High-speed PNP PIN Phototransistors in a 0.18 μm CMOS Process

Plamen Kostov, Wolfgang Gaberl and Horst Zimmermann
Institute of Electrodynamics, Microwave and Circuit Engineering, Vienna University of Technology
Gussauerstr. 25/354, 1040 Vienna, Austria
[plamen.kostov | wolfgang.gaberl | horst.zimmermann]@tuwien.ac.at

Abstract — In this work we present three speed optimized types of phototransistors built in a standard 180 nm CMOS technology without process modifications. An OPTO ASIC wafer consisting of a p-type substrate with a low doped p' epilayer on top of it is used for the implementation. The phototransistors were produced in 40×40 μm² and 100×100 μm² sizes. A gain in responsivity of more than 13 and bandwidths up to 50.7 MHz are achieved. As emitter followers, these phototransistors open the opportunity for application where high-speed photosensitive devices with inherent gain are needed. Possible applications are high speed opto-couplers, optical sensors, image sensors, etc.

I. INTRODUCTION

Since several decades CMOS processes are mature technologies in which sophisticated and cheap implementation of integrated circuits is possible. Furtheron CMOS technology even has some major advantages over III-V compound semiconductors. One advantage is the possibility to combine silicon photodetectors like photodiodes or phototransistors together with signal processing circuity on an optoelectronic integrated circuit (OEIC) for single-chip applications. These OEICs outperform compound photodiodes wire bonded to integrated circuits in many aspects. Silicon OEICs avoid bond pads and bond wires between photodiode and amplifier. They show excellent immunity against electromagnetic interference due to the short distances between the photodetector and the amplifier. Last to mention, they are well suited for cheap mass production [1].

The most common silicon photodetectors are photodiodes (PD) and phototransistors (PT). PDs can be split in several subcategories: pn PD, pin PD and avalanche photodiodes (APDs). Examples of PTs are vertical bipolar phototransistors, lateral bipolar phototransistors, field-effect phototransistors (photo-MOSFETs), photothyristors, and some more.

The most common PD is the pn-junction PD. This kind of PD is the simplest photodetecting device and it is easy to implement into a standard CMOS process. pn-junction PDs can be implemented in three different ways. First, the PD can be built as an n-type source/drain-implant into a p-substrate. This kind of structure has a very shallow and thin space-charge-region (SCR). However, the shallow and thin SCR is not suitable for detection of near infrared light like 850 nm, because most charges are generated deep in the substrate, outside of the SCR. 850 nm light has a 1/e penetration depth of about 16.6 μm [1]. Due to this fact, most photogenerated charges diffuse slowly to the drift zone (SCR-zone) leading to a slow device. Second, the PD can be built as an n-well to p-substrate junction. Such a kind of PD shows a deeper and thicker SCR compared to the n/p-substrate PD and is therefore more appropriate for detection of longer wavelengths. The main reason for the deeper and thicker SCR is the thinner and lower doped n-well. Nevertheless, the responsivity and bandwidth is still not optimal for 850 nm. An n-well/p-substrate PD with a bandwidth of 1.6 MHz and a responsivity of 0.5 A/W at 780 nm for V_D = 5 V is presented in [2]. To avoid the slow charges from the substrate a third PD type can be realized. The whole pn-junction can be isolated from the substrate by using an n-well. For p-substrate a p'/n-well PD avoids charges from the substrate if it is correctly biased. This kind of PD has a much higher bandwidth (mainly carriers from the drift zone contribute to the photocurrent) but due to the thin active layer a much lower responsivity than both first mentioned PD types.

Pin-junction PDs have an additional low doped epilayer between the p- and n-layer. This low doped layer leads to a thick SCR and thus to a larger fraction of drift current compared to diffusion current. The larger drift part leads to high bandwidth of the pin PD. A lateral pin PD with a responsivity of 0.20 A/W and a -3 dB bandwidth of 2.8 GHz at 850 nm for V_D = 20 V is presented in [3]. However the responsivity of pin PDs cannot exceed the maximum responsivity (0.65 A/W for 850 nm) for a quantum efficiency of 1.00 [1].

The responsivity limitation in PDs can be improved by special kinds of photodetectors with inherent current amplification. Such photodetectors are avalanche photodiodes (APDs) or phototransistors (PTs). These devices can achieve quantum efficiencies > 1.

The current amplification in APDs is achieved by avalanche multiplication. The drawback with this mechanism...
is that at least several tens of volts are needed to achieve sufficient electrical field strength for avalanche amplification [4]. Handling such high voltages is a problem in integrated circuits and even more in modern low-voltage processes. Another drawback is the very narrow bias voltage range for linear operation. Reference [5] reports on an APD with 4.6 A/W at 470 nm for $V_{DD} = 19.1$ V.

The most important advantage of PTs over APDs is that they do not need such high voltages for achieving their internal current amplification. Vertical bipolar PTs [6], lateral bipolar PTs [7] and photo-MOSFETs [8] are some examples for building PTs in CMOS technology. A vertical bipolar PT can be described as a PD, which is built by the base-collector (BC) junction and an internal bipolar junction transistor for current amplification (Fig. 1). $pnp$ vertical bipolar PTs in CMOS technology are formed by implanting an $n$-well (base) into the $p$-substrate (collector) and afterwards implanting a $p'$ source/drain (emitter) in the $n$-well. For long wavelength light, most charges are generated in the BC-PD. There the generated charges are separated and swept into base and collector. This leads to an accumulation of electrons in the base and therefore to a negative potential of the base. When the base potential becomes more negative, the base-emitter (BE) diode opens and the $p'$ emitter starts injecting holes into the base. Due to this procedure the primary photocurrent is amplified by the inherent gain of the device. CMOS PTs using a special wafer were presented in [6]. The used starting wafer has a double low doped epitaxial layer, a $p'$ epi and $n$ epi layer. These PTs in 0.6µm CMOS showed responsivities up to 37.2 A/W and bandwidths up to 14 MHz at 850 nm and $V_{CE} = -10$ V.

This paper describes integrated CMOS vertical bipolar PTs with high bandwidth achieved by the means of only one low doped epitaxial layer in the BC area of the PTs. Different types of layouts lead to PTs with different characteristics. These PTs open the opportunity for new cheap and fast devices for applications where high sensitivity together with higher bandwidth are required, e.g. high speed opto-couplers, active pixel sensors, light barriers, position detectors etc.

II. METHODOLOGY

Three speed optimized types of 40×40 µm$^2$ and 100×100 µm$^2$ $pnp$ PIN PTs were fabricated in a 0.18 µm CMOS process without any process modification. The PTs were produced on an OPTO ASIC wafer. To achieve high bandwidths a $p'$ epitaxial layer was placed between the base and collector area. This epitaxial layer leads to a thick BC-SCR and to an increase of the photocurrent drift portion, which furthermore leads to a higher bandwidth. Bandwidths as well as responsivities of the PTs are also depending on the thickness of the base and the size of the emitter area. These characteristics can be changed by changing the design of the base and emitter. To adjust the doping concentration in the base and thus the effective base thickness as well as the BC-SCR thickness, the PTs were designed with two different types of base implants. For analyzing the contribution of the emitter size to the responsivity, each PT has a unique emitter area. Furthermore a lateral 3 µm $p'$ epi-layer gap was added between the $n$-well base and the $p'$-well collector contact to decrease the perimeter capacitance.

A. Phototransistors

Three different layout versions were realized:

1) $50_{\Omega}$Center: The base of this PT is designed by 0.5 µm wide $n$-well stripes with 0.5 µm wide gaps between them. However, these stripes will diffuse into one layer during processing due to the thermal budget. In such a way the base of this PT is 50% lower doped than a full $n$-well. The emitter has a 0.74×0.74 µm$^2$ $p'$ area in the center of the PT. Fig. 2a shows a 3D depiction of this device.

2) $100_{\Omega}$Edge: A homogeneous $n$-well implant without gaps forms the base of this PT. The emitter is formed by a 2.18×0.32 µm$^2$ $p'$ area at the edge of the photosensitive area.

3) $100_{\Omega}$Quad: This device has the full $n$-well base like the device before. The only difference between these two PTs is the number and position of the emitter areas. The later PT has four emitters, placed in the center of each quadrant of the photosensitive area. Fig. 2b shows a 3D depiction of this PT.

All three PT versions were realized in 40×40 µm$^2$ and 100×100 µm$^2$ layouts. However the sizes for the $p'$ emitter and the $n$-well stripe/gap structure are the same.

B. Photocurrent Amplification

The $n$-well stripes of the base were implanted to achieve a different base doping concentration $N_B$ and therefore different Gummel numbers $N_G$. Due to a thermal step in the chip
production, the n-well stripes of the striped base PTs will diffuse into a single slightly inhomogeneous n-well layer with a lower doping concentration. This base doping concentration determines the relevant Gummel number [9]

$$N_G = \int_{0}^{\infty} N_B(x)dx ,$$  

(1)

where \(W\) is the effective width of the base. The effective width of the base is the area between the borders of the BE-SCR and BC-SCR. For all described PTs the resulting base layer is thick enough to prevent reach-through between the collector and emitter area for high voltages, which is illustrated in the following measurement section. Furthermore the base doping concentration and thus the Gummel number are indirectly proportional to the forward current gain \(\beta\) of the PT. This fact will lead to higher current gains for PTs with the lower doped base area. This is due to the fact that

$$\beta \propto \frac{N_E}{N_G} ,$$  

(2)

where \(N_E\) is the emitter doping concentration [9]. For achieving maximum responsivity values, a homogeneous emitter over the whole photosensitive area would be necessary. In such a layout photogenerated charges have to travel a minimum distance corresponding to the effective base width to reach the emitter. For other layouts using small emitter areas some charges have to travel longer distances if the charges are not generated directly under the emitter area. Due to this longer travel distance the probability for recombination in the base increases. This furthermore decreases the gain of the PT.

C. Bandwidth

The bandwidth of the PTs is mainly dependent on their BC- and BE-capacitance. These capacitances are formed by the SCR in both junctions. Since the width of the BC-SCR increases when applying a collector-emitter voltage \(V_{CE}\), the value of the BC capacitance gets smaller. This leads to a change of the PTs’ -3 dB frequency [10]:

$$f_{3dB} = \frac{1}{2\pi \beta \left( \tau_B + \frac{k_B T}{q I_E} \left( C_{BE} + C_{BC} \right) \right)} ,$$  

(3)

where \(\beta\) is the forward current gain, \(\tau_B\) is the base transit time, \(k_B\) is the Boltzmann constant, \(T\) is the absolute temperature, \(q\) is the elementary charge, \(I_E\) is the emitter current of the PT, \(C_{BE}\) and \(C_{BC}\) are the base-emitter and base-collector capacitances. The -3 dB frequency is more dependent on the BE capacitance than on the BC capacitance. This is due to the fact that the BE capacitance is larger compared to the BC capacitance for a full emitter layout. One way to reduce the BE capacitance is to reduce the emitter area of the PT, resulting in a higher -3 dB bandwidth of the PTs. Nevertheless, as described above, a small emitter size will lead to a reduced gain.

III. RESULTS AND DISCUSSION

The presented PTs were optically characterized by three different measurements. The output characteristics of the PTs were characterized by an optical DC measurement. The spectral responsivity of PTs was measured over the visible light spectrum. Optical AC measurements were done for characterizing the PTs’ responsivities and bandwidths, respectively.

A. Output characteristics

The output characteristics of the PTs were measured by applying 850 nm light at different optical light power \(P_{opt}\) and varying the \(V_{CE}\) voltage of the PTs. The base of the PT was left floating. Fig. 3 shows the output characteristic of the 100×100 µm² 500 Center ET PT. The optical power was varied from -37.7 dBm to -8.31 dBm. The output characteristic shows that no reach-through occurs for \(V_{CE}\) up to -13 V.

B. Responsivity measurements

Responsivity measurements were done at an optical power of -15.8 dBm. A with 630 kHz modulated 850 nm laser with an extinction ratio of 1.48 was used. The operating point of the PTs was selected by using three different \(V_{CE}\) voltages (-2 V, -5 V and -10 V) as well as at five different base currents \(I_B\) (0 µA, 1 µA, 2 µA, 5 µA and 10 µA). Due to the small emitter sizes of the PTs the measured responsivities are rather small. However the PTs were mainly optimized for speed. Responsivity values up to 2.34 A/W were achieved for the 100×500 µm² PT. Depending on the operating point of the PTs, responsivities up to 1.50 A/W and 1.34 A/W were achieved for the other PTs. The responsivity values were nearly the same for small and large sized PTs. Furthermore the responsivity for wavelengths between 400 nm and 900 nm was measured using a monochromator. The optical power from the monochromator varied between -35.7 dBm and -26 dBm over the full spectrum. The measured spectral responsivity for the 100×100 µm² 500 Center ET PT is depicted in Fig. 4. The measurements were done at \(V_{CE} = -2 V\) and floating base. This PT shows a maximal responsivity of 6.44 A/W at 783 nm and an incident optical light power of -35.1 dBm. The responsivity spectrum shows the influence of the full oxide and passivation stack. This effect can be eliminated in further designs by applying an optical window etch process step on the photosensitive area.

Figure 3. Output characteristics of the 100×100 µm² 500 Center ET PT at 850 nm with floating base for different \(P_{opt}\).
C. Bandwidth measurements

Bandwidth measurements were done by means of a modulated 850 nm laser and a vector network analyzer. Similar to the responsivity measurements, the bandwidths were measured at different operating points by varying $V_{CE}$ and $I_B$. PT 50$_{b}$Center$_E$ has the smallest emitter and thus the smallest BE capacitance which results in the highest bandwidth. Devices with a size of 40×40 µm$^2$ have higher bandwidths than 100×100 µm$^2$ devices due to a reduced BC capacitance as well as a reduced perimeter capacitance. Fig. 5 shows the frequency response of PT 50$_{b}$Center$_E$ at floating base and $V_{CE}=-2$ V as well as $V_{CE}=-10$ V for the two device sizes. The 40×40 µm$^2$ device achieves a -3 dB bandwidth of 50.7 MHz. PT 100$_{b}$Edge$_E$ and PT 100$_{b}$Quad$_E$ achieved maximum bandwidths of 50 MHz and 31.6 MHz, respectively as shown in Tab. I. PT 100$_{b}$Quad$_E$ shows for low voltages the highest bandwidths due to shorter diffusion length for generated charges.

IV. CONCLUSION

We present three types of 40×40 µm$^2$ and 100×100 µm$^2$ bipolar phototransistors built in a standard 180 nm CMOS process without process modification. Compared to the phototransistors presented in [6], we used an OPTO ASIC wafer with a $p'$ epitaxial layer on top of the $p$' substrate. Due to small emitter areas in the presented phototransistors maximal responsivities of 6.44 A/W were measured. Nevertheless, this is about 13 times larger than the values for $pn$-photodiodes presented in [2] or $pin$-photodiodes presented in [3]. Furthermore bandwidths of up to 50.7 MHz were achieved. The presented phototransistors are more than 30 times faster than the $pn$-photodiodes presented in [2] and more than 3 times faster than the phototransistors presented in [6]. Due to the presented results these phototransistors are well suited for applications where high speed photosensitive devices offering gain are needed. Possible applications are fast opto-couplers, three-dimensional camera, and so on.

### TABLE I.

<table>
<thead>
<tr>
<th>$V_{CE}$</th>
<th>0.0</th>
<th>5.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>50$_{b}$Center$_E$</td>
<td>14.2</td>
<td>25.1</td>
<td>50.7</td>
</tr>
<tr>
<td>100$_{b}$Edge$_E$</td>
<td>12.8</td>
<td>24.0</td>
<td>50.0</td>
</tr>
<tr>
<td>100$_{b}$Quad$_E$</td>
<td>18.6</td>
<td>25.1</td>
<td>31.6</td>
</tr>
</tbody>
</table>

**REFERENCES**


