Integrated Phototransistors in a CMOS Process for Optoelectronic Integrated Circuits

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

***Abstract* — This work presents integrated pnp phototransistors built in a 0.6 µm OPTO ASIC CMOS process using a low doped epi wafer as starting material. Several phototransistors with different designs of the base and emitter area were realized and characterized. For these novel photodetectors responsivities up to 65 A/W for DC light and up to 37.2 A/W for modulated light were achieved. Other transistors reach bandwidths up to 14 MHz. Due to the used standard silicon CMOS process low-cost integration is possible. Analog and digital circuitry can be implemented together with active optical detectors. This paves the way for high performance optical sensors and cost efficient SoC devices. Typical application examples include highly sensitive optical sensors, active pixel image sensors, light barriers and optical distance measurement sensors as well as 3D cameras.**

I. INTRODUCTION

Photodetectors convert optical signals into electrical ones. Different types of them can be built in a standard CMOS process. Often used photodetectors are e.g. PN-, PIN-diodes, avalanche photodiodes (APD) and phototransistors (PT).

PN diodes can be considered as the most common photodetectors. Integrated PN detectors are typically very shallow (only a few µm), which can be a problem for near-infrared light, since e.g. 850 nm has a 1/e penetration depth of around 16.6 µm [1]. Common PN diodes can be designed in two different structures which give either high speed or high responsivity. First, a PN diode might use a common well/substrate structure. This structure will convert charges generated in any depth, even deep inside the substrate. Since the charges in the field free substrate are only carried by diffusion (very slow mechanism) these detectors are rather slow. Therefore, the electrical performance shows a good responsivity but huge diffusion tails. Second, a PN diode can be designed to avoid charges generated deep in the semiconductor. This can be done by means of a p+/n-well structure, which leads to a rather shallow (~1 µm) remaining effective active thickness area. All charges generated below this area (slow charges) can be drained. Drained charges do not contribute to the electrical output signal. This leads to fast but very inefficient detectors.

PIN diodes avoid the limitations of the ordinary PN diodes, because of their additional low-doped intrinsic layer. PIN diodes are integrated in CMOS processes using special starting materials, where a 15 µm low doped epitaxial (epi-) layer is applied on the top of the wafer [1]. This epi-layer forms the intrinsic zone between the p- and n-area and leads to a thick space-charge-region (SCR). Inside this thick intrinsic zone an electrical field can be generated easily by a low voltage. The applied voltage leads to a thick drift zone. In that zone most of the charges are generated. By this mean, the detector gains speed and high responsivity at the same time. Therefore the PIN diode is the mostly used photodetector in applications where high speed is desired. Nevertheless, the responsivity is limited for optimum quantum efficiency ($\eta = 1$) to 0.65 A/W @ 850 nm and 0.55 A/W @ 650 nm [1]. In the case of optimum quantum efficiency all photons are used to generate charges inside the detector. Typical applications for PIN diodes are for example in data communication and optical distance measurement circuits [2].

Photodiodes have no inherent current amplification. Therefore their maximum quantum efficiency is limited to 1. This limit can be exceeded by the use of APDs or PTs. Both use mechanisms for amplifying the received photocurrent. This amplification is desirable and important for detecting weak optical signals. Amplification and very high gain in APDs is achieved by the avalanche multiplication process. However, this process needs voltages of several tens of volts [3]. High voltages for avalanche photodiodes are hard to handle in integrated circuits. This is a big drawback for systems on chip (SoC) applications. Other drawbacks of APDs are a rather high dark count rate in Geiger mode and the very narrow bias voltage range for linear operation. Any change in the bias voltage can lead to a nonlinear behavior. Such a behavior is therefore not feasible for image sensor or optical distance measurement applications, especially when measurements are taken in daylight conditions. Due to all these effects an APD needs a complex control circuitry for many application fields. In [4] a shallow APD with a responsivity of 4.6 A/W at 430 nm with a reverse bias of 19.5 V is reported.

PTs do not need high voltages for amplifying the current. This is a major benefit compared to APDs. A PT consists of a photodiode (base-collector junction) with an internal bipolar transistor for current amplification. In the photodiode (base-

Funding from the Austrian Science Fund (FWF) in the project P21373-N22 is acknowledged.

collector) region generated charges are separated. For a pnp PT electrons are swept into the base area and holes into the collector. The electron accumulation in the base makes the base potential more negative. Therefore the p+ emitter starts to inject holes into the base. This mechanism amplifies the generated (primary) photocurrent. PTs in standard-burried-collector (SBC) technology with a responsivity of 2.7 A/W at 850 nm were reported in [5].

In this paper, we present several silicon integrated PTs with different layouts. Cheap implementation in a CMOS process makes them easy to use as single chip solutions with silicon-based optoelectronic integrated circuits (OEICs). OEICs have many advantages, e.g. no bond wires are necessary, a simple mounting and packaging process can be used and a high mechanical reliability is reached. According to these benefits as well as the low operating voltage and high current amplification, these PTs are appropriate for image sensors, active pixels, optocouplers, light barriers and other SoC applications.

II. Device Structures

In our work, we implement several 100×100 µm pnp-type PTs in a 0.6 µm OPTO ASIC CMOS technology. The OPTO ASIC process used implements PIN photodiodes. The only difference compared to the PIN OPTO ASIC process is the starting material. All PTs use a wafer starting material with a thick (~15 µm) low doped (2×10^{13} cm^{-3}) p-epi layer and a shallow (~1 µm) n-epi layer (10^{14} cm^{-3}). The p-epi layer ensures a thick collector-base SCR even for low voltages. Such a thick SCR device is well suited for deep penetration light (e.g. 850 nm). Different designs of the base (n-type) and emitter (p-type) region lead to varying characteristics for each PT. These designs are illustrated in fig. 1 and 2. These designs as well as the collector area are explained in the following three subsections.

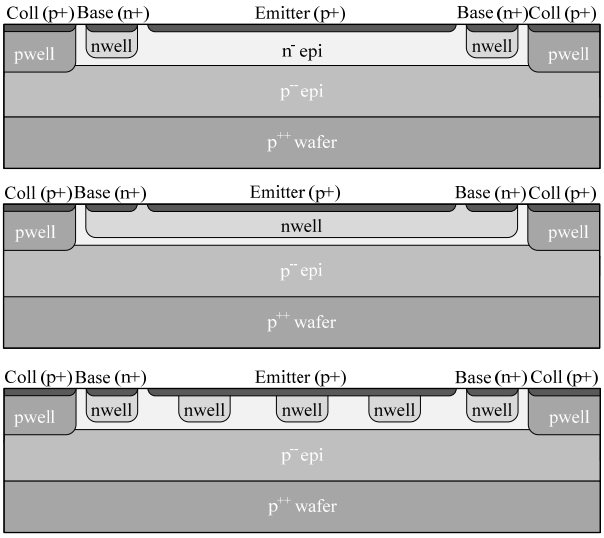

Figure 1. Cross-sections of the different base designs: n-epi (top), n-well (middle) and modulated n-well (bottom)

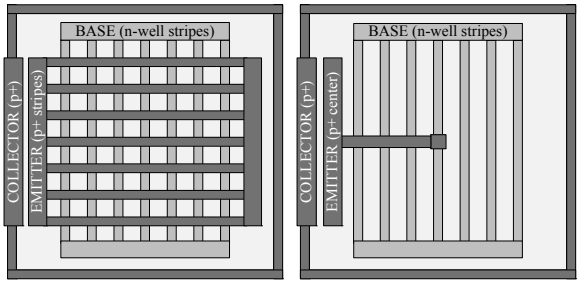

Figure 2. Top view of the different base designs

A. Collector Area

The p-type substrate forms the collector. It is connected via a large-area ring of substrate contacts on the border of the PT. Due to the fact that the collector is tied to substrate potential, the pnp PT can only be used in emitter follower setup.

B. Base Area

The base is formed by the 1 µm thick low doped n-epi layer. To further adjust the base doping additional n-well areas can be implanted. Three different base doping layouts are investigated: a low doped epi-layer base, a full n-well base and a base with n-well stripes.

First, the topmost picture of fig. 1 shows the lowest doped base layer. This 1 µm thick design leads to an increased thickness of both SCRs (BC, BE). The SCR increase results in a reduced effective base width W_B. The reduction of W_B leads to a decrease of the base-collector and base-emitter capacitances C_{BC} and C_{BE}. Accordingly the PT's bandwidth increases due to faster transport of the minority carriers through the (thinner effective) base region. C_{BC} and C_{BE} are the reason why PTs are slower than photodiodes. Both capacitances and the base transit time τ_B are the main parameters defining the –3 dB bandwidth of PTs [6]:

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

The increase of both SRCs causes the effective base width to shrink, which may become a problem when both SCRs touch each other. In this case a reach-through current between emitter and collector will arise. This current is dependent on the collector emitter voltage.

Second, the middle picture of fig. 1 shows the highest doped base. It uses a full n-well. The higher doped base leads to larger capacitances, which results in a slower device. Furthermore, the higher doping increases also the effective base width W_B. Therefore the current gain β decreases as

$$\beta = \frac{1}{\frac{W_B^2}{2\tau_b D_p} + \frac{D_n}{D_p}\frac{W_B}{L_n}\frac{N_D}{N_A}} \qquad (2)$$

where τ_b the minority carrier lifetime in the base, D_p and D_n the carrier diffusion coefficients in the base, L_n the diffusion length of electrons in the emitter, and N_D and N_A the donor and acceptor densities in the emitter and base are [7].

Third, to achieve a good tradeoff between gain and speed, n-well stripes with three different width-to-spacing ratios were designed:

- NW_{33}: 1 µm stripe with 2 µm space,
- NW_{50}: 1 µm stripe with 1 µm space and
- NW_{66}: 2 µm stripe with 1 µm space.

Due to the thermal budget of the process, the different stripes diffuse into an inhomogeneous doped base layer. Thereby it is possible to adjust the effective base doping and the effective base width even in a standard ASIC process.

C. Emitter Area

Different emitter designs have been realized. The largest emitter is a full emitter plane. Other designs with smaller emitters are shown in fig. 2.

The 97×97 µm full plane emitter leads to a low current density resulting in a reduced cut-off frequency. The main advantage of the full plane emitter is a very homogenous electric field inside the PT so generated charges only need to transit vertically. However, the full plane emitter area also increases the base-emitter capacitance C_{BE}. This increase dominates and causes according to (1) a reduced cut-off frequency.

The left structure of fig. 2 shows an emitter layout with stripes (1.4 µm stripes, 8.4 µm gap), whereas the right one shows a small emitter dot (1.4×1.4 µm). These structures have a higher current density and lower capacitance due to the reduced emitter area. Thereby the cut-off frequency is increased. However, generated charges have to travel longer distances for smaller emitter areas which results in a reduced gain due to recombination. Furthermore a device with a small emitter area (7.0×1.4 µm) in one corner of the PT was produced. This structure has no emitter inside the optical area and also no metal structures. Therefore this structure might be combined with an opto-window etching process step to reduce attenuation and reflection effects of the oxide stack. However, to compare the different device layouts against each other, the corner emitter design was not realized with an opto-window. With an additional opto-window the optical performance of the corner layout might be increased by around 2-3 dB depending on the used wavelength.

III. Results and Measurements

Three different measurement setups were used for characterizing the PTs. The electrical current amplification was verified through Gummel measurements. The DC responsivity was measured by sweeping the power of the light source at 675 nm and 850 nm. Finally the AC responsivity respectively the bandwidth was measured at an average optical power of –10 dBm and –21.2 dBm (@ 850 nm). For these measurements the extinction ratio of the light source was chosen at 2:1.

A. Gummel measurements

All PTs with only the low doped epi-layer base show a reach-through effect due to the above mentioned reason. All devices with higher doped base show the expected current amplification β between 57 (for the full n-well base and corner emitter PT) and 176 (for the NW_{33} base and small center emitter PT – fig. 3). A different performance is measured with the NW_{33} base and full emitter PT, which is at the edge of a reach-through scenario. It has an abnormal Gummel plot,

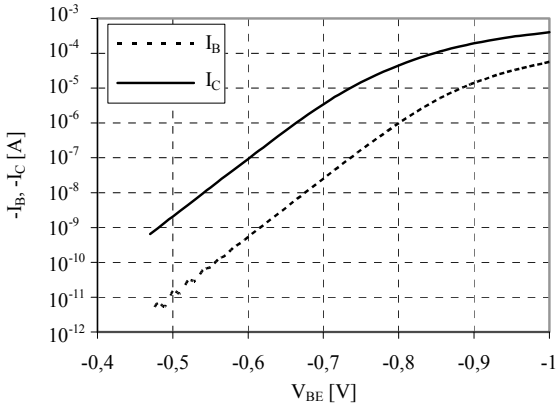

Figure 3. Gummel plot of the NW_{33} base and center emitter transistor

wherein the I_B and I_C graphs run nonparallel. For low currents the gain reaches up to 10^5. Nevertheless, the dark current was only about 4.56 µA at a collector-emitter voltage of –4 V. For all other (non-reach-trough) devices dark currents below 60 pA were measured.

B. DC measurements

The DC responsivity was measured at 675 nm and 850 nm by changing the optical power with an optical attenuator. The optical light power was varied from –55 dBm to –11 dBm and monitored by using an optical 50/50 beam splitter and optical power meter. Increasing the light power shows an exponential decrease of the responsivity due to the operating point variation. Furthermore, the applied collector-emitter voltage was also varied from –1 V to –8 V.

For 675 nm three PTs were measured. The devices show only a small dependence on the collector-emitter voltage. A maximum responsivity of about 65 A/W for the NW_{50} and full emitter PT was achieved.

Since 850 nm light has a larger penetration depth, the effect of the SCR is of higher influence. Values between 20 A/W and 40 A/W were achieved at –55 dBm for small emitter devices. The responsivity for these devices at –11 dBm was about 2 A/W. The smallest responsivity change was measured at the full and striped emitter devices. The responsivity change is below 20 % over the whole light power range. A totally different result shows the near to reach-through PT (NW_{33} base and full emitter). For this PT the corrected responsivity (with subtracted dark current) is in the range of 23000 A/W at –55 dBm optical light power. Moreover at –11 dBm optical light power the device still shows a responsivity of 80 A/W.

C. AC measurements

AC light was used for measuring the dynamic responsivity and the bandwidth of the PTs. The measurements were done at a low optical power level (–21.2 dBm) as well as at a high one (–10 dBm). Eight devices were characterized at three different collector-emitter voltages: V_{CE} = –2 V, –5 V and –10 V. The operating points of the PTs were adjusted by sinking a current I_B from the base (via an on-chip 1 MΩ resistor). Five different operating points, including floating base were investigated. Figure 4 depicts the results of the measurements for 850 nm.

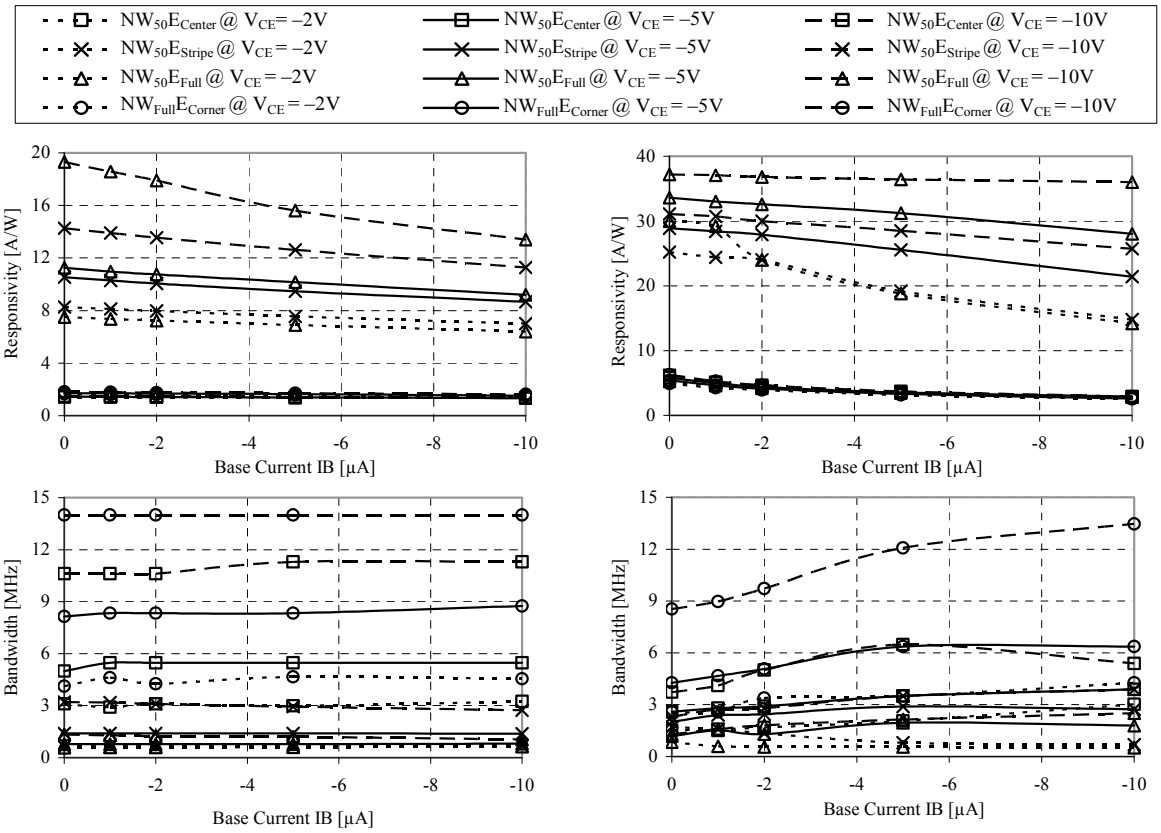

Figure 4. The two left diagrams show responsivity (top) and bandwidth (bottom) for high optical power (–10 dBm). The two right diagrams are respectively for low light conditions (–21.2 dBm).

The measurements at –10 dBm show a decrease of the responsivity for larger base currents. The small-emitter devices have a small but rather constant responsivity. The larger-emitter devices show a collector-emitter voltage depending responsivity which is generally higher. The bandwidth is nearly constant for all operating points, since the incident optical power already ensures a certain operating condition. However, the bandwidth is considerably influenced by the collector-emitter voltages mainly due to increased SCRs, which leads to a thinner effective base.

The measurements at –21.2 dBm show a decrease of the responsivity for higher base currents for all PTs. The responsivity of the small emitter size devices does not depend much on the collector-emitter voltage. The devices with larger emitter sizes (striped and full emitter) show a strong increase of the responsivity for higher collector-emitter voltages. The collector-emitter voltage mainly affects the bandwidth of the devices. Smaller emitter layouts show higher bandwidths than larger.

IV. CONCLUSION

New fully integrated silicon phototransistors in a CMOS OPTO ASIC technology are presented by using special starting material wafers. Several devices with different designs of base and emitter area have been produced and characterized. Electrical Gummel plot measurements showed current gains up to 176. Responsivities up to 65 A/W (about 25 times better than that of SBC npn phototransistors in SiGe BiCMOS) and bandwidths up to 14 MHz were measured. A close-to-reach-through phototransistor showed operating point dependent current amplification up to 50.000 for low optical input power. The responsivity and bandwidth can be adjusted by the layout (n-well stripe widths and spacing as well as emitter area and position). Fully integrated custom tailored phototransistors therefore are well suited for many optical sensing applications, SoCs and also imaging systems.

REFERENCES

[1] H. Zimmermann, “Integrated Silicon Optoelectronics”, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 2010.

[2] G. Zach, M. Davidovic and H. Zimmermann, “Extraneous-Light Resistant Multipixel Range Sensor Based on a Low-Power Correlating Pixel-Circuit”, IEEE ESSCIRC 2009, pp. 236-239.

[3] S. Cova, M. Ghioni, A. Lacaita, C. Samori and F.Zappa, “Avalanche photodiodes and quenching circuits for single-photon detection”, Applied Optics, vol. 35, no. 12, pp. 1956-1976, 1996.

[4] A. Pauchard, A. Rochas, Z. Randjelovic, P.A. Besse and R.S. Popovic, “Ultraviolet Avalanche Photodiode in CMOS Technology”, IEEE IEDM 2000, pp. 709-712.

[5] T. Yin, A.M. Pappu, A.B. Apsel, “Low-cost, high-efficiency, and high-speed SiGe phototransistors in commercial BiCMOS”, IEEE Photonics Technology Letters, vol. 18, no. 1, 2006, pp.55-57.

[6] G. Winstel and C. Weyrich, “Optoelektronik II”, p. 97, Springer, Berlin, Heidelberg, 1986

[7] P. Gray, P. Hurst, S. Lewis and R. Meyer, “Analysis and Design of analog integrated Circuits”, p. 8 ff., Wiley, 2008