High-bandwidth uni-traveling carrier waveguide photodetector on an InP-membrane-on-silicon platform

L. Shen,1,* Y. Jiao,1 W. Yao,1 Z. Cao,2 J. P. van Engelen,1 G. Roelkens,1,3 M. K. Smit,1 and J. J. G. M. van der Tol1

1Photonic Integration Group, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
2Electro-Optical Communications Group, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
3Photonics Research Group, Ghent University-IMEC, B-9000 Ghent, Belgium

*l.shen@tue.nl

Abstract: A uni-traveling carrier photodetector (UTC-PD), heterogeneously integrated on silicon, is demonstrated. It is fabricated in an InP-based photonic membrane bonded on a silicon wafer, using a novel double-sided processing scheme. A very high 3 dB bandwidth of beyond 67 GHz is obtained, together with a responsivity of 0.7 A/W at 1.55 μm wavelength. In addition, open eye diagrams at 54 Gb/s are observed. These results promise high speed applications using a novel full-functionality photonic platform on silicon.

© 2016 Optical Society of America

OCIS codes: (230.5160) Photodetectors; (130.3120) Integrated optics devices.

References and links

#259934 Received 23 Feb 2016; revised 1 Apr 2016; accepted 4 Apr 2016; published 8 Apr 2016
© 2016 OSA 18 Apr 2016 | Vol. 24, No. 8 | DOI:10.1364/OE.24.008290 | OPTICS EXPRESS 8290
1. Introduction

Heterogeneous integration has been demonstrated as a promising technology for realizing photonic integrated circuits (PICs) on silicon. It not only allows for monolithic integration of high-performance InP-based light sources on silicon [1–5], but also provides photodetectors (PDs) with high efficiency and low dark current [6]. Currently, the reported bandwidths of heterogeneously integrated InP-based PDs [7, 8] are lower than those of Ge-based PDs [9, 10]. In order to have a full-functionality photonic platform for high-speed applications, waveguide-coupled PDs with higher bandwidth need to be developed. The InP-based uni-traveling carrier photodetector (UTC-PD) is one of the most promising structures for high-speed operation [11]. Unlike in p-i-n PDs, only electrons act as the active carrier in the depletion region of UTC-PDs. As the velocity of electrons is higher than that of holes (especially in the velocity overshoot regime), a lower transit time and reduced space charge effects can be achieved. In the past decade the UTC-PD and its variants exhibited very high bandwidth and saturation power in InP-based PICs [12, 13]. UTC designs were also investigated in Ge-based PDs on silicon [14]; however, their performance is still inferior to the InP-based devices, primarily due to the lack of velocity overshoot of electrons and the inability for complex bandgap engineering in Ge/Si.
structures. Recently, InP-based modified UTC-PDs heterogeneously integrated on silicon-on-insulator (SOI) waveguide circuits were demonstrated showing excellent high-power performance, thanks to the complex bandgap engineering in their structures [7]. The 3 dB bandwidth was 48 GHz, limited by the RC constant resulting from a relatively large device area. More compact PDs with lower capacitance are demanded for very high data rate (100 Gb/s and beyond) applications in optical interconnects and telecommunications.

In this paper, we present a novel type of heterogeneously integrated UTC-PDs aiming for high bandwidth operation. It is fabricated in an InP-based membrane which is bonded on a silicon wafer using benzocyclobutene (BCB). This InP-membrane-on-silicon (IMOS) platform has exhibited high-density integration of electrically pumped lasers and a large variety of passive devices [5, 15–17]. In this platform, all photonic components are formed in the InP membrane. The silicon wafer is used as a carrier wafer, and in the longer term, will carry electronic ICs. Thanks to the high vertical refractive index contrast, light is tightly confined in the InP membrane, which allows for very compact devices. Moreover, as the PD is butt-coupled to waveguides (WGs) in the same membrane, the light is absorbed immediately after entering the PD. This is in clear contrast to other heterogeneously integrated PDs on SOI, where the light is evanescently coupled from SOI WGs to PDs over a certain length before being absorbed. Consequently, on the IMOS platform shorter PDs with lower junction capacitances are possible, which allows high bandwidth operation. In this paper, a $3 \times 10 \mu m^2$ PD is presented with a 3 dB bandwidth of $> 67$ GHz and a responsivity of 0.7 A/W at 1.55 $\mu$m. Open eye diagrams at 54 Gb/s are obtained. To the best of our knowledge, the bandwidth and the data rate achieved in this work are the highest values reported to date by heterogeneously-integrated PDs on silicon.

Furthermore, a new double-sided processing scheme is used to process the photonic membrane from both sides, before and after bonding to silicon. As will be seen from the design section of this paper, this approach enables an extra degree of freedom in optimizing the device performance. This technology is uniquely suitable for the IMOS platform. As the light only propagates in the InP membrane layer where all photonic components are located, critical alignment to another WG layer in the silicon wafer is not necessary during the bonding, which assures the practicality of this technology. Thanks to the use of BCB as the bonding material, complex topography on the processed samples can be easily accommodated in the bonding process, resulting in high yield. The successful demonstration of this process technology promises new approaches for monolithic integration of UTC-PDs with lasers, by processing them from either side of the membrane. This is of high importance in the longer term for realizing a single membrane-based photonic platform with high performance and full photonic functionality on top of Si-based electronic ICs.

2. Design and fabrication

The cross-sectional design of the UTC-PD is shown in Fig. 1(a). A 150 nm thick p-type doped InGaAs layer is used both as the absorption layer and as the p-contact layer. The photogenerated holes are collected directly by the p-contact while the electrons travel to the depletion region in the i-InP layer. The thickness of the p-InGaAs layer is chosen as a trade-off between the modal absorption coefficient and the transit time of electrons. On the one hand, the thickness assures a sufficient quantum efficiency ($> 97\%$) for a 10 $\mu$m PD according to simulations; and on the other hand, it leads to a transit time through the absorption layer ($\tau_e$) of about 1 ps for a pure diffusion transport of electrons [18]. This time constant is comparable to the time constant for electrons drifting through the i-InP layer, as will be discussed later. To further enhance the transport of electrons towards the i-InP layer, the p-InGaAs layer has a graded doping profile (from $10^{17}$ cm$^{-3}$ at top to $10^{19}$ cm$^{-3}$ at bottom). This induces a quasi-field for electrons to drift instead of diffusing through this layer.
The i-InP layer in the UTC-PD is used both as the collector for electrons and as the passive WG layer. The thickness of the i-InP layer (300 nm) allows for single-mode strip WGs and is compatible with other photonic components in the IMOS platform [5]. As can be seen from Fig. 1(b) and (c), the input InP WG is connected to the UTC-PD through a simple butt-joint coupling. Over 97% of the WG mode is coupled to the fundamental mode and the first order mode in the PD sections according to simulations using a mode solver. The transit time for electrons to drift through the i-InP collector layer (τ_c) is 1 ps, assuming a electron drift velocity of $3 \times 10^7$ cm/s in the regime of electron velocity overshoot. In fact, the electric field through such a thin intrinsic layer is usually above the velocity overshoot regime even with zero bias. In this case, the value of τ_c will increase to 3 ps assuming a saturation velocity of $1 \times 10^7$ [18]. For these two cases, the transit-time limited 3dB bandwidths ($f_{tr}$) are 123 GHz and 85 GHz, respectively, using [6, 19]

$$f_{tr} = \frac{1}{2\pi(\tau_d + 0.29\tau_c)},$$

(1)

where $\tau_d$ is the transit time through the absorption layer as introduced in the last paragraph. The junction capacitance of a $3 \times 10\ \mu m^2$ can be estimated to be about 11 fF using a parallel-plate model. By assuming a total resistance of 100 Ω (including the load), the RC limited bandwidth ($f_{RC}$) will be around 145 GHz. This value of $f_{RC}$ will be lower in the real case, as the probing pads will introduce extra RC terms. To suppress the parasitic capacitance, the ground-signal-
ground (GSG) probing pads are deposited on the surface of a polyimide layer (Fig. 1(d)). Also, the InP membrane below the GSG pads is etched.

NiGeAu is used as the n-contact of the PD due to its low contact resistance. Metal spiking is a well-known problem for annealed NiGeAu contacts, which is particularly critical for the membrane PD, as the n-contacts are located on top of the device. In order to reduce the spiking and the associated optical loss, while maintaining a low contact resistance, a new contact scheme incorporating a thin layer (20 nm) of Au for the annealed NiGeAu n-contact, followed by a thickening metal layer (TiAu) is used [20]. The p-contacts made of TiPtAu are formed at the bottom side of the device, thanks to the use of the double-sided processing. This leads to a low resistance at the p-side, which is crucial for getting a high bandwidth. First, due to the gradient doping profile in the p-InGaAs layer, low contact resistance can only be obtained if the metals are evaporated on the highly doped surface at the bottom side. Moreover, in order to avoid excessive series resistance resulting from the thin p-InGaAs layer, p-contacts should be placed very close the optical mode. This is only possible when they are formed at the lower side of the PD mesa, considering the practicality of a lift-off process used for the metal deposition. The optimal spacing between the p-contacts in a 3 μm wide PD is about 2.5 μm, which is a trade-off between the metal optical loss and the resistance from the p-side [21].

It should be noted that we have not introduced diffusion-blocking layers or band-smoothing layers in this proof-of-concept design, mainly for reasons of processing simplicity. This will lead to limited efficiency and bandwidth at low bias voltage. These effects will be studied in the experimental section of this paper.

The fabrication flow of the double-sided processed UTC-PD is shown in Fig. 2. The InP wafer in Fig. 2(a) is grown using Metal-Organic Chemical Vapor Deposition (MOCVD). In order to suppress the diffusion of Zn (the dopant of p-InGaAs) into the i-InP layer, the epitaxy is grown from n-side to p-side. The fabrication starts from the p-side of the InP wafer. In Fig. 2(b) p-contacts are first formed using electron beam evaporation and a lift-off process, followed by the removal of InGaAs outside of the PD area using a selective etching solution of H₂SO₄/H₂O₂/H₂O. Alignment-markers are etched through the epitaxial layers using CH₄/H₂-based reactive ion etching (RIE). After the processing from the p-side, the InP wafer is bonded onto a SiO₂/Si carrier wafer [22]. A 2 μm thick BCB layer is used for the bonding to accommodate the topography of the chip and achieve a high bonding yield. The total thickness of SiO₂ and BCB is around 2.4 μm, which is sufficient to prevent the optical mode in the InP membrane from coupling into the Si substrate. After the bonding, the flipped sample is immersed in a HCl/H₂O solution for removing the InP substrate (Fig. 2(d)). In Fig. 2(e), the sample is processed from the n-side. NiGeAu-based n-contacts are deposited, followed by a 380 °C annealing for 15 seconds. The n-InP layer is etched with a selective etching solution of HCl/H₃PO₄, using the n-contact as a mask. An InGaAsP layer (not shown in the figure) between n-InP and i-InP is used as an etch-stop layer. Afterwards, electron beam lithography (EBL) is used with a C₆₀/ZEP resist [16] to pattern the PD mesa, passive WGs, and surface grating couplers, followed by a dry etching process using RIE [17]. The width of the PD mesa is designed slightly larger than that of the n-contact, so that the rough edges of the metals will not transfer to the etched PD mesa. In Fig. 2(f) a polyimide layer is spun on the sample for passivation and planarization, followed by a dry etch to open the n-contact. Another etch of polyimide followed by etching the p-InGaAs is performed to open the p-contact. A reflow process is used to created a sloped opening in the polyimide layer. Finally, a 350 nm thick TiAu layer is evaporated to form the GSG transmission lines and probing pads.
3. Experimental results

3.1. Static measurements

In order to determine the responsivity of the UTC-PD, the coupling loss from fiber to WG is first determined using standard WG-loss measurement on reference WG arrays fabricated on the same chip [17], and calibrated out. The responsivity of the $3 \times 10 \, \mu m^2$ UTC-PD measured at -4 V is plotted as a function of wavelength in Fig. 3(a), showing a value of 0.7 A/W at 1.55 $\mu$m, corresponding to a quantum efficiency of 56%. The responsivity is above 0.6 A/W over the entire C-band. The responsivity is limited primarily by the loss from the n-contact on top of the PD mesa. As mentioned in Section 2, the n-contact is specially designed with a 20 nm thin layer of Au to reduce the spiking effect and thus giving low optical loss. In order to evaluate the loss from the n-contact, a propagation loss coefficient of 1.16 dB/$\mu$m is measured from testing WGs on this chip using an approach introduced in our previous work [23]. In that work, the loss coefficients of the standard NiGeAu n-contact (with a 300 nm thick Au layer) measured before and after annealing are 0.93 and 2.91 dB/$\mu$m, respectively. Comparison of these results indicates that the optimized n-contact used here is indeed important for loss reduction. A GeAg-based n-contact proposed in [23] shows an even lower loss coefficient (0.56 dB/$\mu$m) and should therefore allow for a higher responsivity. The dark current measured for the $3 \times 10 \, \mu m^2$ UTC-PD is plotted in the inset of Fig. 3(a), showing a value of 153 nA at -4 V, which is relatively high compared to state-of-the-art InP-based PDs [6]. This can be related to the remaining spiking of the NiGeAu-based n-contact on top of the membrane PD, which can lead to impurity trap levels and thereby enhance the generation-recombination current as well as the tunneling current. The exponential behavior of the measured dark current suggests that the tunneling current might be dominant here. The dark current may be reduced by using the spiking-free n-contact based on GeAg [23].

Fig. 2. Fabrication flow of the double-sided processed UTC-PD. (a) InP layer-stack; (b) p-contact formation, p-InGaAs etch, and alignment-marker etch; (c) bonding to Si wafer using BCB; (d) InP substrate etch; (e) n-contact formation, PD mesa etch, and passive WG/grating etch; (f) polyimide planarization, p-contact opening, and final metalization for GSG pads.
Fig. 3. (a) Responsivity versus wavelength of a $3 \times 10 \, \mu\text{m}^2$ UTC-PD measured at -4 V; the inset shows the measured dark current versus bias voltage. (b) Responsivity as a function of the length of the UTC-PDs. Both measurement data and simulation results are shown.

The responsivities measured for several UTC-PDs are plotted in Fig. 3(b) as a function of the PD length. The simulated responsivity obtained by using a full-vectorial eigenmode expansion method is also indicated. The scattering of the measurement data is mainly due to processing non-uniformity. In particular, the BCB layer used for bonding usually has a non-uniform thickness over the chip. This will lead to uncertainty of the calibrated value of the coupling loss from local gratings, which is sensitive to the thickness variation of the dielectric layer between the membrane and silicon. This variation in the grating coupling loss can be eliminated using a novel metal grating coupler, in which the coupling efficiency is independent of the BCB layer thickness [24]. In the present chip, the reference WG arrays for measuring the local coupling efficiency are placed close to each PD to obtain a reasonable accuracy for the responsivity calculation. It can be seen that the measured data show a reasonable agreement with the trend predicted by the simulation. The responsivity saturates at a length of only $10 \, \mu\text{m}$, thanks to the strong optical confinement in the p-InGaAs layer and to the butt-coupling scheme. For these sufficiently long PDs, the increasing responsivity at a shorter wavelength shown in Fig. 3(a) can be attributed to a larger ratio of the InGaAs absorption coefficient over the metal loss coefficient.

The measured photocurrents for the $3 \times 10 \, \mu\text{m}^2$ UTC-PD at different bias voltages are plotted as a function of the optical input power in Fig. 4(a), showing good linearity up to 3 mA photocurrent. Continuous-wave (CW) measurements with a higher optical power are performed on other PDs from the same chip, showing that thermal failure typically occurs around 3 mA photocurrent. This is primarily due to the low thermal conductivity of the BCB layer, preventing efficient heat dissipation in the Si substrate. Nevertheless, this current level is sufficient for a large range of applications including optical interconnects and telecommunications. For microwave applications, advanced heat dissipation techniques, such as flip-chip bonding to AlN or diamond submount are required in order to reach higher output power levels [13].

It can be also seen from Fig. 4(a) that the responsivity of the UTC-PD has a strong dependence on the applied bias. It is known that partial depletion of the intrinsic layer in PDs at low bias can cause long diffusion time and carrier recombination, which will result in low quantum efficiency. Hence, we first look into the i-InP layer in our UTC-PDs. The thickness of this layer is only 300 nm, and its doping level is around $1 \times 10^{16} \, \text{cm}^{-3}$ measured with an electrochem-
Fig. 4. (a) Bias dependent photocurrent versus optical power coupled to the UTC-PD. (b) Simulated electric field distribution in the vertical direction of the UTC-PD at zero bias and 1 mA photocurrent. (c) Simulated conduction band discontinuity at the InGaAs/InP interface.

ical C-V profiler. The electrical field distribution in the vertical direction of the UTC-PD is simulated using COMSOL, showing a fully depleted i-InP layer even at zero bias (Fig. 4(b)). The electric field strength is sufficient for electrons to drift through this layer. Hence, the low efficiency at low bias should not result from the diffusion in the i-InP layer. Second, the electron diffusion in the un-depleted p-InGaAs layer could be relevant. Particularly, as there is no diffusion-blocking layer in our UTC-PDs, electrons can diffuse to the p-contact resulting in a drop of the quantum efficiency. The graded doping profile in the p-InGaAs layer results in a quasi-field of several kV/cm, which is just sufficient for a drift motion of electrons in the direction towards the i-InP layer [18]. This is useful for an enhancement of the quantum efficiency in UTC-PDs without diffusion-blocking layers [14]. However, the barrier at the InGaAs/InP interface due to the conduction band discontinuity will cause electron trapping and current blocking. Trapped electrons may recombine there or diffuse to the p-contact, resulting in low quantum efficiency. As can be seen in Fig. 4(c), by increasing the bias voltage to -4 V, the barrier becomes narrower, and its height relative to the conduction band in the p-InGaAs away from the interface is reduced. This will lead to an increased emission rate of electrons through the barrier, because of the enhancements in both thermionic emission (TE) and field emission (FE). Eventually, the emission rate will saturate due to the finite electron drift velocity. As a result, the quantum efficiency and the corresponding responsivity increase as the applied bias increases, and saturate at a certain voltage (-3 to -4 V, in our case). To obtain a lower operating voltage, band-smoothing layers with compositional grading can be used at the InGaAs/InP interface [19].
3.2. Small-signal measurements

To determine the 3 dB opto-electrical bandwidth of the UTC-PD, small-signal measurements are performed using a commercial Agilent 67 GHz lightwave component analyzer (LCA) with a 50 Ω load. The average optical output power of the LCA is set to be 2 dBm, corresponding to an input optical power to the PD of about -4 dBm assuming a typical coupling loss of 6 dB from fiber to PD. The normalized frequency response \( S_{21} \) curves of the \( 3 \times 10 \ \mu \text{m}^2 \) UTC-PD measured at various bias voltages from 0 V to -4 V are shown in Fig. 5. Clear bias dependence in the response can be seen. According to the simulation result in Section 3.1, the i-InP layer is fully depleted at zero bias. Hence, the bias dependence is not due to a changing depletion depth. As carrier trapping can also lead to a large contribution to the total transit time [19], the fast roll-off at low frequency measured at 0 V and -1 V might result from the conduction band discontinuity at the InGaAs/InP interface. As discussed previously, the emission rate of electrons through the barrier increases at higher bias, which will also lead to an improved frequency response. This can be observed in our measured response curves: at -2 V the 3 dB bandwidth is increased to 15 GHz, while after -3 V it reaches about 67 GHz. Finally, at -4 V the 3 dB bandwidth exceeds 67 GHz, which is the upper limit frequency of the LCA.

![Fig. 5. Normalized frequency response \( S_{21} \) of a \( 3 \times 10 \ \mu \text{m}^2 \) UTC-PD measured at different bias voltages.](image)

To estimate the electrical circuit component values of the UTC-PD, \( S_{22} \) small-signal reflection coefficients are measured. Figure 6(a) shows the real part and the imaginary part of \( S_{22} \) measured for the \( 3 \times 10 \ \mu \text{m}^2 \) UTC-PD. A simple equivalent circuit model (Fig. 6(b)) is used to extract the junction capacitance \( C_j \) and the series resistance \( R_s \) of the UTC-PD. To calibrate out the S-parameters of the GSG pads, open and short de-embedding structures from the same chip are measured. The results can be used in a 2-port S-parameter component (S2P) in the circuit model. Commercial Agilent Advanced Design System (ADS) software is used to optimize \( C_j \) and \( R_s \) to fit to the measurement results. It can be seen in Fig. 6(a) that good matches are obtained for both the real and the imaginary part of the \( S_{22} \) parameter. Figure 6(c) shows the values of \( C_j \) determined from this model at different bias voltages, giving 14.2 fF at -4 V for the \( 3 \times 10 \ \mu \text{m}^2 \) UTC-PD. The very limited decrease of \( C_j \) from 0 V to -4 V indicates that the i-InP layer is fully depleted for all the bias conditions, which agrees with the electric field simulation result in Section 3.1. \( C_j \) of several UTC-PDs are plotted in Fig. 6(d) as a function the junction area. The data obtained from measurements basically follows the trend calculated...
using a simple model for a parallel-plate capacitor. Capacitances as low as 5 fF are obtained for $3 \times 5 \, \mu m^2$ UTC-PDs. The $R_s$ values extracted from this model show a large scattering (Fig. 6(e)). Since the PD is not connected to an on-chip 50 Ω termination resistor, $R_s$ has to be extracted out of a very large total impedance, which is difficult to do with a good accuracy. This particularly applies to PDs with small $C_j$. Most 3 µm wide PDs show a $R_s$ value below 50 Ω, thanks to the optimized design and to the use of the double-sided processing. 5 µm wide PDs usually give a larger $R_s$ value. As the holes are generated mostly at the center of the p-InGaAs layer, wider PDs can have higher series resistance due to the longer distance from the center to the p-contact.

![Graph](image)

Fig. 6. (a) Real part and imaginary part of the reflection coefficient ($S_{22}$) of a $3 \times 10 \, \mu m^2$ UTC-PD. Results from measurements and modeling are both shown. (b) The equivalent circuit model of the UTC-PD. (c) Bias dependence of the junction capacitance ($C_j$) extracted from the model. (d) $C_j$ of different UTC-PDs as a function of the junction area. (e) Series resistance ($R_s$) of different UTC-PDs versus their length.

3.3. Large-signal measurements

To test the performance of the UTC-PD in communication systems, on-off keying (OOK) data transmission measurements are performed. A commercial bit pattern generator (BPG) is used to generate $2^{11}-1$ long non-return to zero (NRZ) pseudo random bit sequence (RPBS) data pat-
terns. The output RF signal is used to drive a commercial LiNbO$_3$ modulator. The practical upper limit of the data rate for this commercial setup has been tested to be 54 Gb/s. A commercial Agilent 50 GHz sampling oscilloscope is used to measure the output. The optical eye diagrams output from the modulator at 40 Gb/s and 54 Gb/s are shown as references in Figs. 7(a) and 7(b), respectively. The modulated light at 1.55 $\mu$m is then coupled to the UTC-PD. The average optical power coupled to the PD is estimated to be -1.5 dBm. The electrical output signal from the $3 \times 10 \, \mu$m$^2$ UTC-PD is measured. Open eye diagrams obtained at 40 Gb/s and 54 Gb/s with a bias voltage of -4 V are shown in Figs. 7(c) and 7(d), respectively. The rise and fall curves follow the optical pattern defined by the modulator. This high data rate demonstration promises high capacity applications in optical interconnects and telecommunications.

![Fig. 7. Reference optical eye diagrams output from the commercial modulator at (a) 40 Gb/s and (b) 54 Gb/s, along with the associated electrical eye diagrams output from the $3 \times 10 \, \mu$m$^2$ UTC-PD at (c) 40 Gb/s and (d) 54 Gb/s.](image)

4. Conclusion

High bandwidth WG-coupled UTC-PDs integrated on an InP-membrane-on-silicon platform are reported. A double-sided processing scheme is used to optimize the performance of the UTC-PD. Thanks to the butt-coupling and to the strong confinement of light in the membrane, compact and low capacitance UTC-PDs can be realized. A $3 \times 10 \, \mu$m$^2$ UTC-PD exhibits a 3 dB bandwidth beyond 67 GHz, which is the highest value reported in heterogeneously integrated PDs on silicon. The responsivity is 0.7 A/W at 1.55 $\mu$m, and the dark current is 153 nA at -4 V. Open eye diagrams are demonstrated up to 54 Gb/s. These results promise high speed applications in optical interconnects and telecommunications.
Acknowledgments

This work is supported by the ERC project NOLIMITS. The authors acknowledge Nanolab@TU/e for the use of the cleanroom facilities.