Optimization of the 65 nm CMOS Linear front-end circuit for the CMS pixel readout at the HL-LHC

L. Gaioni, M. Manghisoni, L. Ratti, V. Re, E. Riceputi, G. Traversi, G. Dellacasa, N. Demaria, S. Garbolino and F. Rotondo

***Abstract*—The Linear front-end is the analog processor chosen for the final integration into the pixel readout chip for the High-Luminosity upgrade of the CMS experiment at the Large Hadron Collider. The front-end has been included in the RD53A chip, designed by the CERN RD53 collaboration and submitted in 2017. An optimized version of the front-end has been designed, submitted and tested in the framework of the RD53B developments. The paper describes the main design improvements of the front-end channel together with the results from the characterization of a small prototype chip including a 16 $\times$ 16 pixel matrix featuring both the RD53A and RD53B version of the front-end.**

***Index Terms*—CMOS Front-end electronics, High-Luminosity LHC, Pixel readout chip, Low-noise analog front-end, ionizing radiation effects.**

I. Introduction

HIGH -Luminosity (HL) upgrades of the CMS and ATLAS experiments at the Large Hadron Collider (LHC) will require completely new trackers facing unprecedented particle rates and radiation levels. At the core of the systems, advanced pixel sensors will deal with rates close to 3 GHz/cm^2, with a predicted total ionizing dose around 1 Grad(SiO_2) and a 1 MeV neutron equivalent fluence of 2×10^{16} cm^{-2} accumulated in about 10 years of operation [1], [2]. Accuracy in momentum measurement will benefit from an elementary pixel size shrunk down to 50 μm $\times$ 50 μm (or 25 μm $\times$ 100 μm) in the layers closer to the interaction region. Compared to the present pixel detectors, featuring a typical thickness around 300 μm, thinner sensors will be adopted in the High-Luminosity experiment upgrades to improve radiation tolerance and reduce multiple scattering effects [3], [4]. With an active volume thickness close to 100 μm, the pixels will deliver smaller signals, further reduced by radiation damage. For the purpose of preserving the detection efficiency, operation of the front-end chips at small thresholds (around 1000 electrons or lower) has to be envisaged, tightening the requirements on the noise and threshold dispersion performance of the analog processors integrated in the pixel readout chips. Highly efficient readout chips are being developed in a 65 nm CMOS technology by the CERN RD53 collaboration [5], a combined effort of the ATLAS and CMS communities which led to the submission, in 2017, of the RD53A chip [6]. The RD53A integrated circuit is a large-scale demonstrator which paved the way to the design of the production chips for the two experiments, including a matrix of 400 $\times$ 192 pixels (about half the size of the final chips) with an elementary cell size of 50 μm $\times$ 50 μm. The matrix is divided in three regions featuring different analog front-end flavours, called Synchronous (spanning 128 columns of the RD53A matrix), Linear (136 columns) and Differential (136 columns), which have been thoroughly characterized in view of their integration into the production chips [7], [8]. After a detailed review process, the Linear front-end has been chosen for the integration in the CMS chip, whereas the Differential front-end will be integrated in the ATLAS one.

While very good results were obtained in the test campaigns of the RD53A chip, a number of improvements and few design fixes have been implemented in the second generation of the RD53 pixel chips, called RD53B [9]. In particular, the characterization of the RD53A Linear front-end revealed a larger time-walk with respect to the one measured for the Synchronous and the Differential processors, together with an issue related to the in-pixel threshold tuning DAC (usually adopted in front-end channels for threshold equalization) in specific operating conditions.

In this work, the optimization of the Linear front-end will be discussed, together with the results relevant to the characterization of a small prototype including the RD53A and RD53B version of the front-end. Section 2 will introduce the general architecture of the analog processor with a description of the circuits common to both the designs. Section 3 will deal with the specific optimization that led to the design of the RD53B flavour, whereas Section 4 is concerned with the main test results, before and after irradiation up to a total ionizing dose (TID) of 1 Grad(SiO_2), of the prototype chip.

II. Front-end design

The RD53A and RD53B versions of the Linear front-end share the common architecture shown in Fig. 1, implementing a shaper-less analog processor with a charge sensitive amplifier (CSA, or preamplifier) driving a threshold discriminator exploited for time-over-threshold (ToT) conversion of the input signal amplitude. The readout channel is equipped with a tuning DAC for threshold equalization and with a calibration circuit generating a test signal through an injection capacitor

Manuscript received August 5, 2021

L. Gaioni, M. Manghisoni, V. Re, E. Riceputi and G. Traversi are with Università di Bergamo, I-24044 Dalmine (BG), Italy and INFN, Sezione di Pavia, Via Bassi 6, I-27100 Pavia, Italy, (phone: +39 035 2052006; fax: +39 035 562779; email: luigi.gaioni@unibg.it).

L. Ratti is with Università di Pavia, Via Ferrata 5, I-27100 Pavia, Italy and INFN, Sezione di Pavia, Via Bassi 6, I-27100 Pavia, Italy.

G. Dellacasa, N. Demaria, S. Garbolino and F. Rotondo are with INFN, Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy.

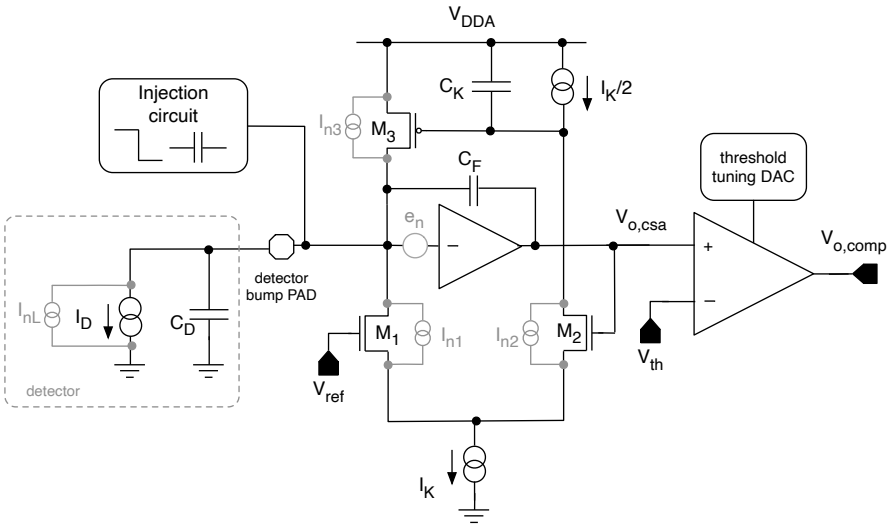

Fig. 1. Simplified schematic diagram of the Linear analog front-end. Components in light grey are used for the noise analysis of the channel.

connected to the CSA input. The sensor, modelled by the current source I_D in parallel with the capacitance C_D, can be connected to the front-end by means of a bump pad laid out in the top metal layer.

A. Charge sensitive amplifier

The CSA features a gain stage with a Krummenacher feedback network [10] including the differential pair M_1-M_2, the M_3 pMOS transistor, the C_K capacitor, and two current generators (I_K, $I_{K/2}$). In quiescent conditions, the drain current of M_1, provided my M_3, is the same as the one flowing in M_2 ($I_K/2$). The negative feedback ensures that the DC output of the charge sensitive amplifier be equal to V_{ref}, applied to the gate terminal of M_1. As discussed later in this section, the Krummenacher network provides, for a relatively large input signal (from about 2000 electrons input charge upwards in simulations), a linear discharge of the feedback capacitance C_F and, at the same time, makes it possible to compensate for radiation-induced detector leakage currents.

The gain stage of the CSA is based on a folded cascode architecture biased with a current of 3 μA flowing in the input branch and a current close to 200 nA in the cascode branch. The transfer function, $A(s)$, of the gain stage can be modelled by means of:

$$A(s) = \frac{A_0}{1 + s\tau}, \tag{1}$$

with circuit simulations revealing a DC gain close to 76 dB and a cut-off frequency around 140 kHz.

For small input charge signals, the behavior of the circuit in Fig. 1 can be studied by means of device small signal models. Assuming that C_K tends to infinity, the CSA transfer function, $T(s)$, can be written as

$$T(s) = \frac{V_{o,csa}}{Q_{in}} \simeq \frac{2/g_{m_n}}{(1 + s\tau_1)(1 + s\tau_2)}, \tag{2}$$

where Q_{in} is the charge delivered by the sensor, $V_{o,csa}$ is the Laplace transform of the preamplifier output signal, $v_{o,csa}$, g_{m_n} is the transconductance of either M_1 or M_2, $\tau_1 = \frac{2C_F}{g_{m_n}}$ and $\tau_2 = \frac{\tau}{A_0}\frac{C_D^* + C_F}{C_F}$. The capacitance C_D^* is computed as the sum of the detector capacitance C_D and the input capacitance of the preamplifier.

Assuming $\tau_1 \gg \tau_2$, the response of the CSA in the time domain to a Dirac delta shaped current pulse, $\delta(t)Q_{in}$, is ruled by

$$v_{o,csa}(t) \simeq H(t)\frac{Q_{in}}{C_F}\left(e^{-t/\tau_1} - e^{-t/\tau_2}\right), \tag{3}$$

$H(t)$ being the Heaviside function. The preamplifier output voltage features a maximum at the peaking time, t_p, given by

$$t_p \simeq \tau_2 ln\left(\frac{\tau_1}{\tau_2}\right). \tag{4}$$

Eq. (4) shows that t_p is an increasing function of $\tau_1 = 2C_F/g_{m_n}$, meaning that the peaking time decreases by increasing the Krummenacher current I_K, which in turn increases the transconductance g_{m_n} of transistors M_1 and M_2. Under the hypothesis of linearity, and this is the case of the circuit discussed here, the charge sensitivity, G_Q, of the CSA can be computed as the peak value of the Laplace anti-transform of $T(s)$. It can be shown that

$$G_Q \simeq \frac{1}{C_F}\left(\zeta g_{m_n}\right)^{\zeta g_{m_n}}, \tag{5}$$

where $\zeta \simeq \frac{1}{2}\frac{\tau}{A_0}\frac{C_D^* + C_F}{C_F^2}$. G_Q is a decreasing function of g_{m_n} as long as the transconductance of M_1 and M_2 is smaller than $\frac{1}{e\zeta}$ (around 2 μS in the actual design of the front-end). For g_{m_n} approaching zero, the charge sensitivity can be written as $G_Q \simeq C_F^{-1}$. For both the RD53A and RD53B Linear front-end the charge sensitivity was designed to be close to 25 mV/ke$^-$.

For large input signals, the differential pair in the Krummenacher feedback network gets saturated: M_1 is switched off and M_2 carries the current I_K. Half of this current flows through the C_K capacitor: nonetheless, the voltage variation at the gate of M_3 and, thus, in its drain current, can be neglected if C_K is sufficiently large. Since M_1 is off, a constant current $I_K/2$ will

linearly discharge the feedback capacitor C_F. The response of the CSA to a large input signal can thus be approximated by means of

$$v_{o,csa}(t) = \frac{Q_{in}}{C_F}\left(1 - e^{-t/\tau_2}\right) - \frac{I_K}{2C_F}t. \quad (6)$$

For $t > t_e$ the discharge process comes to an end and the output voltage settles to the baseline, t_e being such that

$$\frac{Q_{in}}{C_F}\left(1 - e^{-t_e/\tau_2}\right) - \frac{I_K}{2C_F}t_e = 0. \quad (7)$$

The CSA output voltage in response to a large signal thus resembles a triangular pulse with a maximum at the peaking time, $t_{p,ls}$, which is found to be a monotonically increasing function of the input charge, according to

$$t_{p,ls} = \tau_2 ln\left(\frac{2Q_{in}}{I_K\tau_2}\right). \quad (8)$$

As mentioned, the Krummenacher feedback also serves the purpose of compensating for the radiation-induced detector leakage, I_L, which is expected to increase up to levels close to ten nanoamps per pixel in the innermost layer of the tracker in the phase II upgrade of the experiments [11], [12]. Such a compensation mechanism is implemented via transistor M_3, which supplies the drain current of M_1 and, at the same time, provides a DC path for the detector leakage current. In circuit simulations, the CSA was found to cope with I_L currents larger than 100 nA.

B. Noise performance

The noise performance of the charge sensitive amplifier is mainly dictated by the channel thermal noise of the CSA input transistor together with the contribution of the transistors included in the Krummenacher feedback network (M_1, M_2 and M_3). These contributions are represented in Fig. 1 by the voltage source e_n and the current sources i_{n_1}, i_{n_2} and i_{n_3}, while the shot noise contribution of the detector leakage is modelled by means of the current source i_{n_L}. The noise power spectral density associated with the aforementioned generators is given by

$$\frac{\overline{de_n^2}}{df} = \frac{4k_BT\gamma_{in}}{g_{m_{in}}}, \quad (9)$$

$$\frac{\overline{di_{n_1}^2}}{df} = \frac{\overline{di_{n_2}^2}}{df} = 4k_BT\gamma_n g_{m_n} \equiv \frac{\overline{di_{n_{1,2}}^2}}{df}, \quad (10)$$

$$\frac{\overline{di_{n_3}^2}}{df} = 4k_BT\gamma_p g_{m_p}, \quad (11)$$

$$\frac{\overline{di_{n_L}^2}}{df} = 2qI_L, \quad (12)$$

where k_B is the Boltzmann's constant, q is the elementary charge, T is the absolute temperature, $g_{m_{in}}$ and γ_{in} are the transconductance and the channel thermal noise coefficient of the CSA input device, g_{m_p} is the transconductance of M_3 and γ_n and γ_p are the noise coefficients of transistors M_1 (or M_2) and M_3, depending on the degree of inversion of the device channel [13].

The mean square noise at the CSA output, $\overline{v_n^2}$, can be written as

$$\overline{v_n^2} = \int_0^\infty \left\{ \left[(g_{m_n}/2)^2 + \omega^2(C_D^* + C_F)^2\right] |T(j\omega)|^2 \frac{\overline{de_n^2}}{df} + \right. \\ \left. + \frac{1}{2}|T(j\omega)|^2 \frac{\overline{di_{n_{1,2}}^2}}{df} + |T(j\omega)|^2 \left(\frac{\overline{di_{n_3}^2}}{df} + \frac{\overline{di_{n_L}^2}}{df} \right) \right\} df. \quad (13)$$

The equivalent noise charge (ENC) of the front-end can be obtained by dividing the square root of $\overline{v_n^2}$ by the charge sensitivity, G_Q, of the preamplifier. Assuming for simplicity that $G_Q = C_F^{-1}$, it can be shown that

$$ENC \simeq \sqrt{k_BTC_F\left[\frac{A_0}{\tau}\frac{C_D^* + C_F}{g_{m_{in}}}\gamma_{in} + \gamma_n + 2\frac{g_{m_p}}{g_{m_n}}\gamma_p + \gamma_l\right]}, \quad (14)$$

where $\gamma_l = \frac{qI_L}{kTg_{m_n}}$. In such an approximated expression, the ENC depends on the $A_0/(\tau g_{m_{in}})$ ratio, which is a constant function of $g_{m_{in}}$. This suggests that a variation of the current biasing the CSA input device and, thus, of its transconductance, should in principle not affect the noise performance of the front-end. In other words, a variation of the channel thermal noise of the CSA input transistor, induced by a change in its transconductance, is counterbalanced by a variation of the preamplifier gain-bandwidth product.

It is worth noticing that, in (14), the ratio g_{m_p}/g_{m_n} is close to 1 when the front-end is connected to a non-leaky detector: transistors M_3 and M_1, operating in weak inversion in the actual design, indeed carry the same current $I_K/2$. On the other hand, in the presence of a detector leakage current, the transconductance of M_3 becomes larger than the one of M_1, giving rise to a non negligible increase in the ENC.

Eq. (14) has been obtained by assuming an ideal comparator connected to the preamplifier output: obviously the bandwidth of the comparator itself plays a role in the noise performance of the overall channel. Nonetheless, in order to keep the noise analysis straightforward, the effects of the comparator have been neglected.

C. Threshold discriminator and tuning DAC

The block diagram of the threshold discriminator integrated in the RD53A and RD53B versions of the Linear front-end is shown in Fig. 2 a). It includes a transconductance stage, whose output current depends on the difference between the signal coming from the preamplifier and the global threshold (V_{th}), connected to a transimpedance amplifier (TIA) based on a slightly modified version of the Träff amplifier [14]. The signal from the TIA stage is fed to a couple of inverters, supplied in the digital domain, which provides the comparator output signal with sharp transitions and consolidated logic levels. Local tuning of the threshold is performed by properly setting a DC current, I_{trim}, generated by an in-pixel threshold tuning DAC, controlled by the TDAC digital input (4 bits in RD53A, 5 bits in RD53B). The actual implementation of the transimpedance stage and of the tuning DAC is different in the two flavours of the front-end and will be discussed in the next section.

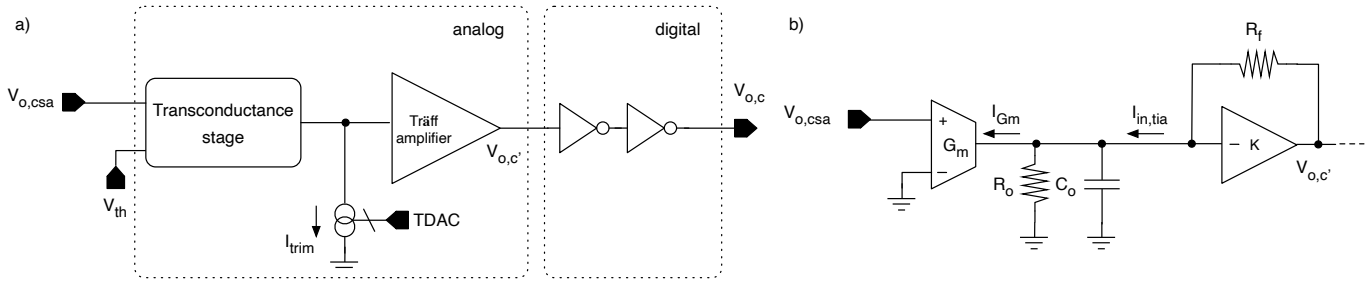

Fig. 2. Block diagram of the comparator integrated in the Linear analog front-end (a) and the equivalent circuit used for small signal analysis (b).

As the signal at the preamplifier output gets larger than the actual pixel threshold, a positive output step is generated at the comparator output, indicating the occurrence of a significant event. When $v_{o,csa}$ crosses the threshold level downwards, the comparator output toggles and the generated time-over-threshold interval is proportional to the front-end input signal amplitude.

Small signal analysis of the comparator can be carried out with the aid of the simplified schematic shown in Fig. 2 b), where the transimpedance stage has been modelled by means of an ideal amplifier with a DC gain $K \gg 1$ and a feedback resistor R_f. By assuming an infinite bandwidth for the TIA gain stage, the transfer function, $C(s)$, of the analog section of the comparator can be easily shown to be

$$C(s) = \frac{V_{o,c'}}{V_{o,csa}} \simeq K G_m \frac{\frac{R_f}{K} \parallel R_o}{1 + s\tau_{c,cl}}, \tag{15}$$

where $V_{o,c'}$ is the Laplace transform of the signal at the output of the TIA, G_m and R_o are, respectively, the transconductance and the output resistance of the comparator input stage, C_o is the total capacitance shunting the TIA input and

$$\tau_{c,cl} = \left(\frac{R_f}{K} \parallel R_o \right) C_o \simeq \frac{R_f}{K} C_o \tag{16}$$

is the closed-loop time constant associated with such a node. On the other hand, the open-loop time constant, $\tau_{c,ol}$, associated with the same node is equal to

$$\tau_{c,ol} = (R_f \parallel R_o)\, C_o. \tag{17}$$

Hence, the closed-loop time constant is smaller than the open-loop one (a factor of K if $R_o \gg R_f$) and the negative feedback in the transimpedance amplifier boosts the response speed of the comparator.

III. Front-end optimization

As mentioned in the Introduction, two main drawbacks associated with the Linear front-end have emerged during the characterization of the RD53A chips: limited time-walk performance and threshold tuning issues in specific conditions. In particular, after irradiation and with the chip operated at low temperatures, the input-output characteristics (namely, with reference to Fig. 2-a, I_{trim} vs TDAC code) of the in-pixel tuning DAC is strongly non-linear, with a saturation effect taking place at the higher TDAC codes.

As far as the response speed is concerned, compared to the RD53A Differential and Synchronous, the Linear front-end features a significantly larger time-walk, close to 40 ns for a small (close to the threshold) input charge. Such a value, to be compared to the HL-LHC bunch crossing period of 25 ns, is actually much larger than the preamplifier peaking time, which is found to be close to 20 ns in circuit simulations. This points to a significant time-walk contribution from the comparator, whose design has been reviewed in the RD53B version of the Linear front-end.

A. Time-walk optimization

A partial redesign of the comparator has been implemented in the RD53B Linear front-end, with the aim of improving its time-walk performance. In particular, the revised front-end integrates a slightly modified version of the RD53A transimpedance amplifier, whose schematic diagram is shown in Fig. 3. The version conceived for RD53B is shown in Fig. 4: in both the designs, the forward gain stage of the TIA is implemented by means of an inverter including transistors M_3 and M_4. The feedback network, which has been reduced to a simple resistor, R_f, in the discussion of Section 2.3, includes two diode-connected transistors (M_{1b} and M_{2b}) in the RD53A version, limiting the inverter crowbar current. Those transistors have been removed in the RD53B version, where M_5, whose gate is kept to a constant bias voltage, sets the static current of the gain stage around 200 nA. The C_m capacitor, implemented by means of a pMOS device, provides a path to ground for the source of transistor M_3.In Figs. 3 and 4, $i_{in,tia}$ is the input current signal of the transimpedance amplifier, which can vary between the saturation values $-I_{sat}$ and $+I_{sat}$, $|I_{sat}|$ being set to 1 μA in the actual implementation of the transconductance stage, with the saturation occurring when the output level of the preamplifier is around 75 mV below or above the comparator threshold. In steady state, if the output level of the preamplifier is equal to the discriminator threshold voltage, $i_{in,tia}$ is zero and $v_{o,c'} \simeq V_{dd}/2$. With reference to the small signal analysis carried out in Section 2.3, it is easy to show that the equivalent resistance R_f for the RD53A design can be computed as

$$R_f = \begin{cases} \frac{1}{g_{m,1a}} + \frac{1}{g_{m,1b}}, & i_{in,tia} < 0 \\ \frac{1}{g_{m,2a}} + \frac{1}{g_{m,2b}}, & i_{in,tia} > 0, \end{cases} \tag{18}$$

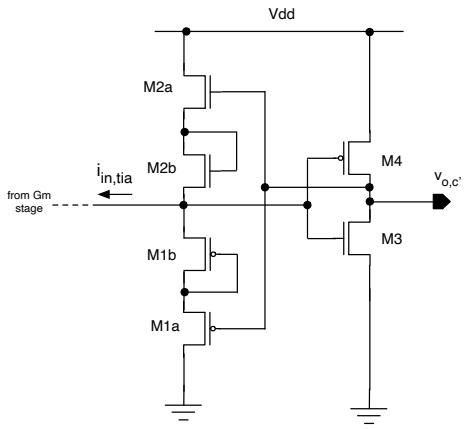

Fig. 3. Schematic diagram of the transimpedance stage integrated in the RD53A front-end comparator.

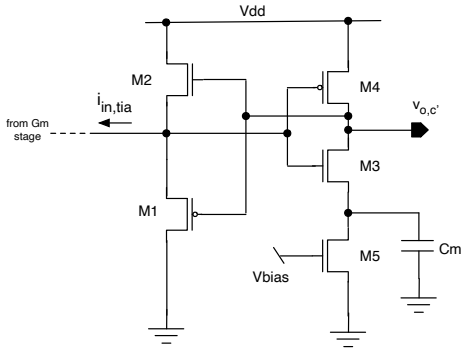

Fig. 4. Schematic diagram of the transimpedance stage integrated in the RD53B front-end comparator.

where $g_{m,i}$ is the transconductance of transistor M_i. For the RD53B version, R_f can be shown to be

$$R_f = \begin{cases} \frac{1}{g_{m,1}}, & i_{in,tia} < 0 \\ \frac{1}{g_{m,2}}, & i_{in,tia} > 0. \end{cases} \tag{19}$$

Since M_1, M_{1a} and M_{1b} are equally sized transistors carrying the same current (the same holds for M_2, M_{2a} and M_{2b}), it is reasonable to assume that $g_{m,1a} \simeq g_{m,1b} \simeq g_{m,1}$ and $g_{m,2a} \simeq g_{m,2b} \simeq g_{m,2}$. In this case, the closed-loop time constant associated with the input node of the TIA, as obtained from (16), is a factor of 2 smaller for the RD53B version of the Linear front-end.

More importantly, on one hand, the presence of the diode-connected transistors in the RD53A design limits the current in the inverter amplifier, on the other hand, it prevents the gain stage from being operated in the linear region for an extended range of the input current $i_{in,tia}$. Assuming weak inversion operation for the transistors integrated in the stage, the same threshold voltage $V_{th,n}$ for M_{1a} and M_{1b} and the same threshold voltage $V_{th,p}$ for M_{2a} and M_{2b}, it can be shown that for $i_{in,tia}$ in

$$\left[-I_{sat}, -I_{D0p}\frac{W_p}{L_p}e^{\frac{\frac{V_{dd}}{4}-|V_{th,p}|}{\phi_t}}\right] \cup \left[I_{D0n}\frac{W_n}{L_n}e^{\frac{\frac{V_{dd}}{4}-V_{th,n}}{\phi_t}}, +I_{sat}\right] \tag{20}$$

the inverter is out of its dynamic range, and the response speed of the TIA, in a sense, is regulated by the open-loop time constant defined in (17). This results in an overall slow comparator response speed, negatively affecting the time-walk performance of the front-end. In the previous equation, ϕ_t is the thermal voltage, W_p/L_p and W_n/L_n are, respectively, the aspect ratios of transistors M_{1a} (or M_{1b}) and M_{2a} (or M_{2b}) while the currents $I_{D_{0p}}$ and $I_{D_{0n}}$ are defined as [15]

$$I_{D_{0p}} = 6\mu_p C_{ox}\phi_t^2,\ I_{D_{0n}} = 6\mu_n C_{ox}\phi_t^2, \tag{21}$$

μ_p and μ_n being, respectively, the hole and electron mobility and C_{ox} the oxide capacitance per unit area.

On the other hand, in the RD53B design, the TIA output accommodates just one gate-to-source voltage (the one relative to M_1 if $i_{in,tia} < 0$ or the one associated with M_2 in the opposite case), and the inverter stage is always operated in the linear region. Hence, the dynamic characteristics of the TIA are regulated by the faster closed-loop time constant, boosting the response speed of the comparator and, ultimately, the front-end time-walk performance.

Fig. 5 shows the simulated time-walk for the RD53A (red curve) and RD53B (blue curve) versions of the Linear front-end as a function of the input charge, ranging from 1000 up to 30000 electrons. The simulation, carried out with a threshold set to 1000 electrons, points to a significant reduction (of the order of 25% for a small input signal) of the time-walk in the RD53B front-end. It is worth recalling that such an improved performance has been achieved with just a marginal increase (200 nA) of the overall static current consumption of the analog processor (around 5 μA).

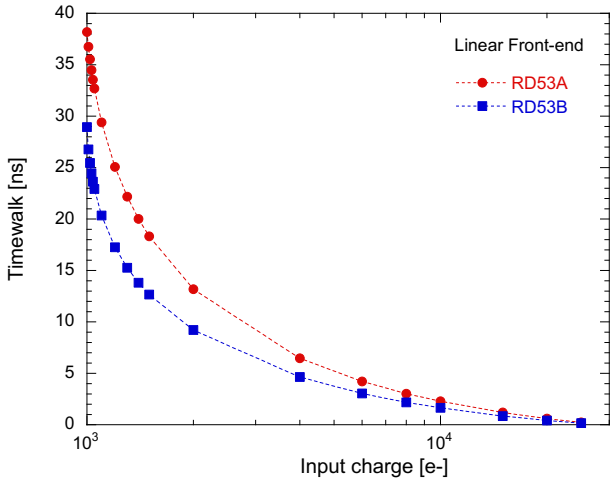

Fig. 5. Simulated time-walk as a function of the input charge for the RD53A (red) and RD53B (blue) versions of the Linear front-end.

B. Threshold tuning optimization

The threshold tuning capabilities of the Linear front-end have been enhanced with the design of a new in-pixel trimming DAC (TDAC). The schematic diagrams of the tuning DACs for the RD53A and RD53B front-ends are shown in Fig. 6 and Fig. 7, respectively. The RD53A Linear front-end features a 4-bit, binary weighted current DAC with cascoded current mirrors, whereas the RD53B version includes a more compact, 5-bit DAC with the same binary weighted architecture featuring regular current mirrors. The TDAC trimming bits (B0-B3

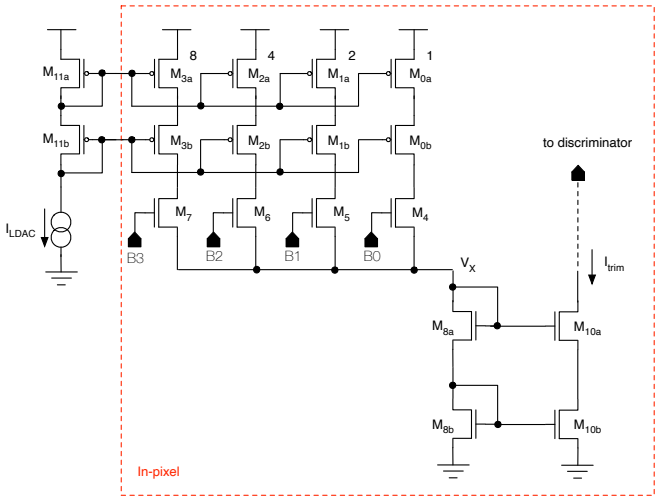

Fig. 6. Schematic diagram of the threshold tuning DAC integrated in the RD53A front-end.

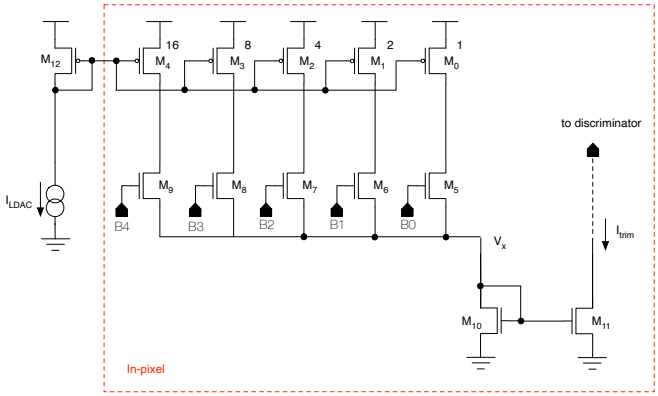

Fig. 7. Schematic diagram of the threshold tuning DAC integrated in the RD53B front-end.

in RD53A, B0-B4 in RD53B) control a number of switches implemented with nMOS transistors in both the designs. The output node of the tuning DAC, generating a current I_{trim}, is connected to the output of the transconductance stage of the comparator, as shown in Fig. 2 a). As discussed in [16], the effective threshold variation, ΔV_{th}, at the comparator input in response to a variation ΔI_{trim} of the TDAC output current can be expressed as

$$\Delta V_{th} = \frac{\Delta I_{trim}}{G_m}. \tag{22}$$

The maximum value of I_{trim} determines the TDAC dynamic range and is set by the current I_{LDAC}, generated in the matrix periphery and mirrored in the pixel cells.

As mentioned, a saturation effect in the input-output characteristics of the tuning DACs was detected during the characterization of RD53A chips. Such an effect emerged during the test of irradiated chips at cold, which needed to be operated with large I_{LDAC} currents as a consequence of a significant increase in the un-tuned threshold dispersion. In these conditions, the voltage V_X at the drain of M_{8a} in the RD53A design is large enough to push switch transistors out of the triode region, resulting in the saturation effect visible in the

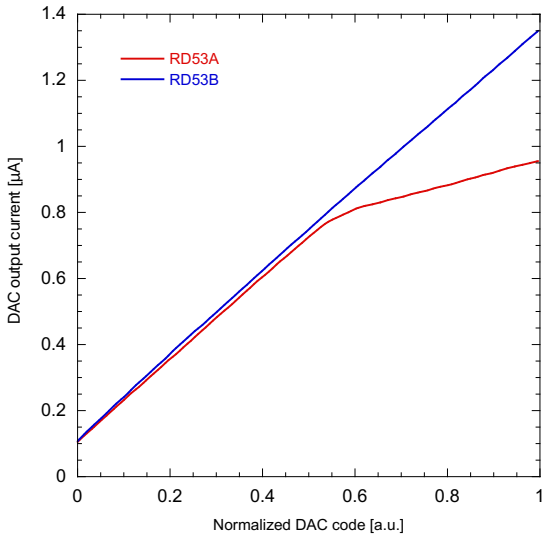

Fig. 8. Simulated TDAC output current (I_{trim}) as a function of the normalized TDAC code for the two versions of the front-end.

red curve of Fig. 8, which shows the simulated I_{trim} current as a function of the normalized DAC input code for the two versions of the tuning DACs. In such a simulation, the I_{LDAC} current was set to 30 μA, which is significantly higher than the current needed for proper operation of un-irradiated chips at room temperature ($I_{LDAC} \simeq 14\mu A$). The simulated input-output characteristics of the DAC integrated in the RD53B Linear front-end is instead represented by the blue curve of Fig. 8: the removal of cascode structures forces the V_x voltage to lay in a range that guarantees switch transistors to operate in triode, preventing the saturation effect from kicking in. Moreover, the choice of regular current mirrors in the RD53B TDAC led to a more compact analog macro layout, enabling the integration of additional digital functionalities in the pixel cell.

IV. Test Results

A prototype chip including the RD53A and RD53B Linear front-end has been fabricated and tested in view of the submission of the RD53B-CMS pixel readout chip, which took place in June 2021. The characterization of the prototype has been carried out before and after exposure to total ionizing doses (TIDs) up to 1 Grad of X-rays. The following sections provide a brief description of the prototype chip and discuss the main results from the testing activity.

A. Prototype chip

The prototype chip, whose microphotograph is shown in Fig. 9, includes a matrix of 16×16 front-end channels with a pitch of 50 μm × 50 μm laid out in the so-called analog island arrangement as in the RD53A chip [17]. The matrix is divided in two equally sized regions featuring the RD53A and RD53B version of the front-end. The analog processor includes the same calibration circuit, featuring an injection capacitance of 8.5 fF, integrated in RD53A, able to deliver programmable test signals at the front-end input, while a sensor bump pad has

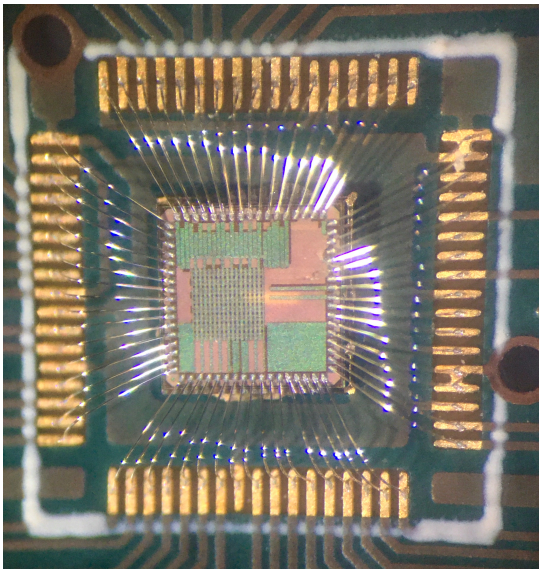

Fig. 9. Microphotograph of the prototype chip including a 16 × 16 matrix featuring the RD53A and RD53B Linear front-end.

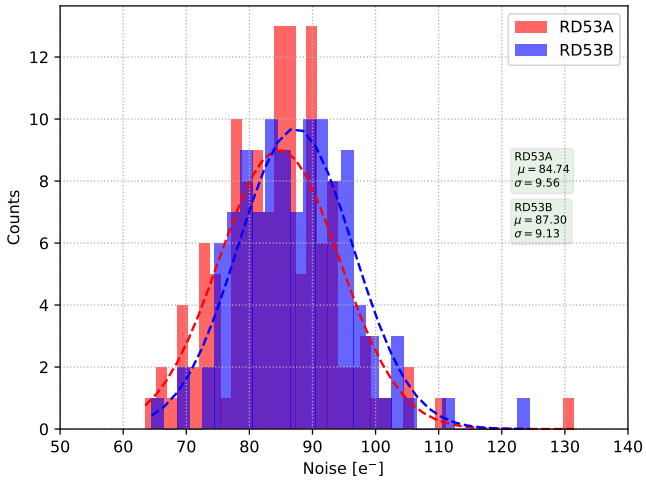

Fig. 10. Measured ENC distributions for the RD53A (red) and RD53B (blue) Linear front-end. The text boxes report the mean value of the ENC and its standard deviation.

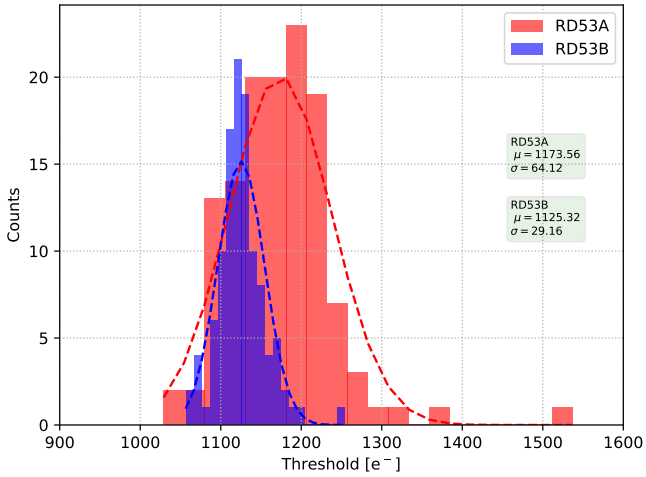

Fig. 11. Measured threshold distributions for the RD53A (red) and RD53B (blue) Linear front-end. The text boxes report the mean value of the threshold and its standard deviation, referred to as threshold dispersion.

been laid out with the same size as in the RD53B-CMS chip. The front-end channel is equipped with two detector emulating capacitors of 50 fF and 100 fF which can be selectively shunted to the preamplifier input, thus enabling 4 different configurations (namely 0, 50, 100 and 150 fF). The gate of the preamplifier input transistor can also be connected to a leakage injection circuit by means of a controllable CMOS switch, activating a simple nMOS current mirror whose output emulates the presence of a DC detector leakage.

A number of samples has been characterized before and after irradiation with a purposely developed test setup based on a Xilinx FPGA board and a data acquisition software written in LabView. The results relevant to the main analog performance parameters, such as noise, threshold dispersion, time walk and time-over-threshold, are discussed in the following subsections.

B. Results before irradiation

The noise performance of the front-end has been assessed in terms of equivalent noise charge, which, together with threshold data, has been obtained from charge scans on the readout channels. In a charge scan, the input charge signal, Q_{in}, of the front-end is made to vary in an adequate range and the hit efficiency of the comparator is measured. The resulting efficiency curve can be fitted by means of the function η_{hit}, defined as

$$\eta_{hit} = \frac{1}{2}\left[1 + erf\left(\frac{Q_{in} - p_1}{\sqrt{2}p_2}\right)\right], \qquad (23)$$

where the fitting parameters p_1 and p_2 correspond to the pixel threshold and ENC, respectively. As an example, Fig. 10 and Fig. 11 show the measured distributions for the ENC and the threshold, evaluated for both the front-end versions at room temperature with the readout channels configured with a detector emulating capacitor of 50 fF. As expected, the threshold dispersion (i.e. the standard deviation of the threshold distribution) for the RD53B Linear front-end is roughly a factor of two smaller than the one measured for RD53A, thanks to increased resolution in the RD53B version of the in-pixel tuning DAC.

Fig. 12 shows the equivalent noise charge as a function of the detector emulating capacitor, C_D, for the RD53A and RD53B front-ends integrated in one sample of the prototype chip. In particular, the plot shows the mean value of the ENC measured at room temperature for 128 pixels featuring the RD53A design and for the same number of pixels implementing the RD53B architecture. The error bars, in this figure and in the following ones, correspond to ±1 standard deviation of the data set. The ENC was found to be smaller than 100 e r.m.s. for both the front-end versions for a detector capacitance of 50 fF, a value close to the one foreseen for the pixel sensors at the HL-LHC experiments. In agreement with (14), the plot shows an increase of the ENC with C_D, with a $dENC/dC_D$ slope in the examined range close to 380 e^-/pF and 420 e^-/pF for the RD53A and RD53B version, respectively. The larger

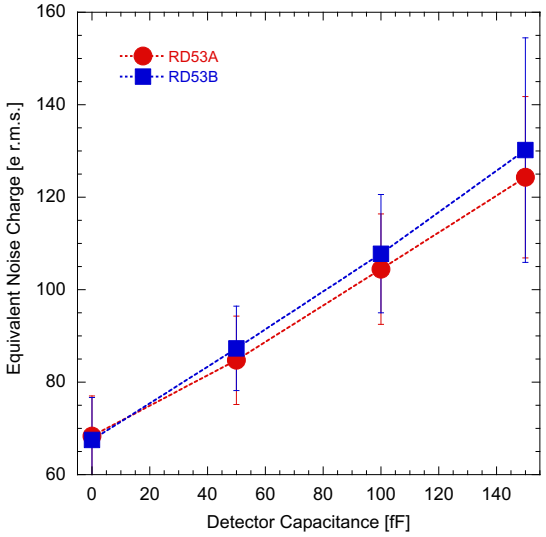

Fig. 12. ENC as a function of the detector emulating capacitor. Blue squares are relevant to RD53A Linear front-end data, red circles to RD53B ones. The error bars correspond to ± 1 standard deviation of the data set.

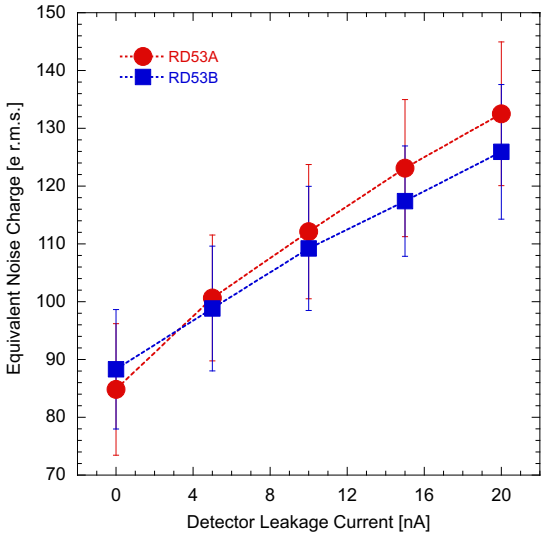

Fig. 13. ENC as a function of the detector leakage current, for C_D=50 fF. Blue squares are relevant to RD53A Linear front-end data, red circles to RD53B ones. The error bars correspond to ± 1 standard deviation of the data set.

slope detected for the latter could be ascribed to the larger bandwidth featured by its comparator stage.

The effect of the detector leakage current on the noise performance of the front-ends is visible in Fig. 13, which shows the ENC for a per-pixel emulated leakage current ranging between 0 and 20 nA. As predicted by (14), the ENC increases with the detector leakage, with a growth which was found to be slightly larger in the RD53A flavour.

The improvement in terms of channel response speed is noticeable in Fig. 14, which shows the time-walk for both the front-end versions as a function of an input charge ranging from the threshold, set around 1150 electrons, to 1600 electrons. In agreement with simulation results, the measured time-walk for an input signal close to threshold is around 10 ns smaller in the

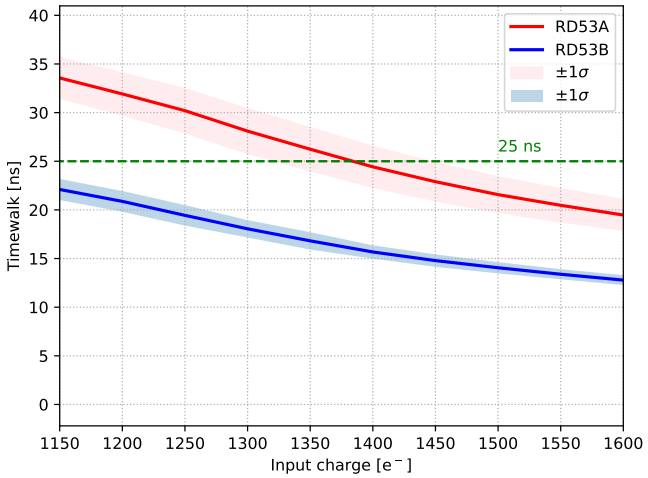

Fig. 14. Time-walk as a function of the input charge for RD53A (red) and RD53B (blue) front-ends. The threshold was set around 1150 electrons and the measurement was performed at -10°C. The filled area boundaries correspond to $\pm 1\sigma$ of the data set.

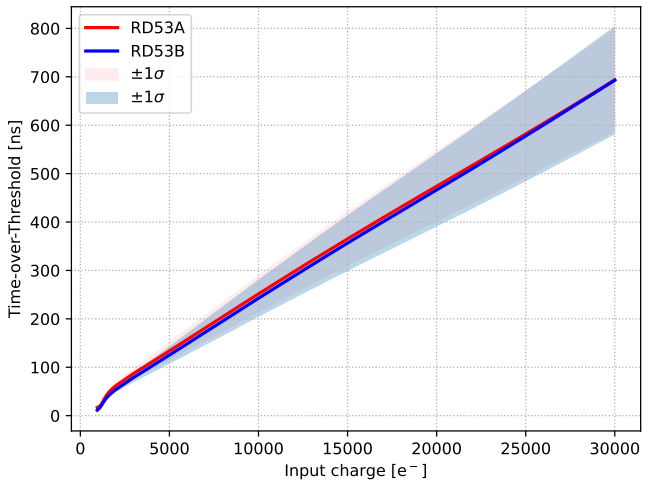

Fig. 15. Time-over-Threshold (ToT) as a function of the input charge for RD53A (red) and RD53B (blue) front-ends. The threshold was set around 1150 electrons and the measurement was performed at -10°C. The filled area boundaries correspond to $\pm 1\sigma$ of the data set.

RD53B Linear front-end. In particular, for the tested sample, a time-walk of 22 ns, smaller than the LHC bunch crossing period, has been obtained for the RD53B flavour, whereas a value close to 33 ns was detected for the RD53A version.

The time-over-threshold characteristics of the two front-ends are shown in Fig. 15, where the measured ToT is plotted as a function of an input signal charge, Q_{in}, ranging from 1150 electrons (the threshold value) up to 30000 electrons. The Krummenacher current (I_K in Fig. 1) was set in such a way to obtain a ToT close to 130 ns in response to an input charge of 6000 electrons, according to RD53A specifications. For both the front-end versions, the response was found to be linear for $Q_{in} > 2000$ electrons, that is, as mentioned in Section 2.1, the charge at which the Krummenacher feedback network of the charge preamplifier gets saturated. The measured dispersion of the ToT (around 20 ns r.m.s. at Q_{in} = 6000 electrons) is very similar for the two flavours of the front-end.

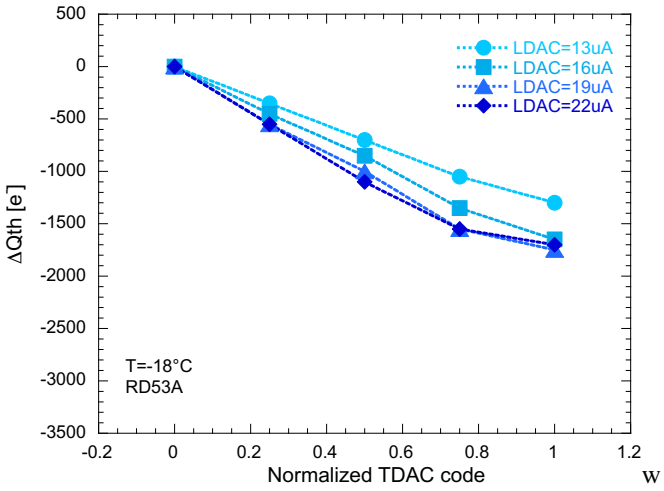

Fig. 16. Threshold variation, ΔQ_{th}, as a function of the normalized TDAC input code, for I_{LDAC} ranging from 13 μA to 22 μA. Data measured for the RD53A Linear front-end at -18°C

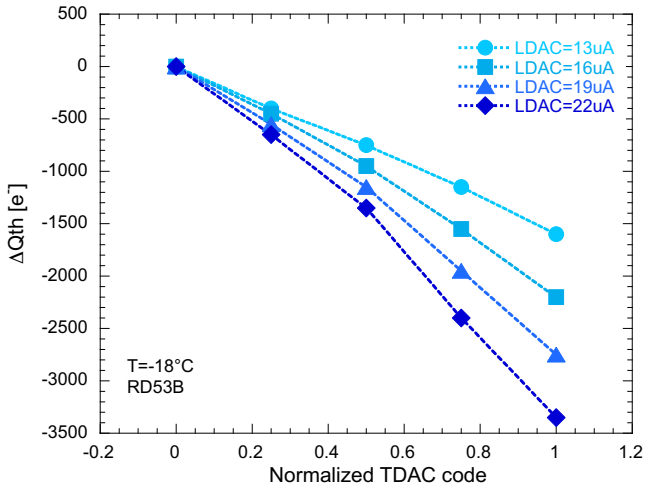

Fig. 17. Threshold variation, ΔQ_{th}, as a function of the normalized TDAC input code, for I_{LDAC} ranging from 13 μA to 22 μA. Data measured for the RD53B Linear front-end at -18°C

The saturation effect taking place in the threshold tuning DAC of the RD53A Linear front-end is clearly visible in Fig. 16, which shows the TDAC input-output characteristics measured at -18°C for different I_{LDAC} settings. In particular the plot shows the threshold charge variation, ΔQ_{th} as a function of the normalized TDAC input code, for I_{LDAC} ranging from 13 μA to 22 μA. As discussed in Section 3.2, the dynamic range (i.e. the maximum value of $|\Delta Q_{th}|$) of the tuning DAC is expected to increase with the I_{LDAC} current: this is, however, not the case of the RD53A tested sample, where the dynamic range is stuck around 1700 electrons for large I_{LDAC} currents. The issue was fixed with the RD53B TDAC design, whose measured input-output characteristics is shown in Fig. 17, and where the tuning DAC dynamic range can be properly set by adjusting the I_{LDAC} current.

C. Results after irradiation

In order to asses the radiation tolerance of the front-ends, two samples of the prototype chip were exposed to total ionizing doses up to 1 Grad(SiO_2) of X-rays. The chips were irradiated at the CERN EP-ESE irradiation system with a Seifert RP149 X-ray machine. The temperature of the chips was kept around -10°C while some basic charge scans were continuously run during the irradiation in order to force some analog and digital activity in the chip. The dose rate was around 11 Mrad(SiO_2)/h and the measurements were performed immediately after irradiation for each TID step.

The two irradiated chips were biased in different conditions: the first one was biased in such a way to achieve a ToT close to 130 ns for an input charge of 6000 electrons and with an I_{LDAC} current of 13 μA, while in the second one the Krummenacher current in the preamplifier feedback network was set to have a ToT of 60 ns in response to a 6000 electrons input charge, with an I_{LDAC} current of 19 μA. In the following discussion and plots, the configuration adopted for the first chip will be called regular ToT mode, whereas the one chosen for the second chip will be referred to as fast ToT mode.

As discussed in Section 4.2, the main analog performance parameters were obtained through charge scans on the pixel matrix, with the front-end threshold set around 1000 electrons before irradiation. As an example of the threshold distributions obtained after irradiation, Fig. 18 shows such distributions for the two front-ends in fast ToT mode as obtained at a TID of 1 Grad(SiO_2). It is worth mentioning that the front-end bias, including the comparator input threshold voltage, was not modified during the irradiation, while the threshold trimming was performed at each irradiation step. From the plot it is clear that the threshold dispersion of the RD53A Linear front-end is significantly larger than the one of the RD53B version. As already mentioned, this is actually due to the lower resolution of the RD53A trimming DAC but, at the same time, also to the saturation effect associated with the RD53A version of the TDAC. Indeed, without such an effect, the ratio between the threshold dispersions for the RD53A matrix and the RD53B one, is expected to be close to 2, while a factor larger than 3 was obtained after irradiation at the maximum TID level.

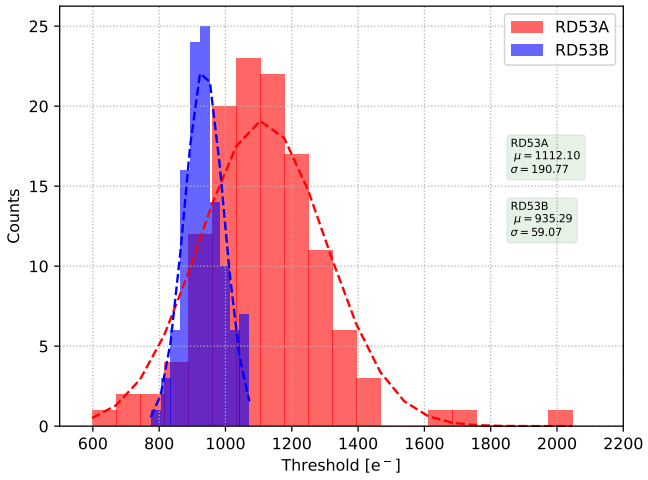

Fig. 18. Measured threshold distributions for the RD53A (red) and RD53B (blue) Linear front-end in fast ToT mode at 1 Grad(SiO_2).

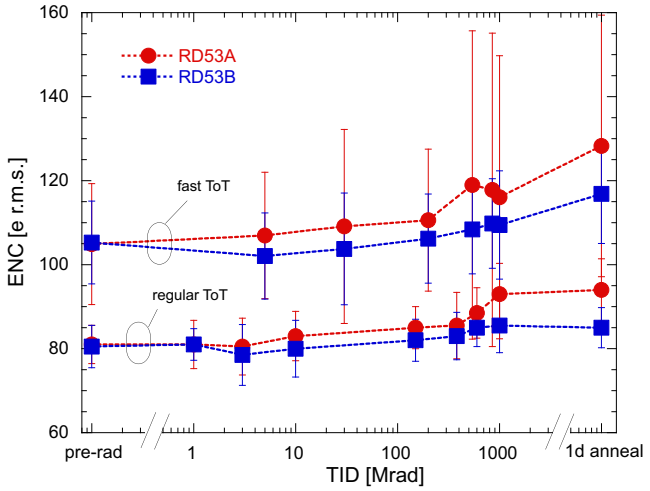

Fig. 19. ENC as a function of the TID for the RD53A (red) and RD53B (blue) front-end in regular and fast ToT mode.

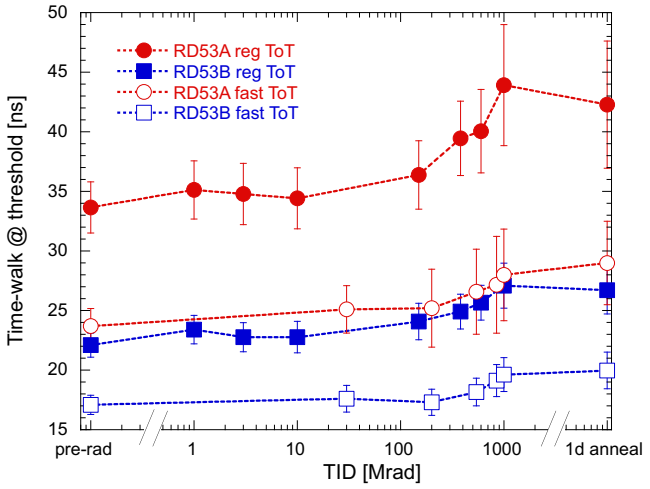

Fig. 21. Time-walk as a function of the TID for RD53A (red) and RD53B (blue) front-ends in regular (solid markers) and fast (blank markers) ToT mode.

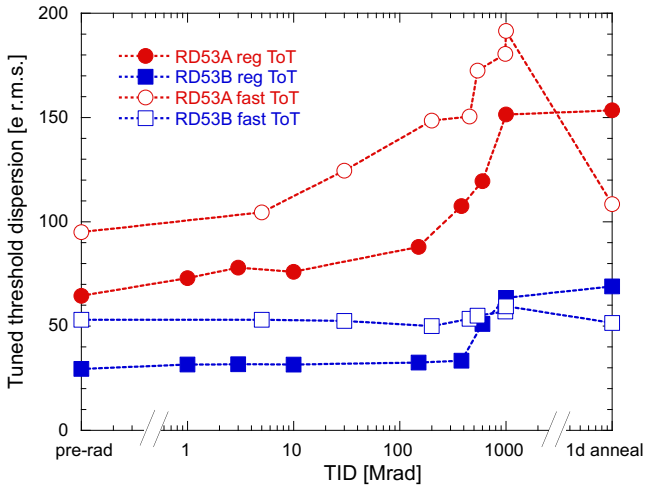

Fig. 20. Threshold dispersion measured after the tuning as a function of the TID for RD53A (red) and RD53B (blue) front-ends in regular (solid markers) and fast (blank markers) ToT mode.

Fig. 19 shows the equivalent noise charge as a function of the TID as measured for the two chips in regular and fast ToT modes. The plot also shows ENC data obtained after 24 hours annealing at room temperature after the last irradiation step. It is worth noticing the larger ENC detected in the chip operated in fast ToT mode: as a matter of fact, with respect to the regular ToT mode, both the preamplifier peaking time and transfer function $T(j\omega)$ get modified and, as shown in (13), the CSA output noise changes consequently. The noise performance is not significantly affected by radiation, with an ENC increase at 1 Grad(SiO_2) around 12% for the RD53A front-end and close to 6% for the RD53B one in both the ToT modes. A larger ENC dispersion has been detected for the chip in fast ToT mode, in particular for the RD53A analog processor at the higher TID levels. After annealing the ENC is larger (in fast ToT mode) than the value measured at the end of irradiation at 1 Grad(SiO_2). However, it has to be noticed that the noise (and the other analog performance parameters) after annealing was measured at room temperature, while the data at the different irradiation steps were obtained at about -10°C: this can, at least partially, explain such an ENC increase.

Fig. 20 shows the tuned threshold dispersion as a function of the TID. In the figure, filled markers are relevant to the chip configured in regular ToT mode, whereas blank markers are associated with the chip in fast ToT mode. A larger threshold dispersion has been detected, also before irradiation, for both the front-ends in fast ToT mode: this is actually expected since in such a configuration the I_{LDAC} current, determining the least significant bit (LSB) of the trimming DAC, is larger than the one set in regular ToT mode. In general, the threshold dispersion of the RD53B Linear front-end is significantly smaller than the one measured for RD53A. In particular, a dispersion close to 30 and 50 e r.m.s has been obtained for RD53B in regular and fast ToT mode, respectively. Such values were found to be pretty stable up to 400 Mrad(SiO_2), with a moderate increase for larger TID levels. Nonetheless, the threshold dispersion of the RD53B channels achieved at 1 Grad(SiO_2) is smaller than 65 e r.m.s. in both the ToT configurations. This data, combined with noise results, makes the RD53B front-end compliant with the noise occupancy requirements of the pixel front-end chips at the High-Luminosity upgrades of CMS and ATLAS experiments, where the quadrature sum of ENC and threshold dispersion has to be lower than around 130 e r.m.s. in order to achieve a noise occupancy smaller than 10^{-6} at 600 electrons threshold [18]. As far as the RD53A version is concerned, a threshold dispersion increase was detected even at the first TID steps, with a tuned threshold dispersion at the maximum radiation level of the order of 2 times the dispersion achieved before irradiation. Interestingly enough, a remarkable recovery after annealing was achieved for the RD53A front-end chip in fast ToT mode. The reason of such a recovery is twofold: on one hand, the annealing at room temperature reduces the un-tuned threshold dispersion, and on the other hand the TDAC saturation effect does not take place at room temperature. The recovery is not present in the case of the chip operated in regular ToT mode, where the measured un-tuned threshold

distribution after annealing was very similar to the one obtained at the end of the irradiation.

Fig. 21 shows the maximum time-walk (i.e. the one measured for an input charge close to the threshold) as a function of the TID for the two front-end versions. Again, filled markers are relevant to the chip configured in regular ToT mode, whereas blank markers are associated with the chip in fast ToT mode. For both the front-end versions the time-walk is significantly smaller in the fast ToT mode. Such a behaviour can be ascribed to a smaller preamplifier peaking time obtained in fast ToT mode as a consequence of a larger Krummenacher current. For the RD53B flavour, the time-walk is smaller than 25 ns over the explored TID range, except for the front-end operated at 1 Grad(SiO_2) in the regular ToT mode, where the time-walk slightly exceeds such a value. On the other hand, the time-walk associated with the RD53A front-end in regular ToT mode is well above 25 ns, with a maximum time-walk, achieved at the largest TID level, close to 45 ns.

In general, it is possible to conclude that the analog performance parameters of the RD53B version of the Linear front-end are very little affected by radiation, in agreement with the excellent radiation tolerance associated with the 65 nm CMOS technology at low temperatures [19].

V. Conclusions

Pixel readout chips for the High-Luminosity upgrades of the CMS and ATLAS experiments are being designed by the CERN RD53 collaboration, which submitted, in 2017, the large-scale demonstrator chip RD53A. The demonstrator embodies three different versions of the analog processor, called Synchronous, Linear and Differential. After a comprehensive test campaign, the Linear front-end has been chosen for the final integration into the CMS pixel chip. The paper discussed two main design improvements proposed for the so-called RD53B version of the Linear front-end, related to its comparator stage and threshold tuning DAC. A small prototype chip including a matrix of 16×16 pixels featuring the RD53A and RD53B versions of the Linear front-end has been submitted and tested, with very promising results obtained for the RD53B flavour also after irradiation at 1 Grad(SiO_2). The Linear front-end has been recently (June 2021) submitted in the RD53B-CMS pixel readout chip, which includes a matrix of 432×336 channels, expected back from the foundry in September for the full characterization.

VI. Acknowledgments

The authors wish to thank Giulio Borghello for his valuable support during the irradiation campaign at CERN and Andrea Galliani for carrying out part of the testing activity on the prototype chip. They also would like to acknowledge the RD53 collaboration members for the support provided throughout this work.

References

[1] G. Sguazzoni, "The CMS pixel detector for the high luminosity LHC", *Nucl. Instrum. and Methods*, vol. 958, 162089, Apr. 2020.

[2] J. Pater et al., "ATLAS Pixel Detector upgrade at High Luminosity LHC", *Proceedings of Science*, PoS (Vertex2019) 011, pp. 1-8, Sept. 2020.

[3] M. Boscardin et al., "Performance of new radiation-tolerant thin planar and 3D columnar n^+ on p silicon pixel sensors up to a maximum fluence of $\sim 5 \times 10^{15}$ n_{eq}/cm^2", *Nucl. Instrum. and Methods*, vol. 953, 163222, Feb. 2020.

[4] M. Meschini et al., "First results on 3D pixel sensors interconnected to the RD53A readout chip after irradiation to 1×10^{15} neq cm^{-2}", *Journal of Instrumentation*, vol. 14 C06018, June 2019.

[5] RD53 web site, http://rd53.web.cern.ch/rd53/

[6] A. Dimitrievska et al., "RD53A: A large-scale prototype chip for the phase II upgrade in the serially powered HL-LHC pixel detectors", *Nucl. Instrum. and Methods*, vol. 958, 162091, Apr. 2020.

[7] L. Gaioni et al., "Test results and prospects for RD53A, a large scale 65 nm CMOS chip for pixel readout at the HL-LHC", *Nucl. Instrum. and Methods*, vol. 936, pp. 282-285, Aug. 2019.

[8] N. Emriskova et al., "Analog front-end characterization of the RD53A chip", *Proceedings of Science*, PoS (TWEPP2019) 021, pp. 1-4, Apr. 2020.

[9] L. Gaioni et al., "RD53 analog front-end processors for the ATLAS and CMS experiments at the High-Luminosity LHC", *Proceedings of Science*, PoS (Vertex2019) 021, pp. 1-7, Sept. 2020.

[10] F. Krummenacher, "Pixel detectors with local intelligence: an IC designer point of view", *Nucl. Instrum. and Methods*, vol. 305 no. 3, pp. 527-532, Aug. 1991.

[11] J. Sonneveld et al. "CMS pixel sensors and modules: lessons learned and the road ahead", presented at the 12th Terascale Detector Workshop, Dresden, 12-15 March 2019.

[12] S. Terzo et al., "Novel 3D Pixel Sensors for the Upgrade of the ATLAS Inner Tracker", *Front. Phys.*, Vol. 9, 624668, Apr. 2021.

[13] M. Manghisoni et al., "Assessment of a Low-Power 65 nm CMOS Technology for Analog Front-End Design", *IEEE Trans. Nucl. Sci.*, vol. 61, no. 1, pp. 553-560, Feb. 2014.

[14] H. Träff, "Novel Approach to High Speed CMOS Current Comparators", *Elect. Lett.*, vol. 28 no. 3, pp. 310-312, Jan. 1992, doi:10.1049/el:19920192.

[15] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. New York: McGraw-Hill, 1999.

[16] L. Gaioni et al., "Threshold tuning DACs for pixel readout chips at the High Luminosity LHC", *Nucl. Instrum. and Methods*, vol. 969, 164025, Jul. 2020.

[17] The RD53A Integrated Circuit. ONLINE: https://cds.cern.ch/record/2287593/files/RD53A_Manual_V3-51.pdf

[18] RD53A Integrated Circuit Specifications. ONLINE: https://cds.cern.ch/record/2113263/files/RD53A_specs_V3_2.pdf.

[19] G. Borghello et al., "Ionizing radiation damage in 65 nm CMOS technology: Influence of geometry, bias and temperature at ultra-high doses", *Microelectronics Reliability*, vol. 116, 114016, Jan. 2021.