Low-Power Amplifier for Readout Interface of Semiconductor Scintillator

Xiao Yun, Student Member, IEEE, Milutin Stanačević, Member, IEEE, and Serge Luryi, Fellow, IEEE

Abstract—We present the design of a readout system, comprising a charge sensitive amplifier and a pulse shaper, that directly interfaces a semiconductor scintillator. The designed amplifier quantifies optical response of a large-area epitaxial photodiode that registers luminescence produced by a scintillating semiconductor wafer when excited by ionizing radiation. The epitaxial photodiode is characterized by a capacitance of 50 pF and a dark current of 10 pA. The presented optimization procedure for the biasing and sizing the input transistor of the CSA directly relates the region of operation of the input transistor with the constraints on power, area and event rate of the readout system. Experimental results of the amplifier implemented in 0.5 µm CMOS technology, verify gain of 71 mV/µC, with the measured linearity of 1.3%. For the parameters of the photodiode, the measured equivalent noise charge (ENC) is 950 electrons with the measured time constant of the pulse shaper of 90 µs and the power consumption of 210 µW. The measured slope of the ENC dependence on the input capacitance is 18 e⁻/pF.

Index Terms—Radiation detection, scintillator, readout IC, charged sensitive amplifier, pulse shaper.

I. INTRODUCTION

The enhancement and deployment of radiological detection capabilities to prevent the illicit use of nuclear devices or materials has been an increasingly important national security issue. Scintillating solid-state radiation detectors [1] register ionizing radiation by converting the energy of the incident radiation into light measured by a coupled photo-detector. Due to inherent non-proportionality response and the resulting poor resolution of the dielectric scintillators [2], [3], the implementation of semiconductor scintillators could potentially bring a significant leap in the energy resolution [4]. The challenge in developing a semiconductor scintillator is how to make the material transmit its own infrared luminescence. Two different approaches have been proposed for the direct-gap semiconductors, like InP and GaAs. In the first, the semiconductor is heavily doped with the donor impurities, in order to introduce the Burstein shift between the emission and the absorption spectra [5]. The second approach is based on the extremely high radiative efficiency of high-quality direct-gap semiconductors. In high-quality InP, a scintillation photon is not completely lost in an act of interband absorption, since a newly created minority carrier generates upon recombination a new scintillation photon in a random direction [4], [6]. Recent progress towards realization of the InP scintillator is described in [7]. An important advantage of the semiconductor scintillator is the ability to integrate an epitaxial photodiode [7] on the scintillator body, ensuring nearly perfect registration of the scintillation photons. Based on the Compton telescope technique [8], a three-dimensional array [4] of semiconductor scintillators (illustrated in Fig. 1) provides an accurate spectroscopic resolution for isotope discrimination and simultaneously an accurate determination of the direction to the source. To measure the optical response of the epitaxial photodiode, an application specific integrated circuit (ASIC) has to be integrated with the detector.

There have been numerous examples in the literature of the readout ASICs [9]–[12] for radiation detection, where both circuit implementation and integration strongly depend on the detector and the application. In the proposed semiconductor scintillator array, the size of the single pixel epitaxial photodiode is 1mm x 1mm. A large area of the photodiode gives rise to a large capacitance, measured at 50 pF [7]. The large diode capacitance leads to a higher noise figure and a potential
instability in the amplifier design. Despite the large diode area, the reverse-bias leakage current is rather low, on the order of 10 pA. The main challenge in the design of the readout ASIC is to minimize the input noise so as to achieve the high sensitivity and prevent false alarms. As the proposed 3D integration calls for a considerable volume with a large number of pixels, strict constraints are imposed on the power consumption and size of the readout ASIC. We envision a 10 mm x 10 mm x 10 mm array, 1 cm$^3$ in volume, with each array element directly interfaced to a thinned single channel readout ASIC. The purpose of the 3D array is not only to enhance the interaction efficiency with incident gamma photons, but most importantly to resolve the energy and the location of the first three interactions, as required for Compton analysis. Careful design of the readout system, especially the charge sensitive amplifier (CSA), is required to optimize noise given the power and area. Previously reported optimization procedures [13]–[15] for sizing the input transistor of the CSA had focused on achieving an unconstrained noise minimum and therefore were primarily interested in accurate noise modeling. Here we shall derive an optimization procedure that directly relates the region of operation of the input transistor to different constraints on the readout system. To do this we shall use tractable models that describe the transistor’s operation and noise. The optimization procedure will be applied to the design of the readout circuit that will be integrated with the described scintillator-photodiode unit into a 3D array for radiation detection.

II. INPUT TRANSISTOR OPTIMIZATION

The readout system, comprising a charge sensitive amplifier and a pulse shaper, is adopted to estimate the theoretical noise floor in the detection of an ionizing radiation event. The optimization of the input MOS transistor of the CSA is derived under the assumption that the total noise of the readout system is dominated by the contribution of the input transistor. The sensitivity of the readout system is expressed through the equivalent noise charge (ENC) [17]. The ENC has three main components: (1) the white and (2) the flicker (1/f) series noise, originated in the input MOS transistor, and (3) the white parallel noise, due to the detector leakage current and the feedback network [13]:

$$\text{ENC}_{ws}^2 = a_{ws} \frac{\gamma^4 kT}{g_m} (C_p + C_g)^2 \frac{1}{\tau}$$  \hspace{1cm} (1)

$$\text{ENC}_{1/f}^2 = a_f A_F (C_p + C_g)^2$$  \hspace{1cm} (2)

$$\text{ENC}_{wp}^2 = a_{wp} 2 q I_{det} \tau,$$  \hspace{1cm} (3)

where $a_{ws}$, $a_f$ and $a_{wp}$ are constants that depend on the type and order of the pulse shaper; $\gamma$ is the thermal noise coefficient that depends on the operation region of the transistor; $g_m$ and $C_g$ are respectively the transconductance and the gate capacitance of the input transistor; $C_p$ is the load capacitance at the CSA input, which is the sum of the detector capacitance $C_d$, the feedback capacitance $C_f$ and possibly other parasitic capacitances resulting from interconnects; $\tau$ is the time constant of the pulse shaper; $I_{det}$ is the detector leakage current and $A_F$ is the coefficient associated with the flicker noise of the input transistor.

As the transconductance and noise parameters of the transistor depend on the region of operation, the optimization procedure for sizing and biasing the input transistor is formulated separately for the strong- and the weak-inversion regions of operation. The transconductance of a transistor biased in the strong-inversion region can be approximated as:

$$g_{m,\text{strong}} = \sqrt{2 \mu C_{ox} I W L},$$  \hspace{1cm} (4)

and in the weak-inversion region as:

$$g_{m,\text{weak}} = \frac{I}{n U_T},$$  \hspace{1cm} (5)

where $n$ is the subthreshold slope coefficient, $U_T = kT/q$, where $k$ is the Boltzmann constant and $T$ is the absolute temperature, $\mu$ is the carrier mobility, $C_{ox}$ is the gate oxide capacitance per unit area, $I$, $W$ and $L$ are respectively the biasing current, the width and the length of the input transistor.

The flicker noise of MOS transistors may have two different origins [14], [16]. The $\delta N$ model assumes that the flicker noise in the drain current results from the fluctuation in the number of charge carriers. On the other hand, the $\delta \mu$ model assumes that the flicker noise is generated by the fluctuation in the mobility of charge carriers. In our optimization procedure it is assumed that the input transistor is PMOS. In the 0.5 $\mu$m CMOS technology, a PMOS transistor has two orders of magnitude lower flicker noise than a NMOS transistor. Studies have shown that PMOS transistors might follow the $\delta N$ model when biased in weak inversion and the $\delta \mu$ model when biased in strong inversion. Therefore, the flicker noise voltage spectral density $S_F$ can be expressed as:

$$S_{F,\delta N} = \left(\frac{q^2 nkT N_T}{2 \beta C_{ox}^2}\right) \frac{1}{WL J} \frac{1}{\sqrt{\mu C_{ox}}},$$  \hspace{1cm} (6)

$$S_{F,\delta \mu} = \left(\frac{nq \alpha H}{2 \mu C_{ox}^2}\right) \left(\frac{1}{\sqrt{\mu C_{ox}}^2}\right) \frac{1}{\sqrt{J}},$$  \hspace{1cm} (7)

where $\alpha_H$ is the Hooge constant, $N_T$ is the oxide/interface trap density at the quasi Fermi level per unit volume, and $\beta$ is the tunneling parameter of the traps. The $\delta \mu$ model predicts that the noise increases with the square root of the biasing current $I$, while the $\delta N$ model predicts that the noise is independent of $I$.

A. Input transistor biased in strong inversion

The derivation of the optimal sizing of the input transistor that operates in the strong-inversion region is presented in [14]. The total ENC is expressed in terms of three design variables: the biasing current $I$, the gate capacitance of the input transistor $C_g$ and the shaping time constant $\tau$.

$$\text{ENC}_{\text{tot, str}}^2 = A_1 \frac{(C_p + C_g)^2}{\sqrt{C_g}} \frac{1}{\sqrt{J}} \tau + A_2 \frac{(C_p + C_g)^2}{\sqrt{C_g}} \sqrt{J} + A_3 \tau,$$  \hspace{1cm} (8)
The optimal shaping time constant is

\[ \tau_{\text{opt}} = \frac{2}{B_3 \chi C_p} \]

1) Unconstrained optimization: In order to have the input transistor biased in the weak-inversion regime, the following condition has to be satisfied [14]:

\[ I < 0.01 I_S, \]

where

\[ I_S = 2 \pi \mu C_{ox} U_T^2 \frac{W}{L}. \]

This leads to an upper limit on the biasing current

\[ I < C_g n \mu U_T^2 \]

Therefore, the optimum \( I \) is chosen as the largest current that keeps the input transistor in the weak-inversion regime:

\[ I_{\text{opt}} = \chi C_g, \]

where

\[ \chi = \frac{n \mu U_T^2}{50 L^2}. \]

The optimum \( C_g \) and \( \tau \) can be found by equating partial derivatives of (12) with zero, which gives the unconstrained weak-inversion optima in the form:

\[ C_{g,\text{opt}} = C_p \]

\[ \tau_{\text{opt}} = 2 \frac{B_1}{B_3 \chi C_p}. \]

When the above optima are feasible, the absolute minimum ENC in the weak-inversion regime is:

\[ \min[ENC_{\text{tot,weak}}^2] = 4 \sqrt{\frac{B_1 B_3}{\chi C_p + 4 B_2 C_p}}. \]
The optimum fixed biasing current, with \( \tau \) shows the dependence of \( ENC \) on the transistor in the weak-inversion regime, or when the input transistor size is sufficiently large to match the gate capacitance of the input transistor cannot be made smaller. Therefore, the power consumption can be reduced by more than two orders of magnitude, while the penalty in the ENC is an increase by a factor of four.

2) Constrained optimization: Here we consider the situation when one of the three design variables is constrained. When the biasing current \( I \) is fixed, the suboptimum \( \tau \) can be calculated as

\[
\tau_{\text{subopt}} = \sqrt{\frac{B_1}{B_3 I}} (C_p + C_g). \tag{20}
\]

After inserting (20) into (12), the suboptimum ENC becomes a function of \( C_g \) as

\[
ENC_{\text{subopt}}^2 = 2\sqrt{\frac{B_1 B_3}{I}} (C_p + C_g) + B_2 \frac{(C_p + C_g)^2}{C_g}. \tag{21}
\]

The first term in (21) is minimized by the smallest value of \( C_g \) that still keeps the transistor biased in weak inversion. The second term in (21) is minimized by \( C_g = C_p \). Figure 2 shows the dependence of \( ENC_{\text{subopt}}^2 \) as a function of the fixed biasing current, with \( \tau = \tau_{\text{subopt}} \) and the value of \( C_g \) selected to minimize (21).

If \( C_g \) is limited due to the area constraint, the optimum biasing current has to be chosen as the largest value to keep the transistor in the weak-inversion regime, or \( I_{\text{subopt}} = \chi C_g \). The optimum \( \tau \) can still be calculated as

\[
\tau_{\text{subopt}} = \sqrt{\frac{B_1}{B_3 \chi}} \frac{C_p + C_g}{\sqrt{C_g}}, \tag{22}
\]

which gives the suboptimum ENC

\[
ENC_{\text{subopt}}^2 = 2\sqrt{\frac{B_1 B_3}{\chi \sqrt{C_g}}} \frac{C_p + C_g}{\sqrt{C_g}} + B_2 \frac{(C_p + C_g)^2}{C_g}. \tag{23}
\]

C. Comparison of weak inversion and strong inversion

When the absolute minimum noise is desired, the transistor must be in the strong-inversion regime. However, in circumstances where the biasing current of the input transistor is limited due to the power constraints, the optimal region of the transistor operation depends on the power constraint. The optimum point also varies with the different sensor parameters and the noise requirement.

In the design of the proposed readout system we have relatively flexible choices over the time constant and the input transistor size. However, the power consumption is constrained. For this case, with fixed biasing current and no constraints on \( \tau \) and \( C_g \), we compute the minimum achievable ENC for the cases of the input transistor biased in the strong-inversion regime and the weak-inversion regime, see Fig. 2. The biasing currents are limited to 100 \( \mu \)A, which is reasonable for most readout systems.

In the strong-inversion regime, the ENC monotonically decreases as the current increases. For a small biasing current, the gate capacitance of the input transistor cannot be made sufficiently large to match \( C_p \). Therefore, the minimum ENC is achieved when \( C_g \) is the largest value that keeps the input transistor in strong inversion. In the weak-inversion regime, a minimum value of ENC is achieved at \( C_g = C_p \) and the

For our specific sensor, the unconstrained optimization leads to a minimum ENC of 348 e\(^-\) at the biasing current of 44 \( \mu \)A. The input transistor width is 33 mm and the shaping time constant is 124 \( \mu \)s. Therefore, the power consumption can be reduced by more than two orders of magnitude, while the penalty in the ENC is an increase by a factor of four.

If the shaping time constant is fixed, due to the rate of ionizing radiation events, the same optimization procedure results in \( I_{\text{subopt}} = \chi C_p \). We can notice that the minimum value of ENC worsens by 10% when the input transistor size is at 30% of its optimum value.

If \( I_{\text{subopt}} \) is scaled by the optimal time constant when the input transistor is in the weak-inversion regime.
biasing current equals the largest value that keeps the input transistor in weak inversion. From Fig. 2, we can conclude that when the biasing current is limited to 100 µA, the weak-inversion regime is preferred over the strong-inversion regime; when the biasing current is higher, the strong-inversion regime becomes the preferred region of operation. The moderate-inversion region of operation of the input transistor is omitted in the comparison due to the lack of tractable noise models.

III. AMPLIFIER IMPLEMENTATION

The detailed circuit implementation of the amplifier, comprising a CSA and a pulse shaper, is outlined in this Section. The optimization technique proposed in Section II is applied for sizing the input transistor.

A. Charge Sensitive Amplifier

Conventional architecture of the CSA with a high-gain amplifier and a small feedback capacitor for the input charge integration is implemented as the first stage of amplification. In order to increase the overall gain provided by the input amplification stage, a two-stage cascaded charge-sensitive amplifier is implemented, as shown in Fig. 5. To ensure high linearity of the first-stage of the CSA, a pole-zero compensation network is formed by a parallel connection of $N_1$ replicas of the feedback network formed by capacitor $C_f$ and transistor $M_f$. The second CSA stage increases the overall gain by a factor of $N_2$, leading to the total gain of $N_1N_2=100$ in the proposed implementation. The value of the DC voltage at the input of both high-gain amplifiers is designed to be the same, so as to provide equal biasing conditions for transistors $M_{f1}$ and $N_1M_{f1}$, which is necessary for the high linearity of the CSA. The choice of feedback capacitors is influenced by the high detector capacitance; the feedback capacitors are set at $C_f = C_{f2} = 500$ fF.

Figure 6 shows the folded-cascode implementation of the high-gain amplifier in the first stage of CSA. The PMOS input transistor is chosen over NMOS due to its lower flicker noise. The width of the input transistor is $16.2$ mm and the gate length is $1.2$ µm. We found by simulation that the DC gain of the folded-cascode amplifier is 94 dB. The feedback transistor $M_{f1}$ operates in the subthreshold regime due to the low (10 pA) leakage current of the detector.

1) Noise contributions from other transistors: The additional noise contributions from $M_2$, $M_3$ and $M_5$ can be approximated as:

$$V_{n,add}^2 = \left( V_{n2}^2g_{m2} + V_{n5}^2g_{m5} + \frac{V_{n3}^2}{r_s} \right) / g_m^2,$$

where $V_{n2}$, $V_{n5}$ and $V_{n3}$ are the input-referred noise sources at the gates of $M_2$, $M_5$ and $M_3$, respectively, and $r_s$ is the dynamic resistance seen at the source node of $M_3$. The input-referred voltage noise from $M_2$ and $M_5$ is scaled by the transconductance of $M_1$. Because $M_2$ and $M_1$ have similar biasing current, and NMOS transistors have at least one order of magnitude higher flicker noise coefficient than PMOS, the noise contribution from $M_2$ cannot be neglected. To reduce the thermal noise from $M_2$, $g_{m2}$ has to be small, which leads to a higher voltage drop for the fixed biasing current. To reduce the flicker noise from $M_2$, the length of the transistor is increased [14]. Noise from $M_3$ can be neglected due to the large resistance $r_s$. The optimization procedure presented in Section II is adjusted to take into account the additional noise sources from (25).

2) Stability analysis: The large input transistor makes the non-dominant pole located at the folding node close to the dominant pole located at the output of the folded-cascode amplifier; thus, the stability of the CSA requires careful examination [18]. Assuming that the source follower is an ideal buffer and that the reset MOS transistor $M_{f1}$ has a very large equivalent resistance $R_f$, the open-loop circuit can be modeled as shown in Fig. 7, where $C_x = C_{gd1} + C_{db1} + C_{gd2} + C_{db2} + C_{gs3} + C_{sb3}$ is the total capacitance at the folding node; $r_o$ is the resistance of cascode of $M_4$ and $M_5$; $C_L$ is the capacitance at the output of the folded cascode amplifier and $C_o$ is the capacitance at the CSA output. Thus, the loop gain can be derived as:

$$A(s)\beta = \frac{V_{out}(s)}{I_{in}(s)} = \frac{g_m r_o (\frac{s}{\alpha_1} + 1)}{\left(\frac{s}{\beta_1} + 1\right)\left(\frac{s}{\beta_2} + 1\right)\left(\frac{s}{\beta_3} + 1\right)},$$

where $\beta = 100$.
where zero and poles are located at
\[ z_1 = \frac{1}{R_f C_f} \]
\[ p_1 = \frac{1}{R_f (C_f + C_p + C_g)} \]
\[ p_2 = \frac{1}{r_o C_L} \]
\[ p_3 = \frac{g_{m3} C_L}{C_x} \]  \hspace{1cm} (27)

We attribute pole \( p_1 \) to the feedback capacitor and the input capacitive load, while \( p_2 \) and \( p_3 \) are the two poles associated with the voltage amplifier. The feed-forward zero comes from the \( RC \) feedback network. The unity-gain frequency \( f_u \) can be approximated as:
\[ f_u = \frac{g_{m1} C_f}{C_L (C_f + C_p + C_g)} \]  \hspace{1cm} (28)

Since the first pole and zero are at relatively low frequency, their phase contribution can be approximately canceled. Therefore, the first non-dominant pole \( p_3 \) is critical and ultimately determines the phase margin of the loop. It is reasonable to assume that the CSA’s non-dominant pole is fixed, thus it is desirable to place \( f_u \) far away from \( p_3 \). A smaller \( C_f \) makes the CSA more stable; however a large amplifier gain is required due to a high \( C_p/C_f \) ratio. Thus, adding a compensation capacitive load to \( C_L \) is the only feasible option.

In our design, the value of \( C_L \) is 2 pF and the achieved phase margin is over 70°. Though the added compensation capacitor slows the circuit’s step response to over 3 \( \mu \)s, a time constant of 100 \( \mu \)s of the following shaper means that the overall readout system is not affected.

Since the input capacitance of the second high-gain amplifier in Fig. 5 is significantly smaller than that for the first amplifier, the second amplifier is designed as a five-times scaled-down version of the first amplifier. The additional compensation capacitor is not required for the second amplifier.

B. Pulse Shaper

The calculated optimum shaping time constant for the proposed amplifier is fairly large, 100 \( \mu \)s, as a consequence of the high capacitance and the low leakage current of the detector. This time constant requires that there be no more than 30,000 photons/sec incident on a pixel. In order to achieve such a long shaping time with the constraints of area and power, we have chosen a filter based on ICON RC cell [19] that provides both the low-area and the low-power shaper implementation. The schematic of the two stage pulse shaper is shown in Fig. 8(a). The ICON cell, shown in Fig. 8(b), makes the equivalent resistance \( N \) times higher than the integrated physical resistance due to the ratio of \( N \) in the current mirrors. The amplifiers in the shaper are a scaled down version of the high-gain amplifier in the CSA, thus ensuring the same amplifier DC level.

IV. RESULTS

The proposed readout system was implemented in 0.5 \( \mu \)m CMOS technology with the measured flicker-noise parameter of the PMOS transistor equal to 7x10^\(-26 \) V^2 F. Figure 9 shows the micrograph of the single channel of the chip highlighting the functional blocks described in Sects. III-A and III-B. The size of a single channel is 1.1 mm \( \times \) 0.4 mm. The characterization of the implemented preamplifier was conducted at 3.3 V supply voltage.

The fabricated preamplifier is characterized without a sensor. The chip-on-board technique was employed to wire-bond the die directly on the printed circuit board. A 1 pF capacitor that is connected externally to the input of the CSA enables...
a controlled charge injection into the readout circuitry, as the current pulse at the input is generated by applying a known voltage step signal to the capacitor. A dc current is injected into the input node of the CSA through a 1 GΩ resistor and is controlled through an on-board DAC. To model sensors capacitance in the characterization of the noise performance of the preamplifier, different external capacitors were used at the input node.

Figure 10 shows the response of the CSA for different voltage steps that correspond to the total input injected negative charge of either 5000 or 10,000 or 20,000 electrons. Due to the small leakage current of the detector and biasing of the feedback transistor in the subthreshold regime, the decay time of the CSA is on the order of milliseconds. Figure 11 shows the measured output signal after the pulse shaper. The shaping time constant is 90 µs. The measured charge gain is 71 mV/fC. The measured linearity of the CSA is 0.4%, while the measured linearity of the amplifier, consisting of the CSA and the pulse shaper, is 1.3%. A plot of the output voltage of the amplifier as function of the input charge is shown in Fig. 12.

The measure of the chip sensitivity is ENC. The measured r.m.s. ENC (for the detector capacitance fixed at 50 pF and the leakage current fixed at 10 pA) is 950 electrons, a higher value than predicted. The minimum detectable signal of 3000 electrons corresponds to gamma energy resolution of 15 keV. Figure 13 shows the measured ENC for different values of the input capacitance and the measured slope is 18 e−/pF. The measured power consumption of a single channel is 210 µW. The performance of the implemented amplifier is matched to the required minimum detectable signal and the input range of the proposed radiation detector.

V. CONCLUSION

We presented the design and the implementation of a low-noise, low-power amplifier to register the semiconductor scintillator signal excited by ionizing radiation. To increase the sensitivity of the readout circuitry under the area, power and event-rate constraints, we have devised a novel optimization technique and applied it to the design of the CSA. The integration of the proposed readout amplifier and semiconductor scintillator in a 3D array will provide both isotope
discrimination and angular resolution in various homeland security applications.

VI. ACKNOWLEDGMENTS

The authors are grateful to Gianluigi De Geronimo (Brookhaven National Lab) for useful discussions and review of the paper.

REFERENCES


