NIEL and the TID effects that affect the transistor characteristics. For a good tolerance to NIEL, the front end has been designed to cope with a wide range of sensor leakage currents. As for the tolerance to TID, minimum dimensions have been avoided for the critical devices to mitigate radiation-induced narrow channel (RINCE) and radiation-induced short channel (RISCE) effects [20]. Leakage currents in the order of a hundred pA have been measured for nMOS transistors in this technology after 20 Mrad of

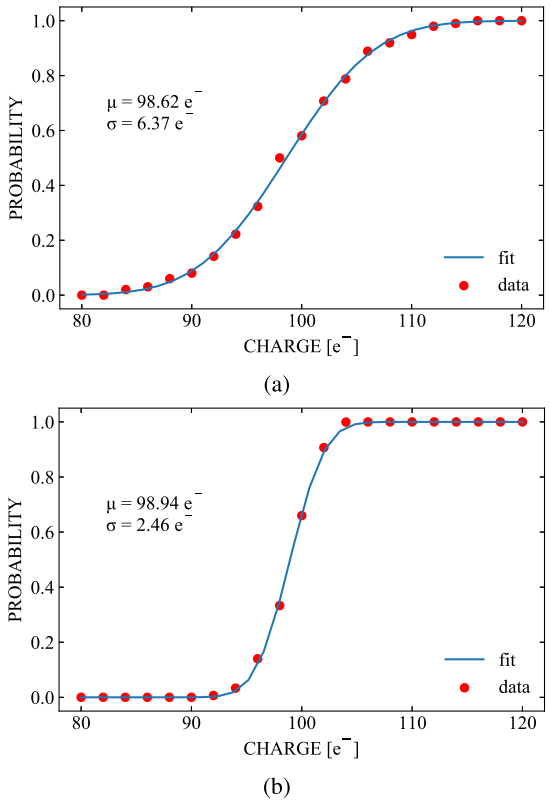

Fig. 5. Hit probability as a function of the input charge with simulated (a) transient noise and (b) transistors' mismatch.

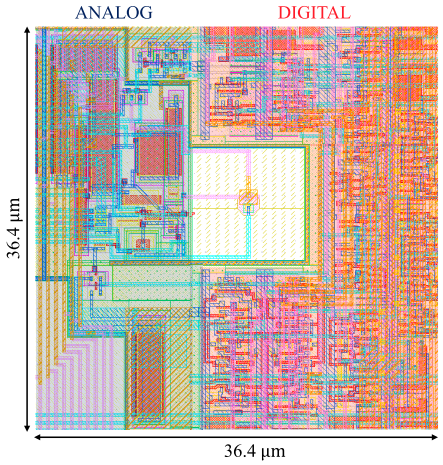

Fig. 6. Layout of the pixel.

TID [21]. Since the I_{THR} current can be below 1 nA, the transistor $M6$ has been designed as an enclosed layout transistor (ELT [22]) and is surrounded by a p$^+$ guard ring to prevent any leakage to neighboring devices. These precautions double its area but they are necessary to ensure radiation hardness. The layout of the pixel is shown in Fig. 6. The 2-μm octagonal collection electrode, distanced 4 μm from the surrounding p-well of the circuitry, is placed in the center of the pixel. The front-end circuit occupies an area of $\sim$160 μm^2 and is placed to the left of the collection electrode with other analog circuitry. The latter includes decoupling capacitors and a testing circuit that can capacitively inject a tunable amount

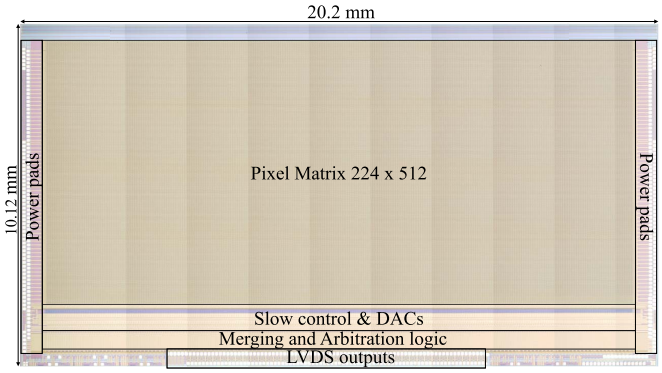

Fig. 7. Micrograph of the MALTA2 chip.

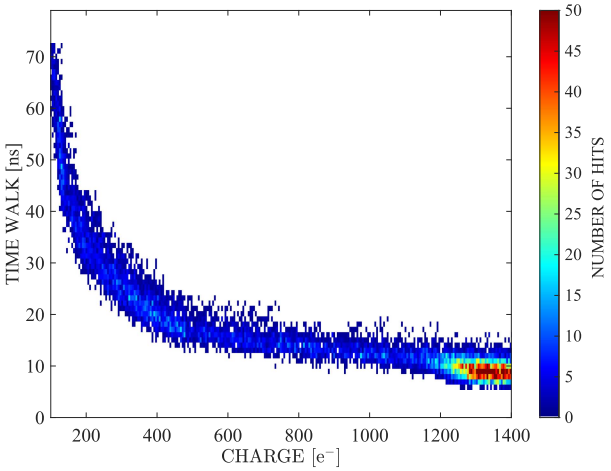

Fig. 8. Time walk curve obtained with a ^{90}Sr source.

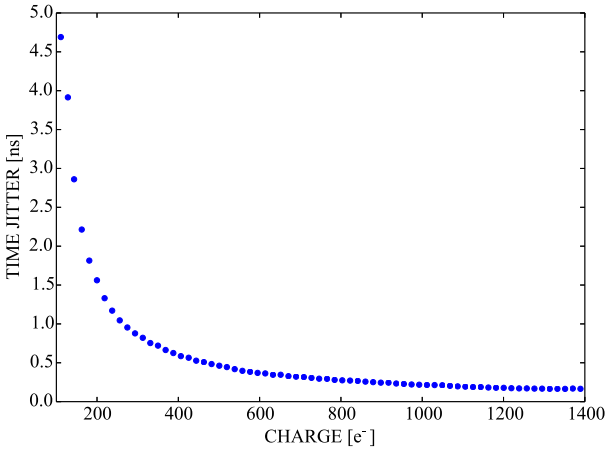

Fig. 9. Dependence of front-end time jitter on charge.

of charge to the collection electrode. The rest of the pixel is occupied by the digital readout circuitry for a total pixel area of 36.4 μm $\times$ 36.4 μm, leading to an analog power density of 75 mW/cm^2 over the matrix.

The MALTA2 chip has a size of 10.12 mm $\times$ 20.2 mm and integrates a matrix of 224 $\times$ 512 pixels. It features an asynchronous readout that avoids propagation of the clock in the matrix to reduce the digital power consumption. Upon a particle hit, the in-pixel digital circuitry sends a pattern of short pulses corresponding to the pixel address to the periphery on a digital data bus. The pixels are organized in double columns, where each double column has a dedicated bus for the transmission of the data down to the periphery. Here, a binary tree-like structure that merges the hits of the whole matrix onto a single bus is implemented. In the case of simultaneous events, this logic delays one of the two hits in time while keeping track of the delay for later reconstruction. The final word is 40 bits wide and is transmitted off-chip via LVDS drivers, which are designed to operate at a maximum speed of 5 Gbps [23], sufficiently high for the pixel detector to cope with the ATLAS ITk outer layer hit rate of 100 MHz/cm^2. A micrograph of the MALTA2 chip is shown in Fig. 7.

IV. Front-End Characterization

To test the performance of the front-end, a special set of pixels that allows the monitoring of its analog output has been included on the left and right sides of the matrix. In these pixels, the front-end analog output is buffered to an output pad with a two-stage source follower with a gain close to 1. The first stage is optimized to match closely the discriminator input capacitance to have the same amplifier output load as in the other pixels of the matrix. An oscilloscope is used to monitor the output pad through a low-capacitance active probe and the full buffering system is designed not to degrade the signal timing. The front-end speed can be evaluated with the plot of Fig. 8 which shows the time walk curve, that is, the time for the amplifier output to reach the discriminator threshold as a function of the charge. The conversion between charge and amplitude is derived through the charge injection circuitry. The injection capacitance was calibrated with ToT measurements of signals from test pulses and an ^{55}Fe source. For this measurement, the front-end operates with the nominal bias settings ($\sim$1-μW power consumption, as confirmed by measuring the total analog current consumption of the matrix), and the oscilloscope is set to trigger with a signal of $\sim$100 e$^-$. The waveforms are collected while exposing the chip to a ^{90}Sr radioactive source which undergoes β^- decay emitting electrons that generate an ionization signal close to a minimum ionizing particle (MIP). The most probable value of charge deposition for an MIP in the 30-μm-thick epitaxial layer is $\sim$1800 e$^-$ [11]. The signal is collected by a cluster of up to four pixels and the seed pixel, the one with the largest signal, has a charge $\geq$1/4 of an MIP charge. Events with high charges ($\gtrsim$1200 e$^-$) have a threshold crossing time close to the minimum value of $\sim$10 ns. With respect to the ATLAS application, an event is considered in time when it falls within 25 ns from this value. As can be noticed in Fig. 8, the in-time threshold corresponds to an input charge of $\sim$200 e$^-$. Less than 10% of the hits are below the in-time threshold. Statistically, these are mostly caused by non-seed pixels, with a neighboring seed pixel that is likely to collect a charge above the in-time threshold.

The front-end timing can also be studied through the matrix digital readout. An increasing amount of charge can be injected

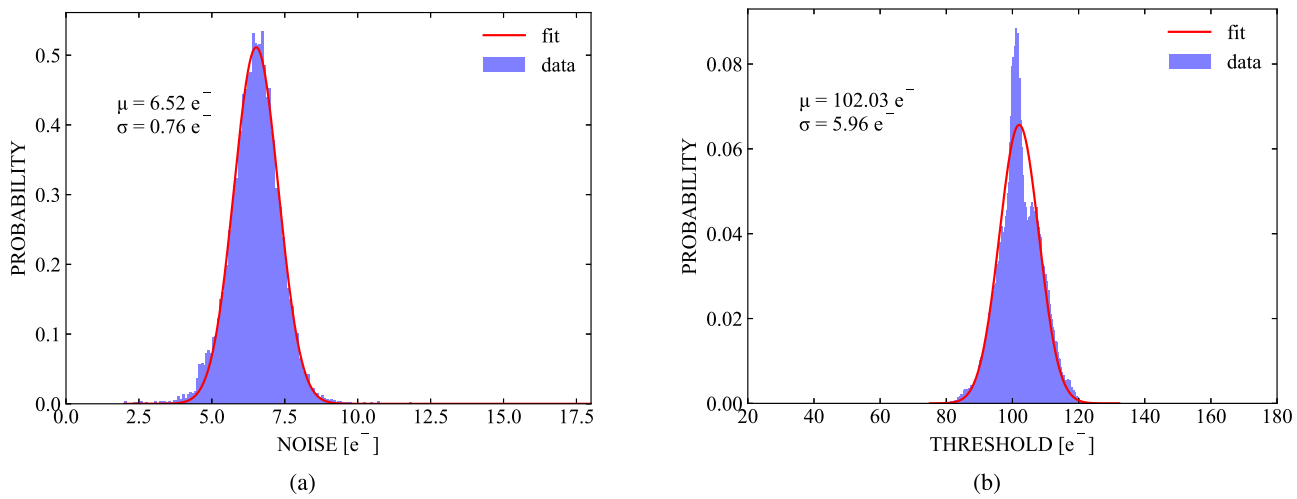

Fig. 10. Distribution of (a) ENC and (b) threshold.

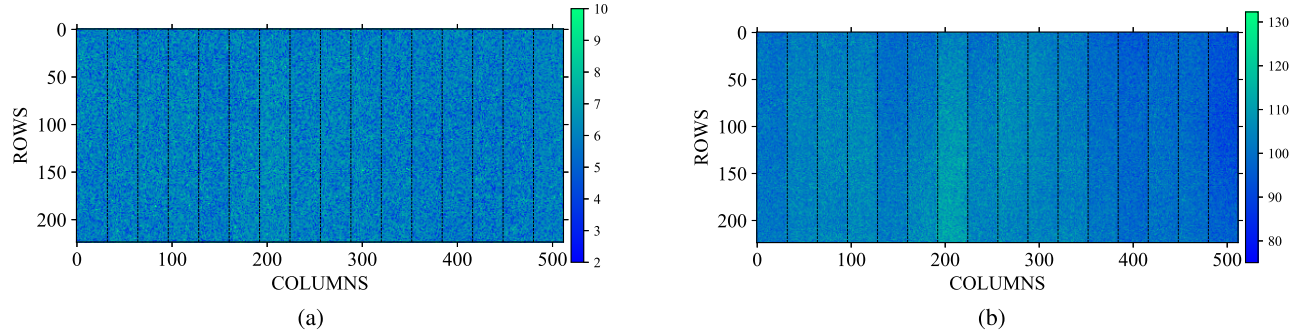

Fig. 11. Two-dimensional map for (a) ENC and (b) threshold.

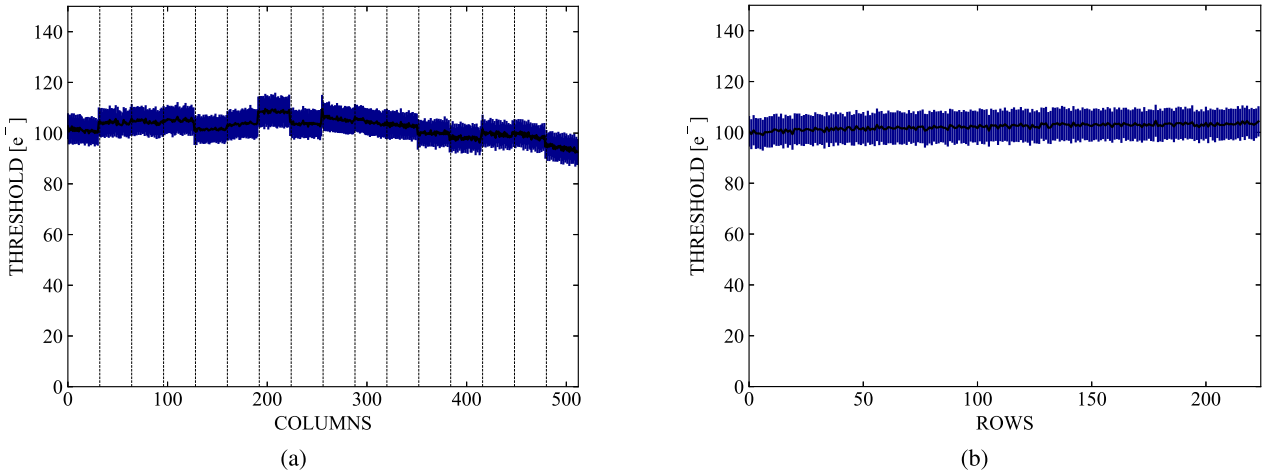

Fig. 12. Distribution of the threshold versus the (a) columns and (b) rows.

into a specific pixel with the aforementioned charge injection circuitry. The time of arrival of the generated hits can then be compared to a time reference. This procedure has been performed using as time reference the charge injection trigger pulse sent to the chip. In order to do so, this signal is also sent to an external 3-ps binning TDC [24] together with a fast-OR signal from the chip. The mean difference between these two signals' time of arrival provides a time walk curve compatible with the one in Fig. 8. This methodology, however, allows for a better study of the front-end jitter by evaluating instead the rms difference between the two signals' time of arrival, which is plotted as a function of the charge in Fig. 9. For each injected charge, 10000 events are acquired. The time jitter of the reference pulse has been estimated to be below 100 ps, and therefore, the values in Fig. 9 are dominated by the front-end jitter which reduces from 4.7 ns at the threshold, down to 0.16 ns for very high input charges.

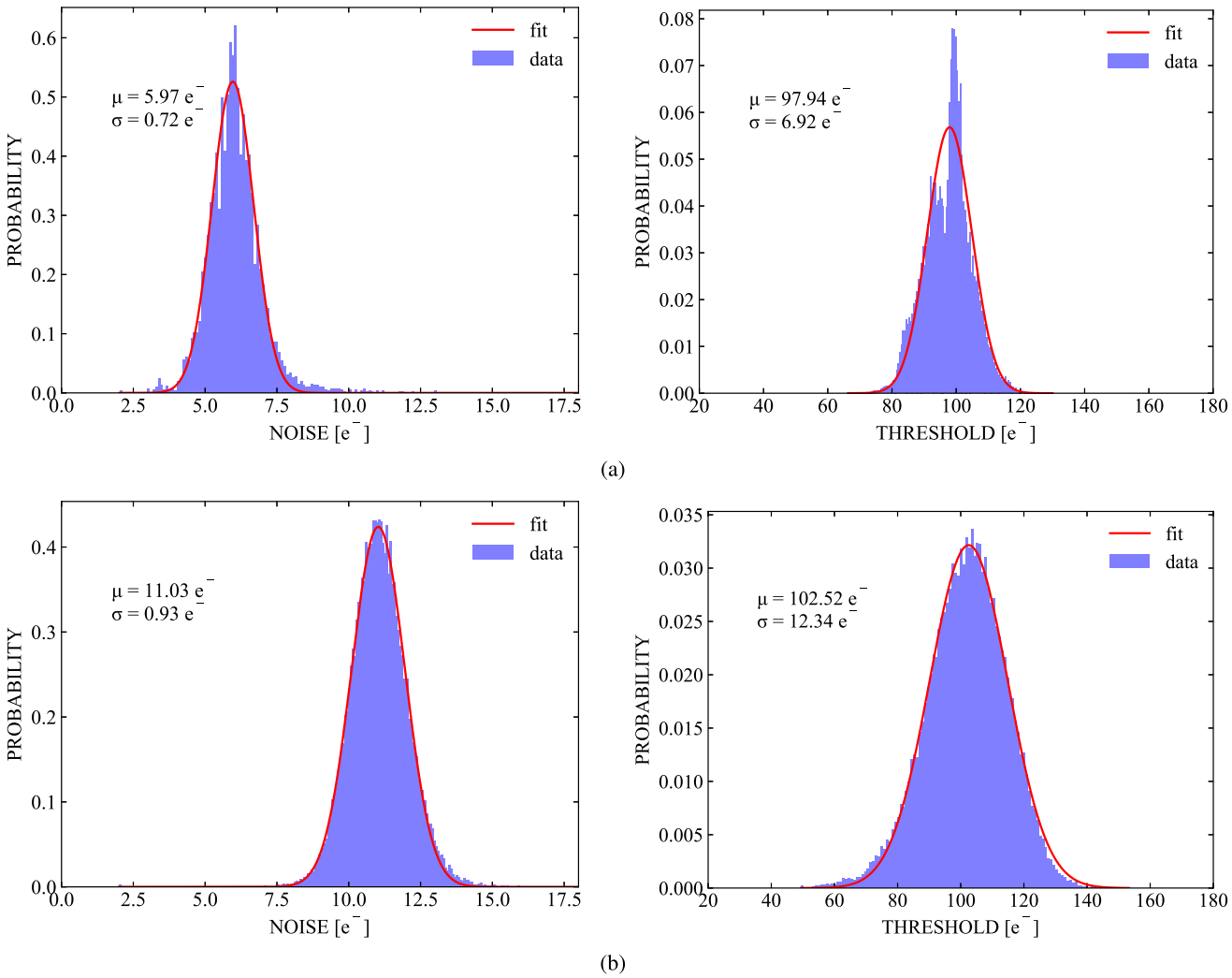

Fig. 13. Distribution of ENC and threshold with a threshold of ~100 e^- for (a) unirradiated sample and (b) sample irradiated at $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 3 Mrad.

The charge injection circuitry also allows us to extract information such as threshold and noise for all the pixels. Fig. 10 shows the threshold and noise distribution for an entire matrix with nominal front-end settings. The average threshold is ~100 e^- with a variation of ~6 e^-, more than a factor of 2 higher than the simulated value shown in Fig. 5. The noise distribution has an average of 6.5 e^- with a low spread, matching fairly well the simulations. Two-dimensional maps of the pixels' threshold and noise are shown in Fig. 11. No systematic effects are observed for the noise. As for the threshold, it is possible to notice a variation in its average over different vertical sections of the matrix. This effect strongly correlates with the scheme of the front-end biasing which is adjusted through DACs in the bottom periphery. The power pads are distributed only along the left and right sides of the matrix. For this reason, a horizontal power voltage drop is inevitably present and is estimated to reach ~12 mV in the middle of the matrix. To compensate for this effect and avoid a systematic threshold gradient, the biasing DACs have a dedicated mirroring stage for every 32 columns of the matrix which shares their local power supply. To better study the threshold behavior, the distribution of the threshold along the columns with one rms error bar is shown in Fig. 12(a) and here the threshold average variation at every biasing group is more clearly visible. A straightforward solution to increase the biasing transistors' area and mitigate this effect is to connect more mirroring stages together, trading off with the power voltage drop compensation accuracy. This is envisaged for a future prototype. Fig. 12(b) illustrates the distribution of the threshold along the rows showing a slight vertical gradient which is caused by the mirroring stages at the matrix bottom which loads the matrix power grid and introduces a vertical power voltage drop. Considering only pixels within the same biasing group and correcting the systematic vertical gradient, the threshold variation is ~5.1 e^-. The variation of the nMOS transistors' output conductance with a high reverse bias to the bulk (beyond the normal supply voltage) is not fully covered by the simulation models and this is thought to be the cause of the discrepancy between the simulated and measured threshold variation. Even with a larger pixel-to-pixel mismatch, the chip can be operated reliably with thresholds of ~100 e^-.

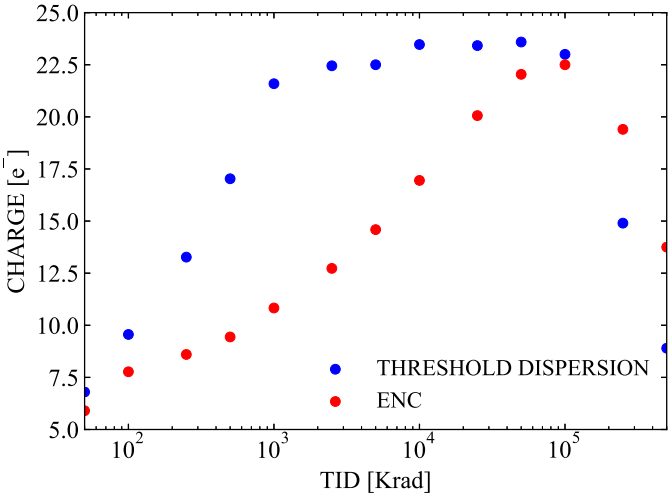

Fig. 14. Dependence of ENC and threshold dispersion on TID with a threshold of ∼100 e⁻. The points at 50 krad represent measurements before irradiation. The points at 250 and 500 Mrad correspond to measurements after 24-h annealing at room temperature and additional 24 h of annealing at 80 °C, respectively.

A number of chips have been irradiated with neutrons at the TRIGA reactor in Ljubljana [25] up to $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 of NIEL fluence. The chips also received a background TID of 1 Mrad for every 10^{15} 1-MeV n_{eq}/cm^2. During irradiation, the chips were not powered. After irradiation, the chips are stored at a low temperature (below −20 °C) to avoid annealing of the radiation damage. For the same reason, all the measurements of irradiated samples are performed at −20 °C, which also helps to contain the increase of the sensor leakage current. The chip still shows full functionality after these levels of irradiation.

Charge injection tests have been performed on these samples with a step of 10^{15} 1-MeV n_{eq}/cm^2 and 1 Mrad. For a fair comparison, the measurement of the unirradiated sample has been repeated at −20 °C and the I_{THR} current setting of the front-end has been adjusted to obtain similar thresholds in all the cases. An increasing level of ENC and threshold dispersion as a function of the irradiation level has been noticed. The distribution of ENC and threshold dispersion for an unirradiated sample and a sample irradiated to $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 3 Mrad are shown in Fig. 13 for reference. The noise average for the unirradiated sample is ∼6 e⁻, slightly lower than in the previous case due to the lower temperature and increases to ∼11 e⁻ for the sample irradiated at $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 3 Mrad. The lack of substantial non-Gaussian tails suggests a negligible contribution of RTS noise. The pixel-to-pixel variation increases from ∼7 e⁻ in the unirradiated case to ∼12.5 e⁻ for the sample irradiated at $3 \cdot 10^{15}$ 1-MeV n_{eq}/cm^2 and 3 Mrad.

To evaluate the front-end performance for higher TID levels, the chip has been irradiated with X-rays at a dose rate of 25 krad/min up to 100 Mrad. The chip still shows full functionality at this TID level. The irradiation was stopped at different doses to perform basic functionality tests and evaluate the front-end performance. In order to reproduce the typical operating conditions, the chip was powered and biased during irradiation. Additionally, to minimize annealing effects, the chip was kept at a low temperature (−10 °C) during the whole process. The measured ENC and pixel-to-pixel threshold variation as a function of TID are shown in Fig. 14. The threshold was adjusted to ∼ 100 e⁻ at each step of the measurement. The ENC grows monotonically from ∼ 5.9 e⁻ before irradiation (first data point) to ∼22.5 e⁻ at 100 Mrad. High levels of RTS noise are present in the ENC distributions for TID levels higher than 1 Mrad. However, already after 24 h of annealing at room temperature, the RTS noise disappears, and the mean ENC drops from ∼22.5 to ∼19 e⁻ (data point at 250 Mrad). After another 24 h of annealing at 80 °C, the mean ENC reduces to ∼14 e⁻ (last data point). As for the pixel-to-pixel threshold variation, it increases from ∼6.8 e⁻ before irradiation to ∼23 e⁻ at 100 Mrad. The threshold dispersion more rapidly increases with TID compared to the noise, but it settles around ∼23 e⁻ already at 1 Mrad. After 24 h of annealing at room temperature, it drops to ∼14.5 e⁻ and it further drops to ∼9 e⁻ after another 24 h of annealing at 80 °C.

TABLE II
FRONT-END SPECIFICATIONS WITH A 100 e⁻ THRESHOLD

Parameter		Value
Area		160 µm²
Power consumption		1 µW
In-time threshold (for a 25 ns time window)		200 e⁻
ENC	unirradiated	6.5 e⁻
	$3 \cdot 10^{15}$ 1 MeV n_{eq}/cm^2, 3 Mrad	11 e⁻
	100 Mrad	22.5 e⁻
Threshold dispersion	unirradiated	6 e⁻
	$3 \cdot 10^{15}$ 1 MeV n_{eq}/cm^2, 3 Mrad	12.5 e⁻
	100 Mrad	23 e⁻
Timing jitter	at threshold	4.7 ns
	for high charges (≳ 1200 e⁻)	0.16 ns

V. CONCLUSION

This article describes the design and characterization of a low-power, radiation-hard front-end circuit for monolithic pixel sensors. The circuit is implemented in the TowerJazz 180-nm CMOS imaging technology and integrated into the MALTA2 chip. The sensor features a small octagonal collection electrode with a diameter of 2 μm to achieve a low sensor capacitance (<5 fF), which is key for low-power operation for a given ENC and bandwidth. Process modifications have been introduced to fully deplete the sensor and enhance the lateral electric field in the pixel corners for good tolerance to NIEL. The front-end is a continuously active open-loop amplifier followed by a high-gain common-source discriminator stage. It is designed for a gain of ∼2 mV/e⁻ and a peaking time in the order of tens of ns, requiring <1 μW per pixel and a layout area of ∼160 μm². It is optimized for low noise and low pixel-to-pixel variation to achieve low charge thresholds. MALTA2 samples were extensively characterized to evaluate

the front-end performance before and after irradiation. The main front-end metrics with a threshold of 100 e^- are summarized in Table II.

ACKNOWLEDGMENT

The authors would like to thank the Jožef Stefan Institute, Ljubljana, Slovenia, for the support during the neutron irradiation.

REFERENCES

[1] J. Ballin *et al.*, "Monolithic active pixel sensors (MAPS) in a quadruple well technology for nearly 100% fill factor and full CMOS pixels," *Sensors*, vol. 8, no. 9, pp. 5336–5351, Sep. 2008, doi: 10.3390/s8095336.

[2] I. Perić, "A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 582, no. 3, pp. 876–885, Dec. 2007, doi: 10.1016/j.nima.2007.07.115.

[3] H. Pernegger *et al.*, "First tests of a novel radiation hard CMOS sensor process for depleted monolithic active pixel sensors," *J. Instrum.*, vol. 12, no. 6, Jun. 2017, Art. no. P06008, doi: 10.1088/1748-0221/12/06/p06008.

[4] I. Peric *et al.*, "High-voltage CMOS active pixel sensor," *IEEE J. Solid-State Circuits*, vol. 56, no. 8, pp. 2488–2502, Aug. 2021, doi: 10.1109/JSSC.2021.3061760.

[5] I. Caicedo *et al.*, "The Monopix chips: Depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS inner tracker upgrade," *J. Instrum.*, vol. 14, no. 6, Jun. 2019, Art. no. C06006, doi: 10.1088/1748-0221/14/06/c06006.

[6] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, document CERN-LHCC-2017-021, ATLAS-TDR-030, Sep. 2017, [Online]. Available: https://cds.cern.ch/record/2285585

[7] I. Berdalovic, "Monolithic pixel development in TowerJazz 180 nm CMOS for the outer pixel layers in the ATLAS experiment," *J. Instrum.*, vol. 13, no. 1, 2018, Art. no. C01023, doi: 10.1088/1748-0221/13/01/c01023.

[8] I. Berdalovic *et al.*, "MALTA: A CMOS pixel sensor with asynchronous readout for the ATLAS high-luminosity upgrade," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Nov. 2018, pp. 1–4, doi: 10.1109/NSSMIC.2018.8824349.

[9] E. J. Schioppa *et al.*, "Measurement results of the Malta monolithic pixel detector," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 958, Apr. 2020, Art. no. 162404, doi: 10.1016/j.nima.2019.162404.

[10] M. Dyndal *et al.*, "Mini-MALTA: Radiation hard pixel designs for small-electrode monolithic CMOS sensors for the high luminosity LHC," *J. Instrum.*, vol. 15, no. 2, Feb. 2020, Art. no. P02005, doi: 10.1088/1748-0221/15/02/p02005.

[11] S. Meroli, D. Passeri, and L. Servoli, "Energy loss measurement for charged particles in very thin silicon layers," *J. Instrum.*, vol. 6, no. 6, Jun. 2011, Art. no. P06013, doi: 10.1088/1748-0221/6/06/p06013.

[12] W. Snoeys *et al.*, "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 871, pp. 90–96, Nov. 2017, doi: 10.1016/j.nima.2017.07.046.

[13] M. Munker *et al.*, "Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance," *J. Instrum.*, vol. 14, no. 5, May 2019, Art. no. C05013, doi: 10.1088/1748-0221/14/05/c05013.

[14] W. Snoeys, "Monolithic pixel detectors for high energy physics," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 731, pp. 125–130, Dec. 2013, doi: 10.1016/j.nima.2013.05.073.

[15] M. Munker, "Test beam and simulation studies on high resistivity CMOS pixel sensors," Ph.D. dissertation, Dept. Math. Natural Sci., Univ. Bonn, Bonn, Germany, 2018.

[16] H. Hillemanns *et al.*, "Radiation hardness and detector performance of new 180 nm CMOS MAPS prototype test structures developed for the upgrade of the Alice inner tracking system," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2013, pp. 1–5, doi: 10.1109/NSSMIC.2013.6829475.

[17] F. Krummenacher, "Pixel detectors with local intelligence: An IC designer point of view," *Nucl. Inst. Methods Phys. Res. A, Accel., Spectrometers, Detectors Assoc. Equip.*, vol. 305, no. 3, pp. 527–532, 1991, doi: 10.1016/0168-9002(91)90152-G.

[18] C. Gao *et al.*, "A novel source–drain follower for monolithic active pixel sensors," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 831, pp. 147–155, Sep. 2016, doi: 10.1016/j.nima.2016.03.074.

[19] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989, doi: 10.1109/JSSC.1989.572629.

[20] F. Faccio, S. Michelis, D. Cornale, A. Paccagnella, and S. Gerardin, "Radiation-induced short channel (RISCE) and narrow channel (RINCE) effects in 65 and 130 nm MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 6, pp. 2933–2940, Dec. 2015, doi: 10.1109/TNS.2015.2492778.

[21] H. Hillemanns *et al.*, "Total ionizing dose effects in 180 nm CMOS structures for monolithic active pixel sensors for the ALICE inner tracker system detector upgrade," CERN internal doc.

[22] W. J. Snoeys, T. A. P. Gutierrez, and G. Anelli, "A new NMOS layout structure for radiation tolerance," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 2, Nov. 2001, pp. 822–826, doi: 10.1109/NSSMIC.2001.1009683.

[23] R. Cardella *et al.*, "LAPA, a 5 Gb/s modular pseudo-LVDS driver in 180 nm CMOS with capacitively coupled pre-emphasis," *Proc. Sci.*, vol. 313, p. 38, 2017, doi: 10.22323/1.313.0038.

[24] L. Perktold and J. Christiansen, "A multichannel time-to-digital converter ASIC with better than 3 ps RMS time resolution," *J. Instrum.*, vol. 9, no. 1, Jan. 2014, Art. no. C01060, doi: 10.1088/1748-0221/9/01/c01060.

[25] L. Snoj, G. Žerovnik, and A. Trkov, "Computational analysis of irradiation facilities at the JSI TRIGA reactor," *Appl. Radiat. Isotopes*, vol. 70, no. 3, pp. 483–488, Mar. 2012, doi: 10.1016/j.apradiso.2011.11.042.