Design and characterization of a 30-GHz bandwidth low-power silicon traveling-wave modulator

Ran Ding a,⁎, Yang Liu a, Qi Li b, Yisu Yang a, Yangjin Ma a, Kishore Padmaraju b, Andy Eu-Jin Lim c, Guo-Qiang Lo c, Keren Bergman b, Tom Baehr-Jones a, Michael Hochberg a,c,d

⁎ Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA
b Department of Electrical Engineering, Columbia University, 500 West 120th Street, New York, New York, USA
c Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR), Singapore
d Department of Electrical and Computer Engineering, University of Singapore, Singapore

A R T I C L E   I N F O

Article history:
Received 22 August 2013
Received in revised form 28 January 2014
Accepted 29 January 2014
Available online 7 February 2014

Keywords:
Silicon photonics
Waveguide modulators
Traveling-wave modulators
Photonics integrated circuits

A B S T R A C T

We present the design and characterization of a silicon PN junction traveling-wave Mach–Zehnder modulator near 1550 nm wavelength. The device shows 30 GHz bandwidth at 1 V reverse bias, with a 2.7 V-cm VπLπ and accordingly a 9-V small-signal Vπ. The insertion loss of the phase shifter is 3.6 dB ± 0.4 dB. The device performance metrics in combination show significant improvement compared to the state-of-the-art in the sense that lower phase shifter loss and higher bandwidth are achieved for the same Vπ or vice versa. We demonstrated low modulation power of 640-fJ/bit at 40 Gb/s with a 1.6-Vpp differential-drive and 0-V DC bias, raising the prospect of direct compatibility with CMOS drive-voltages. Critical design tradeoffs are analyzed and design models are validated with measurement results. We proposed a new figure-of-merit (FOM) VπLπRπCp as the junction design merit for high-speed traveling-wave modulators, and utilized 6 implants to achieve an optimal FOM with lower insertion loss. Several key RF design issues are addressed for the first time using simulation and measurement results. In particular, we discussed bandwidth extension using mismatched termination and closely matched experimental results. A bandwidth-limiting RF multi-mode behavior is noted, which also exists in other results in the literature; we suggested a widely applicable design remedy.

1. Introduction

Silicon optical modulators [1] are critical for data communication related applications [2–5] in silicon photonics. Over the past decade, significant progress has been made in this area, but achieving efficient high-speed modulation in silicon still proves to be challenging, mainly due to the weak electro-optic (EO) effects available in this material [5,6].

The fundamental and key modulator device metrics include insertion loss, device bandwidth, and EO modulation efficiency (for Mach–Zehnder modulators the efficiency is characterized by Vπ). In addition, optical bandwidth, device footprint, temperature sensitivity, fabrication error tolerance and CMOS compatibility are also of great importance for practical designs, design scalability and possibility of CMOS monolithic integration [7,8].

A majority of the high-speed demonstrations thus far have been based on reverse-biased silicon PN junctions. Among these results, high-speed resonator modulators are promising in achieving ultra-low modulation power consumption and compact device footprint. Recently, Li et al. demonstrated a 40 Gb/s 1 V-drive ring modulator [9], although a few issues remain to be fully addressed, such as limited optical bandwidth, and consequently necessary thermal drift stabilization as well as operating wavelength alignment between devices.

The other main category of carrier-depletion PN junction modulators is traveling-wave Mach–Zehnder (TWMZ) modulators. Although in academic demonstrations imbalanced Mach–Zehnder (MZ) modulators are often used (mostly for the convenience of testing), balanced MZ modulators are the true practical devices and have the key advantage of being temperature insensitive, thus do not require active thermal stabilization. Traveling-wave design enables the driving of a long phase-shifter at high speed, therefore can yield low voltage modulators.

40–60 Gb/s channel speed is a logical next step from existing 25–28 Gb/s data rates. The device we present here targets at
applications at these speeds or equivalently a device bandwidth of approximately 30 GHz [9–11]. In recent TWMZ results [10–15] however, it was often found that a similar bandwidth was only achieved with very short devices (∼1 mm or less) as well as high bias voltages (frequently 3–5 V), both of which limit the modulation efficiency. High $V_π$ and associated high drive voltages increase power consumption and make the devices less compatible with advanced CMOS, which is usually constrained by low-breakdown voltages. For long devices [14,15], in addition to smaller bandwidth, they were demonstrated with high insertion loss on the phase shifter making it difficult to fit in a practical system link budget. In summary, further device improvements remain to be made for high-speed low voltage modulators with low loss.

In this paper, we present a 3-mm long, 30-GHz bandwidth differential-drive silicon TWMZ modulator based on a lateral PN junction with low reverse bias. The phase shifter $V_πL_π$ is 2.7 V-cm (a small-signal $V_π$ of 9 V) and the insertion loss on the phase shifter is only 3.6 dB. At similar bandwidth this device shows the lowest $V_π$ and lowest drive voltage requirement due to differential-drive, or at similar $V_π$ it shows the highest bandwidth as well as the lowest insertion loss on the active phase shifter [10–15]. The combined device metrics show significant improvement compared to the state of the art. The main part of the paper is organized as the following. Section 2 describes the fabrication process, identifies key design tradeoffs and presents design details of the device. Section 3 presents the measurement results, compares them to discussions in this paper. The cross-section of the TWMZ shown in Fig. 1 is considered as a PN junction loaded transmission line and its equivalent circuit model is schematized in Fig. 2. $R_{pd}(f)$ is the frequency-dependent metal skin resistance in $Ω$/$m$, and has a $f$ dependence in principle, $C_{tl}$ and $L_{tl}$ are the capacitance and inductance between the metal traces in the units of $F$/m and $H$/m respectively. $C_{pm}$ is the PN junction capacitance in $F$/m; the total amount of silicon series resistance from the electrodes to the edges of the junction depletion region is captured in $R_{pm}$ in $Ω$-$m$. We further define the junction intrinsic RC bandwidth as $f_{rc} = 1/(2πR_{pm}C_{pm})$ and approximate the device impedance as $Z_{dev} = \sqrt{L_{tl}/(C_{tl}+C_{pm})}$, which is accurate when the frequency is well below the intrinsic RC bandwidth, i.e., $f \ll f_{rc}$. It is worth noting that, due to the use the high-resistivity substrate, the substrate conductance due to transverse current flow can be neglected in the frequency range of interest [29].

The bandwidth of a TWMZ modulator is mostly determined by the RF loss due to $R_{pm}$, if RF and optical velocities are closely matched. To make this clear we can look at the overall RF field loss coefficient (in the unit of Neper/m), which can be expressed as [19]

$$\alpha = \alpha_{metal} + \alpha_{silicon}$$

$$\frac{1}{2} R_{pd}(f) + \frac{2\pi f^2 R_{pm} C_{pm} Z_{dev}}{1 + (f/f_{rc})^2}$$

$$\frac{1}{2} R_{pd}(f) + \frac{\pi f^2 C_{pm} Z_{dev}}{f_{rc}(1 + (f/f_{rc})^2)}$$

Where $\alpha_{metal}$ and $\alpha_{silicon}$ are the loss due to metal series resistance and lateral silicon resistance respectively. Let us refer to the first term as $R_{l0}$ loss and the second term as $R_{pm}$ loss. At high frequencies, the second term usually dominates because of its $f^2$ dependence, whereas $R_{pd}(f)$ has a $\sqrt{f}$ dependence. Incidentally, this can be seen clearly in Fig. 4(g), a simulation plot of the actual device under discussion.

For the moment, let us assume perfect velocity match and neglect other non-ideal RF effects (such as reflection, multi-modal behavior etc.), a straightforward relation between EO 3dB bandwidth $f_{EO}$ and achievable device length $L_{dev}$ can be derived as [20]:

$$1 - e^{-\alpha f_{EO} L_{dev}} = \frac{1}{\sqrt{2}} \Rightarrow \alpha f_{EO} L_{dev} = 0.74 \text{Neper} = 6.4 \text{ dB}$$

2 μm buried oxide layer and 750 $Ω$-cm high resistive silicon substrate. The silicon slab thickness was 90 nm and ridge waveguide width was 500 nm. The top metal Aluminum was used for the traveling-wave electrodes, and it was 2 μm thick, mostly situated above dielectric materials in the back-end stack. Other metal and dielectric material properties and thicknesses as well as fabrication steps were identical as they were in [16].

2.2. Overall device design considerations

In this section we briefly review the TWMZ design model and present useful design tradeoff relations to facilitate further discussions in this paper. The cross-section of the TWMZ shown in Fig. 1 is considered as a PN junction loaded transmission line and its equivalent circuit is schematized in Fig. 2. $R_{pd}(f)$ is the frequency-dependent metal skin resistance in $Ω$/$m$, and has a $f$ dependence in principle, $C_{tl}$ and $L_{tl}$ are the capacitance and inductance between the metal traces in the units of $F$/m and $H$/m respectively. $C_{pm}$ is the PN junction capacitance in $F$/m; the total amount of silicon series resistance from the electrodes to the edges of the junction depletion region is captured in $R_{pm}$ in $Ω$-$m$. We further define the junction intrinsic RC bandwidth as $f_{rc} = 1/(2πR_{pm}C_{pm})$ and approximate the device impedance as $Z_{dev} = \sqrt{L_{tl}/(C_{tl}+C_{pm})}$, which is accurate when the frequency is well below the intrinsic RC bandwidth, i.e., $f \ll f_{rc}$. It is worth noting that, due to the use the high-resistivity substrate, the substrate conductance due to transverse current flow can be neglected in the frequency range of interest [29].

The bandwidth of a TWMZ modulator is mostly determined by the RF loss due to $R_{pm}$, if RF and optical velocities are closely matched. To make this clear we can look at the overall RF field loss coefficient (in the unit of Neper/m), which can be expressed as [19]

$$\alpha = \alpha_{metal} + \alpha_{silicon}$$

$$\frac{1}{2} R_{pd}(f) + \frac{2\pi f^2 R_{pm} C_{pm} Z_{dev}}{1 + (f/f_{rc})^2}$$

$$\frac{1}{2} R_{pd}(f) + \frac{\pi f^2 C_{pm} Z_{dev}}{f_{rc}(1 + (f/f_{rc})^2)}$$

Where $\alpha_{metal}$ and $\alpha_{silicon}$ are the loss due to metal series resistance and lateral silicon resistance respectively. Let us refer to the first term as $R_{l0}$ loss and the second term as $R_{pm}$ loss. At high frequencies, the second term usually dominates because of its $f^2$ dependence, whereas $R_{pd}(f)$ has a $\sqrt{f}$ dependence. Incidentally, this can be seen clearly in Fig. 4(g), a simulation plot of the actual device under discussion.

For the moment, let us assume perfect velocity match and neglect other non-ideal RF effects (such as reflection, multi-modal behavior etc.), a straightforward relation between EO 3dB bandwidth $f_{EO}$ and achievable device length $L_{dev}$ can be derived as [20]:

$$1 - e^{-\alpha f_{EO} L_{dev}} = \frac{1}{\sqrt{2}} \Rightarrow \alpha f_{EO} L_{dev} = 0.74 \text{Neper} = 6.4 \text{ dB}$$
This is the “6 dB” rule-of-thumb frequently referred to in publications about TWMZ, i.e., the RF 6 dB (more accurately 6.4 dB) bandwidth is close to the EO response 3 dB bandwidth. For the sake of simplicity let us further assume the desired bandwidth \(f_{\text{EO,3dB}}\) is sufficiently low compared to the intrinsic RC bandwidth of the junction so that \(\frac{f_{\text{EO,3dB}}}{f_{\text{RC}}} < 1\). Usually the TWMZ design is in the “low loss” regime of the transmission line model where this assumption naturally stands. Inserting Eq. 1 into Eq. 2, and we get

\[
L_{\text{dev}} = \frac{0.74}{(1/2)R_{\text{metal}}f_{\text{EO,3dB}}Z_{\text{dev}}} + 2\pi^2 f_{\text{EO,3dB}}^2 Z_{\text{dev}}
\]

(3a)

\[
= \frac{0.74}{2\pi^2 f_{\text{EO,3dB}}^2 Z_{\text{dev}} R_{pn} C_{pn}}
\]

(3b)

The approximation before arriving at Eq. 3a is based on that \(R_{pn}\) loss is much larger than \(R_{Z}\) loss at frequencies near \(f_{\text{EO,3dB}}\).

A few scaling trends can be derived from Eqs. 3a and 3b, which quantify the tradeoffs in the seemingly complicated design space:

(I) It becomes apparent that getting low \(V_z\) at high frequency is increasingly difficult, since \(L_{\text{dev}} \propto 1/f_{\text{EO,3dB}}^2\), consequently \(V_z \propto \sqrt{V_{\text{dev}}/L_{\text{dev}}} \propto f_{\text{EO,3dB}}^{-1}\).

(II) In general, to design for certain impedance and bandwidth the achievable device length \(L_{\text{dev}}\) and \(V_z\) have the following scaling trends with respect to PN junction parameters:

\[
L_{\text{dev}} \propto 1/(R_{pn} C_{pn}^2), \quad V_x \propto (V_x L_x) R_{pn} C_{pn}^2
\]

(4)

Obviously, a reduction in \(R_{pn}\) is directly reflected in a reduction in \(V_z\) (for certain bandwidth design target). Highlighted in Eq. 4 is another key point. In an attempt to improve junction modulation efficiency (reducing \(V_z L_x\)) the factor \((V_x L_x) R_{pn} C_{pn}^2\) should be evaluated. A reduction in \(V_x L_x\) is likely to come at the cost of increased \(C_{pn}\). An increase in \((V_x L_x) R_{pn} C_{pn}\) implies that the reduction in \(V_x L_x\) does not ultimately lead to the reduction of the device \(V_z\). Therefore, a low doped, low capacitance density junction could be a more advantageous for traveling-wave design, which is the case of the design reported in this paper.

(III) Designing at lower device impedance \(Z_{\text{dev}}\) is advantageous, because \(L_{\text{dev}} \propto 1/Z_{\text{dev}}, V_z \propto Z_{\text{dev}}\), holding junction parameters constant. Note that the impedance \(Z_{\text{dev}}\) is not the termination impedance but the RF impedance of device itself. The effect of termination impedance on device performance will be addressed in Section 3.2. Assuming the device is in fact terminated with \(Z_{\text{dev}}\), we can then conveniently evaluate the drawback of using low impedance design, which is mainly some loss of the incoming drive voltage, if the driver is at impedance \(Z_0\) that is higher than \(Z_{\text{dev}}\). The ratio of voltage dropped on the device \((V_0)\) versus the driver output voltage \((V_D)\) is \(V_{\text{dev}}/V_{\text{D}} = 2Z_{\text{dev}}/Z_0\). For example, this voltage-intake factor is 67% for a 25 \(\Omega\) device terminated with 25 \(\Omega\) driven by a 50 \(\Omega\) driver. However, the achievable device length at 25 \(\Omega\) is approximately twice as it would be for a 50 \(\Omega\) design based on Eqs. 3a,b, therefore overall the 25 \(\Omega\) design would be more advantageous. The device length doubling is of course a rough estimate, due to the omission of metal loss.

It is also worth noting that single-drive push-pull modulators [21] demonstrated an elegant way to ease the design of 50 \(\Omega\) device impedance while conveniently maintaining matched RF and optical velocities. Due to halving the junction capacitive loading to the transmission line electrode, the bandwidth-limiting \(R_{pn}\) loss is approximately halved (according to Eq. 1b, keeping \(f_{\pi}\) constant) and therefore very long devices are possible at high speeds. However, doubled device length combined with single-drive operation yields approximately the same drive voltage requirement as it is for a conventional differential-drive TWMZ device, while doubling the device length implies high optical insertion loss, which is a drawback that needs to be weighed against the benefits associated with such single-drive designs.

### 2.3. PN junction design

The waveguide PN junction phase shifter is the core component of an MZ modulator; its metrics largely determine the achievable overall device performance, as shown in Eq. 3a,b and Eq. 4. In addition to providing a low \((V_x L_x) R_{pn} C_{pn}^2\) factor, the PN junction design also needs to achieve low optical insertion loss.

We chose lightly doped P and N to form the junction in the waveguide with the P side average doping concentration being around \(5 \times 10^{17}/\text{cm}^3\) and N side being close to \(3 \times 10^{18}/\text{cm}^3\).
The junction line was designed to be at the center of the waveguide, with no intentionally added intrinsic region width. This low-doped PN junction is in favor of the efficiency versus loss tradeoff, characterized by the figure-of-merit $F_{dB-V}$ (in dB/V) as defined in [22]. Also, it helped achieve a low $V_{\pi L_{\pi}} R_{pn C_{pn}}$ by maintaining a low $C_{pn}$ and competitive $V_{\pi L_{\pi}}$.

To achieve a low $R_{pn}$ with low loss, we employed a 3-level side-doping configuration as illustrated in Fig. 1, with carefully chosen doping density and doping profile. We selected P+ and N+ doping density of $2 \times 10^{18}$/cm$^3$ and $3 \times 10^{18}$/cm$^3$ respectively and optimized the doping profile. The onset of doping to the edge of the waveguide is defined as “clearance”. The clearance of P+ and N+ doping were 120 nm and 140 nm respectively. The P++ and N+++ doping were on the level of $1 \times 10^{20}$/cm$^3$ and their clearance were both 950 nm and the doped region extended laterally until they reached the electrode contact regions that were 3.95 μm away from the edge of the waveguide. A breakdown of optical loss and resistance of the various regions is presented in Table 1 below.

Based on implantation process modeling and device simulation [23], $C_{pn}$ was 280 fF/mm at 0 V and 220 fF/mm at −1 V bias, thus $f_{r, c}$ at −1 V bias reached 100 GHz. The simulated small-signal $V_{\pi L_{\pi}}$ was 1.7 V-cm at −1 V bias.

2.4. Traveling-wave electrode design

In light of the observation (III) in the end of Section 2.1, we continued to use a 33–37 Ω impedance design similar to [16] instead of designing at 50 Ω. This is also a convenient impedance to design with our junction capacitance $C_{pn}$ while achieving velocity matching condition and low RF loss. Another consideration was that if the device were properly terminated with matching impedance, this device would still provide an acceptable
Table 1

<table>
<thead>
<tr>
<th></th>
<th>N and P doped waveguide core</th>
<th>N and P doped slab region</th>
<th>N+ doped slab region</th>
<th>P+ doped slab region</th>
<th>N++ and P++ doped region</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1.5</td>
<td>1.6</td>
<td>1.1</td>
<td>2.5</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>IL</td>
<td>6.8</td>
<td>0.97</td>
<td>0.53</td>
<td>0.49</td>
<td>8.3</td>
<td></td>
</tr>
</tbody>
</table>

Notes: “R” denotes resistance in Ω-mm, “IL” denotes insertion loss in dB/cm. The IL reported for waveguide core includes the loss of N and P doped slab region due to simulation setting. IL numbers only include free carrier absorption loss at 0 V bias, and do not include intrinsic waveguide loss.

Fig. 5. (a) Optical spectra versus applied voltage (b) $V_{IL}$ versus applied voltage.

S11 ($<-10$ dB), when tested in 50 Ω environment or integrated with 50 Ω drivers.

However, we do not deem it a fundamental requirement to comply with 50 Ω standard RF interfaces, because eventually and especially for on-chip applications the TWMZ devices are likely be closely integrated with custom drivers, in which cases the short length (or even zero length) electrical connection as well as the availability of custom designed low impedance drivers would make it possible to freely choose any impedance as long as it advances the overall device and system performance. For example, numerous laser drivers have been designed at very low impedances [24] for efficient driving.

The device layout and microscopic photograph is shown in Fig. 3. We used a Ground-Signal (GS) coplanar transmission line electrode to drive each arm. The GS lines have 50 μm trace width and 5 μm gap. The length of the phase shifter is 3 mm. The separation between the two arms is 400 μm center-to-center. Metal tiling exist around the device to comply with density design rules, the clearance between metal tiling and the RF traces is 75 μm. The traces used the 2 μm thick aluminum layer (M2) in the process. The layout of the via-stack from M2 down to silicon was closely integrated with custom drivers, in which cases the short length (or even zero length) electrical connection as well as the availability of custom designed low impedance drivers would make it possible to freely choose any impedance as long as it advances the overall device and system performance. For example, numerous laser drivers have been designed at very low impedances [24] for efficient driving.

The detailed PN junction and side doping configurations were described in Section 2.3. It is worth noting that the phase shifter in the actual TWM2 only has ~90% of its length loaded with PN junction, due to striation as illustrated in Fig. 3(a). The striation in the wave propagation direction, as opposed to the generated in the metal traces toward the wave propagation direction, as opposed to flowing in the silicon.

At the driving end (the right side in Fig. 3(b)), a GSSGSG pad set was used to deliver differential driving signal to the device. A transition was made on chip from GSG to GS. At the termination end (the left side in Fig. 3(b)), a GSSGSGG pad set was used, offering four 50 Ω ports. All pad pitches are 100 μm. The G pads in the termination pad set are connected together by the lower level metal (M1), which is not very visible in the photograph. Each device arm was associated with two 50 Ω ports in parallel, allowing 25 Ω termination through a GSSGSGG RF probe and off-chip resistors. Other values of termination impedance would be less convenient, but is possible to implement by bonding the photonics chip to a high-speed printed circuit board (PCB) via short-length bond wires, and using surface-mount low parasitic resistors on the PCB. On the other hand, the GSSGSGG pad set allowed us to take the RF S-parameter as discussed in detail in Section 3.2.

Simulated RF characteristics of the transmission line electrodes with and without PN junction loading is presented in Fig. 4(a)-(f). First, the metal traces were simulated in HFSS [25], and then loaded line characteristics were calculated based on the circuit model in Fig. 1(b). The presented curves that involve PN junction, i.e. the dotted traces in Fig. 4(e) and (f) and the entire plot of Fig. 4(g), were calculated at 0 V bias condition and with the 90% junction loading factor taken into account. Near 30 GHz, $R_{jn}$ is 8.1 Ω/mm, $L_{jn}$ is 420 pH/mm, and $C_{jn}$ is 120 pF/mm. Loaded with PN junction, the impedance is 34 Ω and RF index is 3.78 at 30 GHz. The RF loss is presented in Fig. 4(g) with the total loss, and loss due to $R_{tl}$ and $R_{pj}$ plotted separately and it is clear that $R_{pj}$ loss is the dominating source of RF loss at high frequencies. In this plot Case 1 trace was obtained with simulated $R_{jn}$ and $C_{jn}$, i.e. 7.2 Ω-mm, and 280 fl/mm, whereas Case 2 was using 15 Ω-mm and 230 fl/mm to approximate the measured device. The increase in $R_{pj}$ in fabricated device is mainly attributed to the higher sheet resistance in the light P and N doping as well as higher contact resistance from metal to P++ doped silicon, which will be discussed in Section 3.

3. Measurement results and discussions

3.1. DC $V_{IL}$ and insertion loss

We first measured the optical transmission and DC performance of the device. The optical test setup used an Agilent 81980 A tunable laser and an 81636B detector to record the device transmission spectra. The intentional imbalance of the device was 100 μm and the free spectral range was about 5.7 nm. We tracked the null in the spectrum to generate phase shift versus applied voltage on one arm, from which $V_{IL}$ was calculated. The results are shown in Fig. 5. Incidentally, there are some ambiguity and
discrepancy about the definition of $V_d L_d$ in the literature. One way to report $V_d L_d$ is based on an actual $\pi$ phase shift [26,27]. This test protocol requires devices of various lengths to get a $V_d L_d$ versus $V_d$ (or $I_d$) curve and could be difficult for short devices (high voltages). Here we adopted an alternative approach: the $V_d L_d$ versus applied voltage relation is generated by the measurement of the phase shift $\Delta \phi_{dut}$ versus applied voltage $V_{applied}$ on one phase shifter of certain length $L_{dut}$, and simply $V_d L_d = \frac{\pi}{\Delta \phi_{dut}} V_{applied}$. The curve $V_d L_d$ versus $V_{applied}$ is basically $V_d L_d$ versus $V_d$. This approach was widely used in publications on silicon modulators, including several in the references list, for example [11], [13], [14], and [24].

At $-1$ V bias the measured phase shifter $V_d L_d$ was 2.7 V-cm. Further measurements over 5 different chips showed a good uniformity. Taking into account the 90% loading factor, the $V_d L_d$ of a fully doped waveguide phase shifter would be 2.43 V-cm, higher than the simulated 1.7 V-cm. The measured $C_{pn}$ was 230 fF/mm at 0 V and 190 fF/mm at $-1$ V bias for a fully doped waveguide phase shifter, lower than the simulated value 280 fF/mm at 0 V and 220 fF/mm at $-1$ V. We further measured silicon resistors with various doping types, doping concentrations and silicon thicknesses, and revealed that P++ or N++ doped silicon resistance agreed with simulation within 5%. However, P+ or N+ doped silicon showed $\sim 30\%$ higher resistance than simulation, and P or N doped resistance showed $\sim 50\%$ higher resistance than simulation. Combining the higher $V_d L_d$, lower $C_{pn}$ and higher resistance of P or N doped silicon resistance, we suspect the discrepancies could be due to: physically applied implantation dose being smaller than simulated, and possible incomplete ionization of the dopant. Other fabrication inaccuracies are harder to be verified or excluded.

The on and off chip coupling were through grating couplers. The device insertion loss was obtained by comparing the maximum transmission of the MZ spectrum to a grating coupler loop to de-embed the coupler insertion loss. Then the routing waveguide intrinsic loss was measured to be 0.38 dB loss. Considering the 90% loading coefficient, we can back calculate the free carrier absorption loss was 11.1 $\pm$ 1.5 dB/cm, whereas the simulated value is 8.3 dB/cm. The discrepancy could be excess loss due to P+ and N+ misalignment error (the 3σ of the overlay accuracy was a large fraction of the doping clearance) or a slight change in waveguide dimensions inducing change in optical mode overlap with P+ and N+ dopants. Incidentally, we measured PN junction doped waveguides that are without the side doping P+, N+, P++ and N+++. The measured free carrier loss was 6.3 dB/cm on average, which was very close to the simulated 6.8 dB/cm. This made it more likely the extra loss seen in the TWMZ phase shifters was due to P+, N+ doping etc.

3.2. Small-signal bandwidth measurement

The EO frequency response of the TWMZ was characterized using an Agilent 67 GHz Vector Network Analyzer (VNA), and an U2T XPDV3120R-VF-VP 70 GHz bandwidth photodetector. The EO S21 is presented in Fig. 6. This measurement was done through on-wafer probing with a 40 GHz rated Cascade ACP GSGSS probe driving and a 20 GHz rated Cascade Unity GSSGSS probe with off-chip resistors for 25 Ω termination. The VNA drove only one device arm at a time; the unused arm was properly terminated at both the driving and termination end with 50 Ω and 25 Ω respectively. This “full-termination” scheme closely resembles the actual operating condition of a differential drive modulator. The probes were not de-embedded from the frequency response and the RF S21 roll-off of the GSSG probe was 0.5 dB near the device bandwidth. Reverse bias was applied through the bias-tee in the VNA.

30 GHz bandwidth was achieved for both arms at 1 V reverse bias. The signal-to-noise ratio is better than 10 dB near the bandwidth. The same bandwidth was almost reached with 0 V bias (a long plateau in the EO S21 coincided with the $-3$ dB line between 23 GHz and 30 GHz). We observed two distinctive features of the EO S21: the peaking near 5 GHz and the notch near 17 GHz. These two features were consistently observed across many devices on both arms. We will explain them based on RF measurements.

The RF S-parameter measurements were done with the 40 GHz GSSGSS probe at the driving side with one GSG port connected to VNA Port 1 and the other GSG port attached to 50 Ω to terminate the unused arm. At the termination end, a 40 GHz GSSG probe was used with one GS port connected to VNA Port 2 and one GS port attached to 50 Ω. It presents 25 Ω to the modulator arm under test and receives RF through signal. We had to leave the termination end of the untested arm open due to the lack of a high speed GSSGSSG probe. The RF calibration included the cabling but could not include the probes, due to the lack of a GSG-to-GS calibration substrate. The uncalibrated probes were responsible for the fine-pitch ripples on the S-parameters traces, most visible on S11 due to their low magnitude. The measurement results are shown in Fig. 7(a).

First of all, we can use RF S11 and S21 to back calculate the RF index and RF loss. The RF index can be obtained by $n_f = c/(2L_{eff} \Delta f)$, where $c$ is the speed of light in vacuum, $L_{eff}$ is the effective RF length (it is the device length with small corrections to take into account of the input and termination wiring that has an RF index of roughly 2.5, $L_{eff}$ is estimated to be 3.3 mm based on simulation), and $\Delta f$ the spacing of S11 null locations (measured to be 12 GHz). Therefore RF index was calculated to be 3.8, and is close to the optical group index 3.9.

To calculate RF loss and RF 6.4 dB bandwidth, we should note that the measured RF S21 starts at $-3.9$ dB rather than close to 0 dB. This was due to the 50 Ω-driving, 25 Ω-termination testing configuration, which results in a 67% voltage-intake at low frequencies (see Section 2.1). In principle it should be $-3.5$ dB,
an approximate device impedance (i.e. $Z_{\text{term}}$) bias. In the case where the termination was exactly matched to bandwidth was almost exactly at the RF 6.4 dB bandwidth.

In the simulations, we chose three different termination scenarios and they are shown in Fig. 8. In these simulations, we assumed the GS-G structure was multi-modal. In more detail, when we were testing one device arm, i.e. driving one GS transmission line, we suggested a RF 6.4 dB bandwidth of 20 GHz at 0 V, in excellent agreement with measurement.

The difference between RF 6.4 dB bandwidth and EO S21 of 3 dB bandwidth is due to the mismatch between device impedance and termination impedance. This assessment can be evaluated analytically by expressing the RF voltage amplitude on the transmission line as forward and reverse propagating waves due to reflection then an overlap integral with the forward propagating optical wave would generate the EO response [20][29]. For the 33 Ω impedance device we simulated three different termination scenarios and they are shown in Fig. 8. In these simulations, we chose an approximate $f^2$ dependent RF loss function to have RF 6.4 dB bandwidth of 23 GHz to emulate measured performance at −1 V bias. In the case where the termination was exactly matched to device impedance (i.e. $Z_{\text{term}} = 33 \, \Omega$), as expected the EO 3 dB bandwidth was almost exactly at the RF 6.4 dB bandwidth.

For high impedance termination case (i.e. $Z_{\text{term}} = 50 \, \Omega$), the device bandwidth was very limited and the last crossing of the -3 dB line was only at 13 GHz, this is caused by the reflected wave with an undesired phase shift. On the other hand, for the low impedance termination case (i.e. $Z_{\text{term}} = 25 \, \Omega$), which is close to the measurement condition, the reflected wave enhanced the EO response. Near 5 GHz we observed peaking of similar magnitude as that in the measurement, the peaking effect is most pronounced at low frequency S11 nulls primarily due to less phase mismatch (and secondarily due to less RF loss). The device showed close to 30 GHz bandwidth. This is a close match to measurement results, validating our theory. This indicates an advantage of using termination impedance that is lower than the device impedance. As we discussed earlier, at low frequencies the voltage on the device can be evaluated by $V_{\text{term}}/V_0 = Z_{\text{term}}/(Z_{\text{term}} + Z_0)$ and a voltage-intake factor of 67% is obtained for 25 Ω termination versus 50 Ω termination. Considering a bandwidth extension ratio of 2.2 (30 GHz versus 13 GHz), the 25 Ω termination case appeared to be more advantageous overall.

Finally, we address the other distinctive signature in the EO S21: the notches in the spectrum. Due to its low magnitude, the notch in the EO S21 would not noticeably degrade eye-diagrams, as was suggested by our eye-diagram simulation comparisons between artificially smoothed EO S21 and measured EO S21. However, the notches are certainly detrimental features and could lead to serious problems in certain designs if we do not develop a good understanding of its root cause.

Observing the RF S21, it is evident the notches and steps in EO S21 occur at similar locations as the notches in RF S21. The most visible ones on EO S21 are the notch near 17 GHz and the step near 30 GHz. The magnitude of the notch near 17 GHz was 3–4 dB, which could mostly explain the corresponding notch in EO S21. To explain the notches in RF S21, we propose a hypothesis that the GS-SG structure is multi-modal. In more detail, when we were testing one device arm, i.e. driving one GS transmission line, we were effectively driving a GS-G structure. The G trace far away has a non-negligible RF interaction with the S trace. This GS-G structure supports two modes due to asymmetry (structure, dielectrics distribution and PN junction loading). A full structure HFSS simulation was challenging due to the large computation domain (long device length and large lateral span), but an analytical analysis could be formulated based on [30]. As a simplification for weakly coupled lines we assume the GS-G supports the 33–37 Ω mode ("main" mode) we originally designed between the S and the nearby G trace, and the GS-G also supports a stray mode that is mostly between the S and the far away G trace. We model the two modes as two independently propagating modes only connected at the beginning and end of the line.

![Fig. 7. (a) RF S-parameters measurement (b) circuit simulation to verify hypothesis of RF multi-modal behavior.](image)

![Fig. 8. Analytical modeling of EO response with respect to termination impedance.](image)
The combined line was driven by a 50 Ω source and terminated by 25 Ω. We used the TWMZ parameters described earlier for the main mode, and estimated the capacitance between the S trace and the far away G trace is $C_{stray} = 20 \text{ fF/mm}$, using two-wire transmission line equations [19]. To avoid complicated circuit models, we picked reasonable constant metal skin resistance as opposed to a frequency-dependent multi-resistor-inductor network (which is responsible for the large simulation errors in $S_{21}$ notch amplitude and $S_{11}$ amplitude and null location). But, our simple model appeared to be sufficient as a qualitative validation [31]. Using a circuit simulator [32], we obtained the traces in Fig. 7(b) with stray inductance between S and side G as $L_{stray} = 4.2 \text{ nH/mm}$. Qualitatively, Fig. 7(b) agrees well with Fig. 7(a), and validates our multi-modal hypothesis.

3.3. Eye-diagrams

We further demonstrated high-speed eye-diagrams with differential drive. The available sample for the eye-diagram measurements exhibited similar $V_p$ and EO $S_{21}$ bandwidth as reported in previous sections. The measurement was set up as following. A pair of differential 40 Gb/s 2$^{15} - 1$ pseudorandom binary sequence (PRBS) was generated by a Centellax TG1P4A PRBS source, amplified by a pair of Centellax OA4MV3 driver amplifiers, and then attenuated by passive attenuators (rated for DC–23 GHz) to the desired amplitude before being applied to the device under test through a Cascade 40 GHz rated GSGSG probe. DC bias voltages were applied to each modulator arm through bias tees inserted before the attenuators. The device probing and termination configuration were the same as the EO S-parameter measurements.
with slightly different cabling. The optical output of the device was passed through an Erbium Doped Fiber Amplifier (EDFA), an optical bandwidth filter with 3.5 nm bandwidth and then sent to the optical module (Agilent 86109B) of an Agilent 86100B digital communication analyzer (DCA).

We report on eye-diagrams with different drive voltages and modulator bias losses in Fig. 9(b)–(d). The bias loss is the excess loss due to modulator biasing, to quantify how much the bit “1” output is below the maximum transmission. For example, 3 dB bias loss means bit “1” is at the quadrature point. The differential drive voltage signals before connected to the GSCSG probe were verified by electrical eye-diagram measurements, based on which we can accurately report the drive voltages in a 50 Ω test environment. A typical 40 Gb/s electrical eye-diagram of the driving signal is shown in Fig. 9(a). The actual drive voltage received by the device is slightly lower than the reported drive voltage due to the voltage-intake factor that we discussed before. We observed a slight output asymmetry of the electrical amplifiers and attenuators on the two data paths, but the difference was within 5%, and we report the averaged drive voltage values below.

Using a 1.6 Vpp drive voltage and 0 V DC bias, 3.1 dB extinction was achieved with a very low 1.4 dB bias loss, as shown in Fig. 9(b). The power consumption can be calculated as following. At an impedance of 50 Ω, an ideal 1.6 Vpp NRZ signal centered at 0 V carries 12.8 mW of power. At 40 Gb/s, the energy per bit is 640 fJ/bit. Due to the availability of test equipment only 40 Gb/s operation was demonstrated, although we expect the device to be capable of passing higher data rate bit streams based on simulations, which would result in further reduction of energy per bit.

The achievable extinction ratio (ER) in the eye-diagrams can be traded off with drive voltage and bias loss. For example, with a 2.5 Vpp drive voltage and a 0.25 V reverse bias, 5.1 dB extinction was obtained with 1.7 dB bias loss, as shown in Fig. 9(c); with a 4.7 Vpp drive voltage and a 3 V reverse bias, 7.1 dB extinction was obtained with 0 dB bias loss, as shown in Fig. 9(d). The observed ER is lower than anticipated based on the phase shifter static performance (phase shift and dynamic loss versus applied voltage). The reduced extinction and reduced eye vertical opening is partly due to the limited bandwidth in the testing system in addition to the device itself. The electrical eye-diagrams mentioned above measured the drive signal immediately before the RF probe connected to the device, that is included the PRBS source, amplifiers, bias tees, attenuators and cables. From the electrical eye-diagram in Fig. 9(a), we measured a 10–90% rise time of roughly 15 ps, which amounts to approximately 23 GHz bandwidth and leaves a tight margin for the device. In addition to the limited system bandwidth, optical noise due to the EDFA further degraded the eye opening. In contrast, we recently demonstrated a modulator with similar RF layout at 1300 nm with the testing done at a different testing facility where higher speed components were available and EDFA was not used [33]. 50 Gb/s open eye-diagrams were demonstrated with similarly low drive voltages, suggesting the 40 Gb/s eye-diagram quality is highly dependent on and likely limited by our test equipment.

4. Conclusions

To summarize, in this paper we address the design and characterization of a high performance silicon PN junction traveling-wave Mach–Zehnder modulator. The key underlying design tradeoffs are identified and incorporated into the device design. A device is then presented with significant performance improvement compared to the state of the art.

A PN junction design should maintain a low (V_Eff<sub>p</sub>)<sup>2</sup> factor to enable low V<sub>T</sub> with high device bandwidth. Modulation efficiency needs to trade effectively with optical loss, indicated by the F<sub>3dB</sub> figure-of-merit. For both considerations we chose lightly doped PN junction with optimized 3-level side doping.

We have shown that lower device impedance is advantageous for overall device performance, and a 33–37 Ω device impedance was chosen for the TWMZ design. Having termination impedance lower than the device impedance has benefit in terms of bandwidth extension while only incurring modest voltage loss. A bandwidth extension factor of 2.2 was achieved while the voltage-intake factor was maintained at a reasonable 67%. Multimodal RF behavior is proposed and qualitatively verified to be the possible cause for the 521 notches observed in our device as well as several recent GSG modulator demonstrations. A remedy is suggested to combat this. Therefore, performance enhancement and impairment are both possible with respect to the RF design aspect of a TWMZ device, and 6.4 dB RF bandwidth is not always a good indicator of 3 dB EO response bandwidth.

A 3 mm long 30 GHz bandwidth differential-drive TWMZ modulator is demonstrated with 1 V reverse bias. The device exhibits a small-signal V<sub>T</sub> of 9 V and has a low 3.6 dB insertion loss on the phase shifters. 40 Gb/s eye-diagrams were demonstrated, and 640 fJ/bit energy per bit was achieved with 1.6 Vpp drive voltage and no DC bias; higher data rates were expected to be possible although not demonstrated due to the limitation of test equipment. Compared to the state of the art, at similar bandwidth this device shows the lowest V<sub>T</sub> and lowest drive voltage requirement due to differential-drive, or at similar V<sub>T</sub> it shows the highest bandwidth as well as the lowest insertion loss on the active phase shifter. The combined device metrics show significant improvement compared to the state of the art. Taking into account of the fabrication imperfections, the device behavior was accurately described by device models.

Acknowledgements

The authors gratefully acknowledge support from an AFOSR STTR grant (FA9550–12–C-0038) and NRF Fellowship NRF2012NRF–NRFPPP01–143. The authors would also like to thank Gernot Pomrenke, of AFOSR, for his support of the OpSIS effort, through both the PECASE award (FA9550–13–1–0027) and ongoing funding for OpSIS (FA9550–10–P-0439). The authors gratefully acknowledge the loan of critical equipment for this project from AT&T.

Reference


