Sub-nanosecond Gating of Large CMOS Imagers

Octavian Maciu, Wilfried Uhring, Jean-Pierre Le Normand, Jean-Baptiste Kammerer, Foudil Dadouche, Norbert Dumas
ICube, UMR 7357, University of Strasbourg and CNRS
23, rue du Loess BP 20, F-67037 Strasbourg Cedex 2, France
Tel.: +33 (0) 3 88 10 68 27, fax: +33 (0) 3 88 10 65 48
E-mail: wilfried.uhring@unistra.fr

Received: 31 August 2015   /Accepted: 5 October 2015   /Published: 30 October 2015

Abstract: Ultra-fast gating of large array imagers can be quite challenging to implement due to the distributed RC (Resistance Capacitance) nature of the metal wires used in all ICs (Integrated Circuits) for electrical connections. For the transmission of a signal across a long path, the metal line reduces the electrical bandwidth and adds a delay. The behavior of these lines has been modeled and a new solution is presented to circumvent these limitations. In this paper, we present an edge-based approach to the gating circuitry that allows sub nanosecond gating with a very low skew across the whole imager. Simulation data shows that our solution is an efficient way of reducing the effect of the distributed RC line delay with a small penalty on surface area and consumption. Copyright © 2015 IFSA Publishing, S. L.

Keywords: SPAD, Ultra-fast gating, Edge-triggered, Fast pulse, Skewless.

1. Introduction

Over the last decade, ultra-fast imaging has been a booming field with many considerable breakthroughs. A key advancement to this technology has been the ability to design image sensors with a sub-nanosecond temporal resolution. These imagers could be configured in either single-shot mode in [1] and [2], or can be repetitive with photodiode or photogate in [3] and [4] or based on Single-Photon Avalanche Diode (SPAD) arrays rather than classical photodiodes in [5] and [6]. In this case, the temporal resolution can be even lower. Currently, integrated streak cameras operate at the fastest frequency, therefore also requiring robust acquisition signals. These fast signals are used in order to have a sub-nanosecond shutter speed. Depending on the design, the temporal resolution of an integrated streak camera can vary from a few hundred picoseconds to several nanoseconds. A 1D or 2D approach of integrated streak camera solution can be found in [3], where a delay generator based on the propagation delay of logic gates is used for sub-nanosecond shuttering. The delay could be customized using current starved inverters. In [1], [4] and [7], we see the use of edge-based control signals for fast gating.

Similar to a clock signal in a synchronous sequential circuit, the gating signals are distributed to the entire imager array. Due to the distributed RC (Resistance capacitance) line delay, the integrity of the signal is compromised along the row of pixels. This would degrade the performance as a pixel at the end of the array would not perceive the same signal as another near the beginning of the array. Moreover, it would be completely unusable for ultra-fast gating purposes as the rise time and fall time would be far greater than the gating time. This is especially true for ultra-fast image sensors as a signal commutes at the
order of 100 picoseconds. It behooves us then to ensure that the signal is identical and usable for every single pixel in the array. Therefore, a more reliable approach would be to reduce the dependency on the pulse width and have an edge triggered gating. This solution was implemented in a 0.18 \( \mu \)m CMOS (Complementary Metal Oxide Semiconductor) process. These approaches have been applied for smaller dimension image sensors and it is therefore interesting to propose a solution that is extendable to full resolution imagers.

2. Theoretical Approach

As mentioned before, in an imager array, the gating signal is being driven at the beginning of a row. Therefore, the signal perceived at the end of the row will no longer be a clean square pulse. Hence, it is interesting to study what the maximum row length for which the pulse will still be usable. Firstly, we can model this problem by looking at a single distributed RC model. Fig. 1 shows the equivalent distributed line representation of a row containing \( N \) pixels where \( R_D \) and \( C_D \) are the resistance and the capacitance per length unit of the metal line respectively. The localized pixel input capacitances \( C_{pixel} \) act like a distributed capacitance according to the pixel pitch \( C_{pixel} \).

\[
C_{D_{pixel}} = \frac{C_{pixel}}{Pixel\ Pitch} \tag{1}
\]

The total distributed capacitance \( C_{D_{tot}} \) is thus given by the sum of the equivalent distributed input pixel capacitance and the line distributed capacitance, \( C_{D_{tot}} = C_D + C_{D_{pixel}} \).

![Fig. 1. Distributed RC line representation of a sensor row including N Pixels.](image)

The open-circuited Laplace transfer function from the beginning of the line \( V_1 \) to the end of the line \( V_2 \) can be written as [8, 9]:

\[
H(s) = \frac{V_2}{V_1} = \frac{1}{\cosh\sqrt{s \cdot R_D \cdot C_{D_{tot}} \cdot I}}. \tag{2}
\]

where \( s \) is the Laplace variable, and \( I \) is the length of the row. The behavior of the line can be approximated by the simplified circuit model depicted in Fig. 2 [8].

![Fig. 2. Simplified circuit of entire line.](image)

Where \( R = R_D \cdot I \) and \( C = C_{D_{tot}} \cdot I \) are the total resistance and capacitance of the line. The unit step response gives us a clear indication that while not perfect, this approximation is sufficient to illustrate the distributed RC delay problem [8]. The distributed RC effect across the line therefore affects the driver rise time. The fastest rise time of \( V_2 \) can thus be obtained by the following expression:

\[
T_r = 0.35\pi \cdot R_D \left(C_D + C_{D_{pixel}}\right) \cdot I^2 \tag{3}
\]

Equation (3) states that the rising time increases with the square of the row length \( I \) and then can dramatically reach a value that makes it impossible to transport a nanosecond pulse across a large array sensor.

2.1. Parameters Extraction

The parameters \( R_D \) and \( C_D \) can generally be found in the design kit documentation. Otherwise, the unit line and surface capacitance can be obtained by an analog extracted view simulation. Based on variants (square and rectangular) of the diagram in Fig. 3, these parameters can be deduced with two different sets of equations (4).

\[
C_{line} = 2 \cdot C_{edge} \cdot (l + w) + C_{surface} \cdot w \cdot l, \tag{4}
\]

where \( w \) and \( l \) are the respectively the width and the length of the line, \( C_{line} \) is the extracted line capacitance, \( C_{edge} \) is the edge capacitance, given in
F/m, and $C_{\text{surface}}$ the surface capacitance of the line, given in F/m².

Then, the distributed capacitance for a long line can be approximated by:

$$C_D = 2 \cdot C_{\text{edge}} + C_{\text{surface}} \cdot w$$  \hspace{1cm} (5)

where $w$ is the width of the considered line. This capacitance did not take account of the parasitic capacitance added by the neighboring wires. This last cannot be easily determined \textit{a priori}. A first rule of thumbs should be to multiply about 2 or 3 times this capacitance. A post extracted simulation, see section 6, is mandatory to obtain an accurate value.

The resistivity can be extracted in the same way or it can be obtained by process data, such as the metal resistivity $\rho$ and thickness $T$. Therefore, the resistance per length unit $R_D$ of a line of width $w$ is given by:

$$R_D = \frac{\rho}{w \cdot T}$$  \hspace{1cm} (6)

3. Ultra-Fast Gating for Large Array Sensor

Fig. 4 shows the shape of a 1 ns pulse propagating in a sensor row at the beginning of the line, where the pulse is applied, in the middle and at the end of the line. If we assume the classic case where the logic of the pixel reshapes the pulse with a threshold voltage of half the power supply (solid line in Fig. 4), we clearly see that after a certain length, the conventional gating is no longer adequate.

Furthermore, it is interesting to create a model in order to anticipate this signal distortion with respect to the length of the sensor. Based on the distributed model and a voltage threshold of half the power supply, Fig. 5 compares the Full-Width at Half Maximum (FWHM) ratio of the in-pixel reshaped pulse versus the original pulse according to the original pulse FWHM to the rise time. The effective FWHM inside the pixel decreases as soon as the original FWHM pulse is below the rising time and is reduced to zero for a ratio of 30%.

Using (3), the relationship between the increasing rise time and the length of the array is shown in Fig. 6 for a metal line, level 2, of a typical 0.18 µm CMOS process with a resistivity $\rho$ of $2.65 \times 10^{-8}$ Ω·m, a thickness $T$ of 425 nm, a pixel pitch of 35 µm, an input pixel capacitance of $C_{\text{pixel}}$ of 2 fF, a surface capacitance $C_{\text{surface}}$ of 0.015 fF/µm², an edge capacitance $C_{\text{edge}}$ of 0.032 fF/µm and a width $w$ of 3 µm. Sub-nanosecond rise times are not allowed for sensor dimensions above 16 mm.

A way to enhance the line bandwidth is to increase the line width $w$ because it reduces the resistance of the line. Fig. 7 shows the calculated rise time in the middle of the previous line according to the metal width $w$. Increasing the width of the metal track above 3 µm has minimal impact on rise time and thus becomes irrelevant.
Fig. 6. Rise Time of a signal seen across a pixel row through a metal line of up to 20 mm and a pixel pitch of 35 µm with the following parameters: $\rho = 2.65 \times 10^{-8} \, \Omega \cdot \text{m}$, $T = 425$ nm, $C_{\text{pixel}} = 2 \, \text{fF}$, $C_{\text{surface}} = 0.015 \, \text{fF/} \mu \text{m}^2$, $C_{\text{edge}} = 0.032 \, \text{fF/} \mu \text{m}$ and $w = 3 \, \mu \text{m}$.

Therefore, generating 200 ps FWHM gating within large sensor arrays over 10 mm is impossible with pulse propagation techniques. To ensure signal integrity, the use of edge-sensitive logic is mandatory.

4. Proposed Design of Edge-Triggered Gating

4.1. Edge-based Circuit for Ultra-fast Gating

In SPAD array image sensors, for ultra-fast gating mode there are three critical signals needed for optimal operation: Quench, Reset and Gate. Moreover, SPADs could be gated immediately after the reset to avoid the detection of photons arriving before the investigation time slot. While Quench is a much slower signal, the Reset and Gate signals can be sub-nanosecond pulses.

Having shown clear limitations for ultra-fast gating in large SPAD array image sensors, a different approach is needed. This approach consists of an additional signal $\text{GATE\_DELAY}$ (cf. Fig. 8). $\text{GATE\_DELAY}$ is delayed thus creating the effective pulse of width $d$ seen above through a logic gate. It can therefore be extended to the Reset signal of the SPAD as seen in Fig. 9. Hence, we are able to achieve gating in the range of 100 ps up to 1 ns while maintaining signal integrity throughout the SPAD array. Quenching time is much longer than the reset and gating signals and do not require an ultrafast signal generation.

Fig. 7. Rise Time with respect to metal track width at midway along a row of pixels.

Fig. 8. Timing diagram of typical SPAD quench, reset and gating with delayed ($d$) edge-based signals below.

Fig. 9. SPAD Edge-triggered circuit with active quench, reset and gating.

4.2. Edge-based Driver for Ultra-fast Gating

In Fig. 10, it can be seen that the GATE signal is delayed to obtain a GATE\_DELAY signal. In order to generate the delay, we used a delay generator to drive the signals. The delay generator is constructed using a series of current starved double inverters as seen in [10].

The delay of both signals (cf. Fig. 10) can be determined by a modifiable control voltage ($V_{\text{ctrl}}$). We are able to modify the pulse width by splitting the signal in two branches with one providing minimal propagation delay and the other being delayed with respect to the other. Both signals are then fed into an in-pixel logic gate to obtain the desired pulse derived from the delay between the two signals.
4.3. Simulation of Edge-based Technique

The presented solution can be demonstrated and validated through a Cadence simulation with two buffered lines of 20 mm long and a width of 3 µm based on the model presented in this paper in order to obtain a 300 ps pulse through a logic gate.

In Fig. 11, the FWHMs of the two pulses at the beginning (solid line) and at the end (dashed line) of a 20 mm length line are both equal to 300 ps. Our solution performs well in a Cadence simulation using the model presented in Fig. 3 and the defining equations shown previously. Consequently, thanks to the edge pulse technique generation, we can ensure that all the pixels of the sensor exhibit the same temporal gate width. However, a non-negligible skew is still present.

Fig. 12 shows the simulation of needed $V_{TH}$ variation across a 20 mm image sensor row using above mentioned parameters. The threshold voltage at the beginning of the sensor is arbitrary set at 1.1 V (above $V_{DD}/2$ where $V_{DD} = 1.8$ V) in order to maintain threshold levels across the array well above the $|V_T|$ of each transistor. For a required voltage $V_{TH}$, the inverter should be adjusted by modifying the geometry of the PMOS and NMOS transistors, thus changing the beta ratio.

5. Analysis and Elimination of Skew

For large array ultra-fast image sensors, ensuring signal integrity across the array is only half of the challenge. As can be seen in Fig. 6, a 20 mm sensor will introduce a skew of over 1.5 ns. Hence, the edge based technique has to be improved for large dimensions if nanosecond or sub-nanosecond gating are targeted.

In order to present a more robust approach on ultra-fast gating, we proceed to eliminate the large skew present across the array mentioned. This can be done by introducing two skewed inverters on each branch (GATE and GATE_DELAY) before the logic gate. The threshold voltage of each inverter can be modified by altering the beta ratio effects inside each inverter. By compensating the voltage drop at a given time along the sensor lines thanks to the inverted threshold voltage, the skew across the entire sensor array can be compensated.

This can be obtained by solving the following equation:

$$I_s = I_p, \begin{cases} V_{out} = V_{in} \\ V_{out} = V_{DD}/2 \end{cases}$$

(7)

In this case, both the NMOS and the PMOS transistor are in the saturated operation region, thus the currents are given by:

$$I_s = \frac{1}{2} \mu C_{ox} \left(\frac{W}{L_s}\right) (V_{TH} - V_{IN} - \frac{V_{DD}}{2}) \left(1 + \lambda \frac{V_{DD}}{2} \right)$$

$$I_p = \frac{1}{2} \mu C_{ox} \left(\frac{W}{L_p}\right) (V_{TH} - V_{IN} - V_{DD} - \frac{V_{DD}}{2}) \left(1 + \lambda \frac{V_{DD}}{2} \right)$$

(8)
Replacing $I_n$ and $I_p$ in equation (7) leads to:

$$\mu_n \left( \frac{W}{L_n} \right) (V_{TH} - V_f) = \mu_p \left( \frac{W}{L_p} \right) (V_{DD} - V_{TH} - |V_f|)$$

Hence, we obtain the ratio $\rho$:

$$\rho = \frac{\left( \frac{W}{L_n} \right)}{\left( \frac{W}{L_p} \right)} = \frac{\mu_n}{\mu_p} \left( \frac{V_{TH} - V_f}{V_{DD} - V_{TH} - |V_f|} \right)^2$$

Furthermore, in order to keep the best fill factor, the inverter area $A$ given by:

$$A = W_p L_p + L_n W_n = \rho \left( \frac{W}{L_n} \right) L_n^2 + \left( \frac{L}{W_n} \right) W_n^2$$

has to be kept as low as possible. For a threshold voltage $V_{TH}$ above $V_{DD}/2$, the PMOS transistor has to be more conductive than the NMOS one, thus the ratio $\rho$ should be greater than one. For this, we can consider using $W_n = W_{min}$ and $L_p = L_{min}$ as the best choice. The inverter area becomes:

$$A = \rho \left( \frac{W_{min}}{L_n} \right) L_{min}^2 + \left( \frac{L_{min}}{W_{min}} \right) W_{min}^2$$

Minimizing $A$ according to the variable $\alpha$ leads to best sizes of the MOS transistors for the smallest use of surface area:

$$\frac{W_n}{L_n} = \frac{W_{min}}{\sqrt{\rho \cdot L_{min}}} \quad \text{and} \quad \frac{W_p}{L_p} = \frac{W_{min}}{\sqrt{\rho \cdot L_{min}}}$$

In a similar way, the optimal sizes of the transistors for a threshold voltage $V_{TH}$ under $V_{DD}/2$ can be computed as:

$$\frac{W_n}{L_n} = \frac{W_{min}}{\sqrt{\rho \cdot L_{min}}} \quad \text{and} \quad \frac{W_p}{L_p} = \frac{W_{min}}{\sqrt{\rho \cdot L_{min}}}$$

In both case, the area of each threshold adjusted inverter is given by:

$$A = W_{min} L_{min} \sqrt{4\rho}$$

The more ratio $\rho$ is distant from 1, the higher the surface area.

It would be ideal to designate an area of a pixel for the placement of automatically generated skewed inverters based on their position in the sensor row.

By employing this technique, it can be seen in Fig. 13 that the pulse at the beginning of the sensor row and the pulse at the end are synchronized and the skew seen in Fig. 11 is eliminated. The adjusted threshold voltage are indicated with dotted and dashed line. This ensure that a sub-nanosecond gated operation can be proceeded uniformly across the whole sensor.

**5.1. Variable Slew Rate Driver**

To be able to apply the previously described method to eliminate skew across the sensor, a custom driver is used. This driver should have a variable slew rate to further compensate skew and ensure that the skew eliminating current starved inverters are effective. This is due to the fact that when operating at sub-nanosecond time intervals, the skew elimination is only possible for a relatively slow driving signal as can be seen in Fig. 13. Otherwise, it would not be possible to compensate the threshold voltage across the sensor array as the signal’s rise time would be too fast. Furthermore, it is desirable to be able to control the slew rate of the signal being driven across the sensor array. This can be done using a voltage controlled current starved inverter element as seen in Fig. 14.
As only the rise time of the edge signals is used to generate the fast pulse generation, the current starved inverter is degenerated only on the PMOS branch. A symmetrical slow driver would create significant limitations in operating frequency due to the slow fall time. Hence it is best to design an asymmetrical driver with a slow rise time and fast fall time. This can be done by altering the geometry of the NMOS transistor in Fig. 14. As can be seen in Fig. 15, the rise time is just over 9 ns and the fall time is three times faster at around 3 ns for a maximum period of 20 ns (50 MHz).

6. Application to an Ultra-fast Pixel

This new concept has been applied to a new chip designed to operate with a sub-nanosecond gating taking into account the aforementioned problems and solutions that arise during the design process. This new chip has a total row length of approximately 13 mm split into 10 sub-row. Each sub-row consists of 33 pixels (1.3 mm).

In Fig. 16, we have shown the layout for our SPAD gating system and pixel. For this application, there are three critical signals, therefore six lines needed for edge-based pulse propagation (seen in cyan and red). It can be seen that while they have a considerable impact on the pixel fill factor, it remains reasonable for pixel pitch of 35 μm.

6.1. Post Extracted Parameters (Additive Parasitic Capacitances)

For a more accurate picture of total parasitic capacitance across a row of pixels, a post extracted simulation is needed. This is due to the fact that there are more $C_{edge}$ parasitic capacitances in parallel than originally accounted for in our theoretical model. In the design presented in Fig. 16, these extra parasitic elements triple the total parasitic capacitance for a line. Therefore, the impact of neighbouring parasitic elements inside the pixel are also very significant. Thus, the adjustment of the threshold voltage of the skew inverter has to be done after a post extracted simulation. Even if the pixel capacitance remains low compared to the line capacitance, altering the inverter’s transistor geometry will have a little impact on the distributed pixel capacitance and consequently, the post extracted simulation has to be reiterate. A good rule of thumbs for the first iteration is to considered that the distributed capacitance is about 3 times the single line capacitance described in section 2.

![Figure 15](image1.png)

**Fig. 15.** Variable slew rate driver response with a period of 20 ns.

![Figure 16](image2.png)

**Fig. 16.** Layout of SPAD pixel using ultra-fast technique.
6.2. Monte-Carlo Simulations

After performing a Monte-Carlo simulation on the possible dispersion existing between the two most spaced pulses within a single line (beginning and end of line), we obtain a sigma of only 18.42 ps rms. However, it is important to take into account the possible dispersion found between lines, i.e. across the entire sensor. By doing a similar simulation, we obtain a sigma of 37.32 ps rms, i.e. a mismatch of less than 90 ps FWHM. This result is illustrated in the histogram in Fig. 17.

In order to circumvent the inevitable mismatch, it is conceivable to modify (post-fab) the supply voltages of the inverters added in order to fine tune the additional skew.

![Interline Monte Carlo simulation showing possible dispersion due to fabrication mismatches.](image)

Fig. 17. Interline Monte Carlo simulation showing possible dispersion due to fabrication mismatches.

7. Conclusion

In the increasingly relevant field of ultra-fast imaging, gating signals play a key role in assuring optimal operation. When designing large resolution sensors, acquisition pulses must be uniformly distributed throughout the entire array. Due to the delay effect of a distributed RC line, this becomes troublesome. For this, a conventional driver at the beginning of the line is no longer adequate and a more robust solution needs to be implemented. This new approach is independent to pulse-based gating and instead uses an edge-triggered gating. Moreover, the designer could place such edge-based drivers on either side of the sensor to obtain a near-perfect gating signal distribution. While this technique is still hindered by signal skew, it can be eliminated using the method presented in this paper. As mentioned earlier, these effects can as much as triple the parasitic effects seen by one line. Therefore, our approach is crucial in ensuring the effective transmission of a sub-nanosecond pulse across a large sensor array. The maximal pixel operating frequency can be in the range of 50 MHz and the gating operation well below 1 nanosecond. The process mismatch show skew spread of about 90 ps FWHM on the entire sensor.

Acknowledgements

We thank the French agency “Agence Nationale de la Recherche” for financial support under the grant ANR-14-CE26-0024-01.

References


