CMOS photodiodes based on vertical p-n-p junctions

Philippe Lalanne and Jean-Claude Rodier
Institut d'Optique Théorique et Appliquée, CNRS, BP 147, F91 403 Orsay Cedex, France
e-mail: philippe.lalanne@iota.u-psud.fr

abstract

A device composed of two junctions, but operating as a photodiode is designed and implemented in a pre-production 1 µm complementary-metal-oxide-semiconductor silicon technology foundry service. No process modification is performed. Tests are performed at a wavelength of 783 nm. Rise and fall times in the nanosecond range are reported along with sensitivity and bandwidth measurements. The suitability of the device for optical interconnects is discussed.

Key words: CMOS photodetector, optical interconnects.

1. Introduction

For the past ten years, it has been noted in the literature that integration of a fast photodetector in complementary-metal-oxide-semiconductor (CMOS) technology is a critical concern for optical interconnection in silicon very-large-scale-integrated (VLSI) circuits. There are several advantages to this approach. Firstly CMOS circuitry is less expensive and generally dissipates less power than Gallium Arsenide or bipolar silicon technologies. For instance, CMOS receivers implemented in 0.8-µm technology and operating at 550 Mbit/s consume ≈5 mW of power and with forthcoming technologies, lower consumptions can reasonably be predicted. Second, in comparison to employing discrete hybridised detectors, the detector integration lowers the power dissipation and time delay needed to drive hybridised components. The speed and the sensitivity of a simple p-n junction photodiode which can be easily integrated with CMOS technology are largely functions of the depth of the p-n junction, the bias voltage applied across the junction, and the wavelength of the light. As it is know from p-i-n type photodetectors, for fast photodetection, nearly all the photons must be absorbed in the intrinsic region in which an electric field will sweep the optically generated electrons and holes across the junction to form a photocurrent. Because only a few process steps are available with CMOS technology, the intrinsic region of p-i-n photodiode cannot be fabricated with standard
CMOS technology, and only p-n junctions can be implemented. Even for a large bias voltage, the depletion-region thickness is no larger than approximately 3 µm. For wavelengths in the range of 800 nm, a large part of the photocarriers are generated outside the depletion region. These photocarriers first diffuse to the depletion layer then drift across the junction to generate a diffusion-type photocurrent. Because the diffusion process is driven by the random motion of carriers, it is a slow transport mechanism. This was perfectly interpreted and experimentally confirmed in the work of reference where the rise and fall time of a vertical CMOS photodiode were measured to be approximately 45 ns and limited by diffusion of deep photocarriers. Similar results were reported later on. In Ref. 7, two different detector types implemented with a 3-µm CMOS p-well process are studied. With a laser diode emitting at 819 nm, rise and fall times of approximately 60 and 45 ns were observed for lateral and vertical p-n photodiodes, respectively. In Ref. 8, the monolithic integration of optical waveguides and CMOS circuitry is investigated. Because of the slow photocarrier diffusion generation in the substrate, a maximum operating frequency of approximately 2 MHz was obtained. More recently, a significant improvement was achieved with a 1.25 µm CMOS process slightly modified in order to incorporate "intrinsic" p-well doping of 2.8 \(10^{14}\) cm\(^{-3}\), and rise and fall times of 2.6 and 7.9 ns respectively were observed at a wavelength of 790 nm.

In this paper, we report on a photodetector device fabricated in a production 1 µm CMOS technology. As shown in Fig. 1, the device is fully compatible with CMOS circuitry; no process modification is performed. The photodetector is based on a vertical p-n-p structure. It is composed of two p-n junctions. The p+ source diffusion and the n-well are used to form the top junction. The bottom junction is formed by the common cathode (n drain well) and the bottom substrate anode. The two junctions are operated with a reverse bias and no transistor effect occurs. We denote by \(V_1\) and \(V_2\) the two voltage supplies applied to the top and bottom junctions, respectively. Under normal operation, \(V_1\) and \(V_2\) are equal to 5 V. To our knowledge, it is the first time that the operation of this p-n-p vertical structure is studied as a photodetector and that nanosecond fall and rise times are observed with standard CMOS technology.

2. Experimental results

With the 1µm CMOS (E.S.2) n-well technology we used, the substrate is moderately doped (impurity concentration of \(10^{15}\) cm\(^{-3}\)), and the n well and the p+ diffusion doping concentrations are 2 \(10^{16}\) cm\(^{-3}\) and 2 \(10^{19}\) cm\(^{-3}\), respectively. The area of the top diffusion is 20x20 µm\(^2\), and the area of the n-well is 28x28 µm\(^2\). Two photocurrents are observed with the device of figure 1. The photocurrent flowing through the top junction is collected at the top anode. It will be refereed to as the top photocurrent. The photocurrent flowing only through the bottom junction (or photodiode) will be referred to as the bottom photocurrent. No experimental results are provided for this photocurrent. The photocurrent flowing through the two junctions can be collected at the cathode. It will be referred to as the global photocurrent. Consistently, this global photocurrent will be said to be generated by an effective global photodiode which is formed by the two junctions in parallel. Measurements performed on the device to determine dark current, responsivity as well as rise and fall times are reported.
Fig. 1 Sketch of the cross-section of the vertical p-n-p structure studied in the paper. The two junctions are operated with a reverse bias. Three photocurrents can be observed: the top and the bottom photocurrents flowing through the top and the bottom photodiodes, respectively, and the global photocurrent collected at the common cathode.

For the measurements, the light reflected from the test integrated circuit carries the superimposed images of the focused infrared laser beam and of a white-light widely illuminating the circuit; both images are observed with a CCD video camera and a video monitor. This setup allows high-magnification viewing capabilities to facilitate identifying the sample area under investigation and the exact location of the infrared spot. The infrared measurements were performed with a Mitsubishi ML-3101 semiconductor laser diode emitting at 783 nm. The laser spot focused onto the integrated circuit was observed to have a full width at half intensity of 4 µm in both directions. The photodiode response-time measurements are performed with a 50 Ω probe, an oscilloscope and a square-wave pulse generator with a temporal resolution in the nanosecond range (rise and fall times down to 1.5 ns). Since the parasitic capacitance of the 50 Ω probe is 3 pF, the pad and photodiode junction capacitances which are smaller than 1 pF are swamped out, yielding response times limited by the instrumentation. This was confirmed by measuring the overall instrumentation response time with a high-speed avalanche photodetector from Mitsubishi (model PD1002-01).

The overall response times of the top and global photodiodes are shown in Figs. 2 and 3, respectively. In Fig. 2, the bottom trace represents the input from the square-wave pulse generator and the top trace, the photodiode output. Small oscillation amplitudes are visible on the top trace; they are due to parasitic inductance effects at high frequencies. The visual comparison of the two traces shows that, in our experimentation, the rise and fall times observed on the top trace are not intrinsic to the photodiode but are rather imposed by instrumentation limitations. No slow tail effect is observed at a lower frequencies with the top photodiode. Thus it is concluded that the top photodiode exhibits a high response speed with sub-nanosecond rise and fall times. In Fig. 3, the global photodiode response time is shown. The time scale, 20 ns per division, is twice that of Fig. 2. This longer temporal scale is chosen in order to properly visualise the slower impulse response. Once again, a symmetric operation is observed and both rise and fall times are equal to 8 ns. The knowledge of the rise and fall times is not sufficient to provide a satisfactory description of the global-photodiode response since two different contributions with very different time constants are clearly identified in the temporal measurement. First, as for the top photodiode, a high speed response in the nanosecond range is observed. Second, a long tail
which amounts to approximately 25 % of the whole collected photocurrent is observed. The temporal measurements of Figs. 2 and 3 are typical of those obtained for a large range of optical power inputs from 100 µW to 1 mW.

**Fig. 2** Oscilloscope traces showing the response of the top photodiode at a wavelength of 783 nm. The lower trace represents the input from the square-wave pulse generator.

**Fig. 3** Oscilloscope traces showing the response of the global photodiode at a wavelength of 783 nm. The lower trace represents the input from the square-wave pulse generator.

The responsivities at 783 nm of the top and global photodetectors were estimated to be 0.048 A/W and 0.38 A/W, respectively. These responsivities are affected by a 11% reflectivity loss at the interface air/oxide/silicon. If this Fresnel loss is compensated for, the internal quantum efficiencies of the top and global photodetectors are 0.085 and 0.67, respectively. The two photodiodes in Fig. 1 have a very low dark current smaller than 1 pA. In addition, while maintaining their speed performance, they can be operated at low bias voltages directly compatible with voltage supplies of sub-micron resolution circuitry. Figure 4 represents the sensitivity (mA/W) of the top photodiode as a function of the bias voltage $V_2$ of the bottom photodiode. Two curves are shown. The solid and dashed curves are obtained for bias voltages $V_1$ equal to 5 and 3V, respectively. It has to be noted that, while measuring the top-photodiode sensitivity, we carefully checked that no speed-performance degradation occurs. Specifically, over the whole range of bias voltages ($V_1$, $V_2$), temporal characteristics similar to those of Figs. 2 and 3 were observed. From Fig. 4, it is concluded that the influence of the bias voltage $V_2$ on the top-photodiode sensitivity is rather weak; a few percent variation is obtained while varying $V_2$ over a large range from 1 to 5V. For any value of $V_2$, lowering the bias voltage $V_1$ from 5 to 3V only results in a 5% loss sensitivity of the top photodiode. Thus, we conclude that the sensitivity and speed performance of the top-photodiode are weakly dependent on the bias voltages applied to the vertical p-n-p structures.

The spot positioning effect was examined for the top and global photodiodes. We found that the device overall performance is independent of the spot position as long as the infrared light beam is focused within the 20x20 µm$^2$ area of the top diffusion layer. A very fast response and a good sensitivity of the global photodiode is obtained for a light spot incident onto depletion layer 2. As the spot position is moved away from the p-n-p device, the sensitivity and the speed of the global photodiode decrease. We also noted that, when the light spot is incident outside the depletion layer 2, no photocurrent is flowing through
the top photodiode; for a 2 mW incident power, we measured a 40 pA photocurrent flowing through the top junction. It is believed that this photocurrent is rather induced by a parasitic reflection illumination (visible on the video monitor) of the top photodiode than by electrical crosstalks.

**Fig. 4** Sensitivity of the top photodiode as a junction of the bias voltage $V_2$ applied to the bottom photodiode. The solid and dotted curves are obtained for $V_1=5$ and 3V, respectively.

### 3. Discussion

The difference in behaviour between the two photodetectors can be understood by considering the simple model shown in Fig. 1.

The photocurrent flowing through the top junction of the top photodetector is due to photocarriers which diffuse from the top anode and cathode regions to the depletion layer and to the photocarriers directly generated into depletion layer 1. Because the bottom reverse-biased p-n junction collects the deep-generated photocarriers, no slow tail effect in the time response of the top photodiode is observed. Neglecting electron-hole recombination, the internal quantum efficiency $Q_{E_1}$ of the top photodiode can be directly related to the critical depth value $d_c$ above and below which the photocarriers created in the cathode region diffuse to the top and bottom junctions, respectively. We have

$$Q_{E_1} = 1 - \exp(-\alpha d_c).$$

For $\alpha = (9\mu m)^{-1}$, a reasonable value$^{10}$ at a wavelength of 783 nm, the critical depth value $d_c$ is equal to 0.8 $\mu$m. Although the response speed observed in Fig. 2 for the top photodiode is in the nanosecond range, we believe that it is much faster and that our observation is strongly limited by the instrumentation. This assumption is mainly motivated by the fact that the operation of the top photodiode is not relying on a slow diffusion transport-phenomenon with transit time of typically 50 ns, but on a fast drift process with transit time typically less than 0.1 ns. The electron-hole pairs generated below the junction do not contribute to the top photocurrent but are swamped away in the bottom junction. Based on this operation principle, photodiodes with large bandwidths of approximately 700 MHz at -
3 dB were recently reported\textsuperscript{11,12} with BiCMOS technology. As for the top photodiode studied in this paper, these photodiodes exhibit rather small sensitivities; for a wavelength of 850 nm, sensitivities of 0.045 A/W and 0.07 A/W are obtained in Refs. 10 and 11, respectively. These sensitivity values are very similar to that obtained at 783 nm with the top photodiode studied in this paper. In the perspective of high data-rate optical interconnects, the top junction is of main interest because high speed response is potentially available at any wavelength of the near-infrared region (700-900 nm). However, even though optical links with Gbit/s data transmission rates were previously demonstrated\textsuperscript{10,13} by using low-sensitivity detectors, we believe that, except if burdensome high-speed and high-gain amplifier circuitry is implemented, the use of an optical receiver with a sensitivity as low as 0.048 A/W is an important penalty for the link power budget especially for optical interconnections with near-infrared lasers.

According to the model of Fig. 1, the photocurrent flowing through the bottom junction is generated by electron-hole pairs which are created in the depletion layer \textsuperscript{2} and by those which do not experience an electric field and diffuse from the cathode region and the substrate below the bottom junction. The deep-generated photocarriers which slowly diffuse from the substrate back to the bottom depletion layer are responsible for the slow tail observed in the time response of Fig. 3. To obtain a more quantitative understanding of the bottom photodiode performance, the simple model of Fig. 1 based on abrupt junctions has to be modified to incorporate the actual doping concentration profile of the p-n-p device. According to the E.S.2 datasheet, two modifications have to be taken into account. Firstly, while the carrier concentration is constant from 3 µm to 7 µm, an increase of three decades in carrier concentration is noted for depths increasing from 7 to 12.5 µm. Secondly, in the cathode region, the n-type carrier concentration is not constant. These two regions which experience an electric field make the interpretation of the temporal and static characteristics of the global photodiode difficult. For instance, the internal quantum efficiency of the global photodiode not only depends on deep-photocarriers recombinations, but also on the spatial variation of the doping concentration below the depletion layer \textsuperscript{2}. The expected existence of an electric field in the cathode region of Fig. 1 and the observation of a fast response in Fig. 3 which amounts to 75\% of the global photocurrent poses an intriguing question: How fast is the high-speed contribution of the global photocurrent? To answer this question, we measured the spectral response of the global photodiode for small signal modulations. The power spectrum obtained with a spectrum analyser shows a nearly-flat response from DC to 500 MHz. For modulation frequencies larger than 500 MHz and up to 700 MHz, an attenuation of -3 dB/100 MHz was observed. At higher frequencies, inductance resonance effects due to the circuit mounting did not permit reliable measurements. The observation of a nearly-flat response confirms that the fast contribution in the global photocurrent is potentially able to deal with data transmission rates of several hundreds of MHz. Thus, when operated at high frequency (100-400 MHz), the photocarriers which slowly diffuse to the bottom junction will provide a DC photocurrent, whereas the photocarriers which experience an electric field will provide a fast response. According to the previous measurements, the photocarriers which contribute with a fast response to the global photocurrent are responsible for an effective internal quantum efficiency which amounts to 0.5 (0.75x0.67). This quantum efficiency is much higher than that of the top photodetector. In our opinion, the global photodiode offers a good tradeoff between speed and sensitivity and is potentially interesting for optical interconnects with moderate data-rates of a few hundreds of Mbit/s.
4. Conclusion

An optoelectronic device consisting of two p-n junctions was designed and implemented in a 1 µm CMOS (n-well) technology without any modification of the fabrication process. The top junction is formed by a p+ source diffusion and a n-drain well. The n-drain well and the substrate are used to form the bottom junction. Two photocurrents are of interest; the top photocurrent flowing through the top junction and the global photocurrent flowing through the two junctions and collected at the common n-well cathode. The operation of the device was tested with a "CD" laser diode emitting at 783 nm. Because the photocarriers collected by the top junction do not experience a slow diffusion-transport mechanism, a high response speed was observed for the top photodiode. Although instrumentation limitations prevented us from observing rise and fall times shorter than one nanosecond, sub-nanosecond response times can reasonably be predicted. This very high speed response is obtained at the expense of a low sensitivity of 0.048 A/W. The global photodetector has rise and fall times of 8 ns. Its sensitivity is 0.38 A/W. The 8 ns rise and fall times are due to photocarriers which slowly diffuse back to the bottom junction. The contribution to the global photocurrent of the photocarrier generated in the upper part of the p-n-p device provides a high internal quantum efficiency along with a large bandwidth of a few hundreds of MHz.

To our knowledge, the two photodetectors reported herein offer the highest response speeds obtained with a pre-production CMOS technology. More important is that this work indicates that the performance of CMOS photodetectors strongly depends on the technology parameters like the doping profile. With forthcoming technology, the characteristics of the p-n-p device studied in this paper are likely to be slightly modified, but we expect that this work will contribute to effectively reduce the speed gap between CMOS transimpedance amplifiers and CMOS photoreceivers.

5. Acknowledgements

We are indebted to Antoine Dupret and Eric Belhaire from the Institut d'Electronique Fondamentale in Orsay and to Patrick Garda from the Université Pierre et Marie Curie in Paris for stimulating discussions and their help for the design. Furthermore, we are grateful to Pierre Chavel for his comments and for his care proofreading. The circuits were implemented thanks to the CMP service in Grenoble (France).
6. References