Review article

Graham T. Reed*, Goran Z. Mashanovich, Frederic Y. Gardes, Milos Nedeljkovic, Youfang Hu, David J. Thomson, Ke Li, Peter R. Wilson, Sheng-Wen Chen and Shawn S. Hsu

Recent breakthroughs in carrier depletion based silicon optical modulators

Abstract: The majority of the most successful optical modulators in silicon demonstrated in recent years operate via the plasma dispersion effect and are more specifically based upon free carrier depletion in a silicon rib waveguide. In this work we overview the different types of free carrier depletion type optical modulators in silicon. A summary of some recent example devices for each configuration is then presented together with the performance that they have achieved. Finally an insight into some current research trends involving silicon based optical modulators is provided including integration, operation in the mid-infrared wavelength range and application in short and long haul data transmission links.

Keywords: modulators; silicon photonics; depletion.

*Corresponding author: Graham T. Reed, Optoelectronics Research Centre, University of Southampton, Highfield Campus Southampton, Southampton, Hampshire SO17 1BJ, UK, e-mail: g.reed@soton.ac.uk
Goran Z. Mashanovich, Frederic Y. Gardes, Milos Nedeljkovic, Youfang Hu and David J. Thomson: Optoelectronics Research Centre, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK
Ke Li and Peter R. Wilson: School of Electronics and Computer Science, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK
Sheng-Wen Chen and Shawn S. Hsu: High-Speed Devices and Integrated Circuits Group Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan

Edited by Haiseng Rong

1 Introduction

The performance of silicon optical modulators has improved dramatically in recent years, coinciding with significant research effort from both academia and commercial organisations worldwide. The motivation for developing high performance optical modulators in silicon is clear. The optical modulator is central to the majority of photonics systems for datacom applications and silicon photonics provides a low cost platform in which to form these systems, making optical based communication feasible for a number of short reach links. Various mechanisms for achieving modulation in silicon have been investigated throughout the years. Excellent progress has been made in hybrid devices where other materials are incorporated with the SOI waveguides to achieve modulation, including group III–V materials [1], germanium [2–6], polymers [7, 8] and graphene [9–13]. Furthermore the plasmonic based approach is also interesting area to achieve modulation [14–17]. This review will however focus on devices based upon the plasma dispersion effect and more specifically those using carrier depletion. Plasma dispersion effect devices are attractive as they can be formed using CMOS like processes making their integration into such fabrication facilities more straightforward. Employing carrier depletion as the mechanism to electrically modulate free carrier densities in the waveguide allows relatively simple fabrication whilst achieving high speed. Following a review of recent carrier depletion devices [18–51] we will look more into the progress of their integration with other components.

2 Performance metrics

It is useful to start by considering the different performance metrics by which the devices can be assessed. For the phase modulator there are four core metrics: the optical loss and capacitance per unit length, the phase efficiency and the speed. Design choices over the passive structure used to house the phase modulator, the phase modulator length and the drive voltage then dictate what the overall performance of the modulator will be. For example for a Mach-Zehnder modulator, the use of a short phase modulator and a low drive voltage will result in low optical loss and low power consumption. However, the extinction ratio achieved will be lower.
Conversely, if a larger extinction ratio is required either the phase modulator length or the drive voltage can be increased (or a combination of both), although this will result in an increase in optical loss or power consumption, respectively. The different performance metrics for the overall modulator structure are listed below with description.

### 2.1 Modulator drive voltage/power consumption

The power consumption of the device is an important metric to consider particularly in applications which have multiple short-reach channels. The convention with silicon optical modulators is to report the power consumption in units of energy consumed per bit of data. The opinion of the power consumption targets varies throughout the field. Some state the other benefits that optical based interconnects bring could balance an increase in power consumption compared to their electrical counterparts. The power consumption of a typical server has been reported as being on the order of 10–30 pJ/bit [52] and therefore an improvement of an order of magnitude may not be essential to justify the added value of optical interconnects. When considering longer haul applications the power consumption of the device becomes less and less important. On the other hand some state that in order for silicon photonics based interconnects to be attractive, the power consumed should not be larger than electrical interconnects which they are proposed to replace (~1 pj/bit) [53]. Some extend this argument stating that the energy should not exceed 100 fJ/bit to surpass the efficiency of the 45 nm process node equalized on-chip electrical link [54]. The optical modulator is not the only element in the system and therefore can only consume a proportion of this target power. Some state that the maximum power consumption for the modulator should therefore be on the order of 10 fJ/bit [55] which is clearly very demanding although some devices have shown potential [51, 56]. Ultimately the power consumption required is likely to be driven by the application in which the modulator is to be employed and this will also determine the modulator configuration to be employed.

The power consumption of the device can be calculated in different ways depending on how it is driven: as a lumped element or with travelling wave electrodes. However, in both cases the drive voltage has a significant impact. Throughout this analysis we assume non-return-to-zero modulation. With the phase modulator incorporated into a Mach-Zehnder interferometer (MZI), a travelling wave electrode is typically used to allow for high speed operation over a long device. A termination resistance (usually 50 Ohms) is placed at the end of the travelling wave electrode to prevent reflection of the RF signal. Any power not consumed by the phase modulator will be dissipated in the termination and therefore the power consumption of the device will be dictated by the drive voltage, the impedance of the system and the bit rate. This is described by equation (1) where $V$ is the peak to peak drive voltage, $Z$ is the impedance of the system and $BR$ the bit rate.

\[
P = \frac{(V/2)^2}{Z \times BR}
\]

(1)

Carrier depletion modulators are essentially a diode operated in reverse bias and therefore a DC level is added to the RF signal to ensure that the drive voltage across the modulator remains negative. If a DC block is employed before the termination, equation (1) can still be used, otherwise equation (2) should be used where $V_{\text{max}}$ and $V_{\text{min}}$ are the voltages of the two data levels in the drive signal.

\[
P = \frac{V_{\text{max}}^2}{2 \times Z \times BR} + \frac{V_{\text{min}}^2}{2 \times Z \times BR}
\]

(2)

With the phase modulator incorporated into a resonant structure, for example, a ring resonator, the modulator is driven as a lumped element with no additional termination element. In this case the capacitance and the drive voltage typically dictate the power consumption. This is described by equation (3) where $C$ is the capacitance of the device and $V$ the peak to peak drive voltage.

\[
P = \frac{C \times V^2}{4}
\]

(3)

It should be noted that typically a large reflection of the drive signal exists in the case of the ring. Consequently the drive voltage could be as much as double that applied to the device. This should be taken into account during the power calculations. Furthermore in the case of temperature sensitive devices such as ring resonators, an active tuning mechanism is typically required to ensure the modulator operating point is maintained with changes in device temperature. The power consumption of such a tuning mechanism should also be included in the calculation of the overall device performance. Alternatively different approaches have been demonstrated to make such structures athermal [57, 58].
2.2 Modulator loss

Aside from the phase modulator loss which is dictated by the loss per unit length and the length of phase modulator selected, there are contributions from the passive elements in the structure for example waveguide splitters, combiners and bends. The phase modulator loss comprises loss due to passive waveguide roughness, free carrier absorption due interaction with the different doped regions and any interaction with the electrode metal.

2.3 Modulation speed

This is often characterised either by its electro-optic bandwidth or its ability to modulate data at certain rates. The intrinsic speed of the phase modulator is governed by its RC time constant, however, the design and resultant electrical performance of the electrodes can also significantly influence the device transient operation. In resonant devices the photon cavity lifetime can also place a limitation on the operating speed of the device.

2.4 Modulation phase efficiency

It is the amount of phase shift produced for a given phase shifter length and drive voltage. It is often reported in units of V.cm (the voltage required on a 1 cm long phase shifter to achieve a pi radian phase shift). Alternatively an effective index change per volt is sometimes reported. In the case of carrier depletion based devices the phase shift achieved with voltage is typically non-linear, reducing with increased reverse bias. The voltage at which the phase efficiency is measured should therefore be stated.

2.5 Modulation depth or extinction ratio

This is the difference in optical output power between the 1 and 0 levels of the modulator. As the modulation rate is increased the extinction ratio can be less than in the DC case (for the same applied voltage) if the electro-optic roll-off of the device prevents the optical response completely reaching the 1 and 0 power levels. The extinction ratio produced relates to the amount of phase shift achieved and ultimately dictates the bit error rate performance of the device.

2.6 Device footprint

It is the physical size of the modulator on the chip. Often in the case of Mach-Zehnder devices this is reported as a length since this is the dominant dimension. In the case of the ring modulator, which is much more compact, the ring radius/diameter is usually stated.

2.7 Optical bandwidth

This is the band of wavelengths for which the device can operate without any tuning of the device. For example in the case of a ring resonator based modulator, the device is operated in the resonance of the device and therefore the operating bandwidth is typically of about <1 nm. On the other hand, symmetrical Mach-Zehnder based devices offer optical bandwidth larger than 80 nm.

2.8 Temperature sensitivity

It shows how sensitive the device performance is to temperature changes. Since silicon has a relatively large thermo-optic coefficient (1.8×10⁴ K⁻¹), the optical response of passive structures such as ring resonators is sensitive to temperature changes. In a symmetrical Mach-Zehnder device, the temperature induced phase changes in either arm are balanced and therefore the device performance is not affected.

2.9 Chirp

This describes slight variations in modulator output wavelength during transitions between data levels. The result of this is dispersion in the fibre which dictates the maximum distance over which data at a specific rate can be transmitted. The majority of applications that silicon photonics is considered for are short-reach, and therefore chirp is not a concern. However, as performance improves, longer haul applications become more viable and chirp then needs to be considered. Chirp will be discussed in more detail in Section 4.1.

3 Design of carrier depletion modulators

Silicon optical modulators based upon carrier depletion consist of a pn diode formed inside the optical waveguide,
which under reverse bias conditions, depletes the waveguide of carriers and causes a change in refractive index and absorption. To produce intensity modulation it is more effective to use the phase response of the device and to employ a passive structure to convert between phase modulation and intensity modulation rather than to use the modulation of the absorption coefficient directly. The two most common structures are the MZI and the ring resonator (RR) as discussed above. To some extent these structures will influence the overall performance of the devices. However, for this work we focus our analysis on the different types of carrier depletion based phase modulator elements.

There have been a wide variety of device designs reported in recent years, each of which can be generally categorised into one of three types. The first category of device is the one with a vertical pn junction similar to that shown in Figure 1 (top). The second category of device has a horizontal pn junction similar to the diagram in Figure 1 (bottom). The final category has interleaved pn junctions (or interdigitated pn junctions). An example diagram of this type of device is shown in Figure 2.

Vertical junction devices have a pn junction which runs parallel to the top and bottom surfaces of the waveguide. The first proposed [18] and demonstrated [19] waveguide based carrier depletion modulators were of this type. The simulated performance by Gardes et al. [18] demonstrated that modulation in the tens of GHz could be achieved in carrier depletion devices, which was a break through at that time [18]. Two years later the Intel Corporation demonstrated this version of the device experimentally with modulation at 30 Gbit/s presented [19]. Later 40 Gbit/s [20] modulation was produced albeit with an extinction ratio of only 1 dB. In their device the doping configuration was reversed compared to that shown in Figure 1 and contact to the top section of the device was made on just one side of the waveguide. The concentration of the p-type region was 1.5e17 cm⁻³ and the n-type region concentration varied from 1.5e17 cm⁻³ at the junction to 3e17 cm⁻³ at the top of the waveguide. A phase efficiency of 4 V·cm was measured at 4 V with optical loss of 1.8 dB/mm.

Horizontal junction devices where the pn junction runs perpendicular to the top and bottom surfaces of the waveguide are a lot more common. The first designs of this type were demonstrated in 2008 [21–23], and since then there have been numerous versions published with different variants in the design [24–31, 33–37]. The most straightforward form of the device has the pn junction positioned in the waveguide equidistant from the rib edges [25–32]. Table 1 summarises the design parameters and performances of some example horizontal junction devices demonstrated recently.

The table focuses on the performance of the phase modulator rather than the overall modulator structure and therefore the device speed can be a little misleading since as described in the previous section this is dependent on more factors than just the phase modulator itself. As can be seen in Table 1 even with the horizontal design selected and the pn junction position fixed in the middle of the...
waveguide there are still many design parameters which can be changed in order to optimise certain performance metrics. Firstly the rib waveguide geometry in terms of its height, width, slab height and sidewall angle can be varied (although in most processes the sidewall angle is fixed at around 90°). When selecting which waveguide dimensions to use there are different factors that should be considered. Firstly, the waveguide should support only single mode propagation to avoid performance degradation in interferometric and resonant structures. Secondly, if polarisation independence is required without a polarisation diversity scheme, the dimensions should be chosen to support both fundamental TE and TM modes and the effective index and optical loss for both modes should be the same [38]. The waveguide geometry also affects the confinement of the optical mode within the waveguide and therefore the degree of interaction with the region of changing refractive index (depletion region). Ultimately the optical mode confinement affects the phase efficiency of the device but also a well confined mode is likely to mean that interaction with the highly doped regions positioned away from the waveguide can be reduced, benefiting optical loss due to absorption from free carriers. The height of the slab influences the access resistance to the pn junction and therefore the speed of the device. A thick slab therefore lends itself to a higher speed. Interaction of the optical mode with side wall roughness is also reduced and consequently scattering losses can be reduced. On the other hand a thick slab region causes high losses in waveguides bends such as those required in MZI and ring resonators.

The next consideration is the concentration and positioning of the different doped regions. The distance separating the waveguide rib and the highly doped regions affects the access resistance to the device and the interaction with the optical mode. In terms of performance this translates into device speed and optical loss. Close proximity of these regions to the waveguide will reduce access resistance to the device and therefore facilitate high speed operation, however it will also result in an increased loss. The doping concentration of the highly doped regions should be high enough to allow ohmic contacts to be formed with the device electrodes as well as providing low access resistance to the junction. The concentration of the lower dose regions influences many of the performance metrics including optical loss, phase efficiency, speed and capacitance. High doping concentrations generally provide a reduced access resistance and higher junction capacitance. Its effect on the overall RC time constant could therefore be positive or negative depending on the design, however if the junction is correctly positioned it will allow for a higher efficiency since a greater change in free carrier concentration will interact with the core of the optical. The down-side to a high doping concentration is the increase in absorption loss and capacitance. Therefore another trade-off exists which can also be used to tailor the device performance towards a profile of metrics as required.

Finally the design of the electrodes is a very important consideration to optimise the speed of the device. For MZI based devices coplanar waveguide electrodes (CPW) are normally employed with the diode connected between the signal track and the ground plane to one of the sides. In order to minimise reflection losses of the drive signal these electrodes should be designed to have a characteristic impedance which is matched to that of the drive equipment (usually 50 Ohms) with the effect of the pn diode on the CPW impedance taken into account. The electrodes should also be designed such that the velocity of the electrical signal is matched to the velocity of the light in the waveguide. Any velocity mismatch will mean that the drive signal will not act on the same portion of

---

Table 1  Comparison of horizontal pn junction modulators with the pn junction positioned in the centre of the device.

<table>
<thead>
<tr>
<th>References</th>
<th>WG width (nm)</th>
<th>WG height (nm)</th>
<th>Slab height (nm)</th>
<th>P conc. (cm⁻³)</th>
<th>N conc. (cm⁻³)</th>
<th>P⁺,N⁺ to WG separation (nm)</th>
<th>Efficiency (V.cm)</th>
<th>Loss (dB/mm)</th>
<th>Capacitance (pF/μm)</th>
<th>Speed (Gb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[24]</td>
<td>480</td>
<td>300</td>
<td>80</td>
<td>4e18</td>
<td>4e18</td>
<td>–</td>
<td>–</td>
<td>7–8</td>
<td>–</td>
<td>40</td>
</tr>
<tr>
<td>[25]</td>
<td>500</td>
<td>220</td>
<td>90</td>
<td>5e17</td>
<td>5e17</td>
<td>950</td>
<td>1.9 (at 3V)</td>
<td>1.2</td>
<td>0.3–0.4</td>
<td>50</td>
</tr>
<tr>
<td>[26]</td>
<td>500</td>
<td>220</td>
<td>90</td>
<td>7e17</td>
<td>5e17</td>
<td>950</td>
<td>3.5 (at 3V)</td>
<td>1*</td>
<td>0.27</td>
<td>20</td>
</tr>
<tr>
<td>[27]</td>
<td>500</td>
<td>220</td>
<td>150</td>
<td>2e18</td>
<td>2e18</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>35</td>
</tr>
<tr>
<td>[28]</td>
<td>500</td>
<td>220</td>
<td>150</td>
<td>1e18</td>
<td>1e18</td>
<td>1000</td>
<td>1.7 (at 4V)</td>
<td>2.2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>[29]</td>
<td>500</td>
<td>220</td>
<td>100</td>
<td>1e18</td>
<td>3e18</td>
<td>–</td>
<td>1.59 (Linear)</td>
<td>3.2</td>
<td>0.84</td>
<td>30</td>
</tr>
<tr>
<td>[30]</td>
<td>450</td>
<td>340</td>
<td>80</td>
<td>2e17</td>
<td>2e17</td>
<td>1000</td>
<td>1.7 (at 3V)</td>
<td>1.2*</td>
<td>–</td>
<td>30</td>
</tr>
<tr>
<td>[31]</td>
<td>500</td>
<td>220</td>
<td>150</td>
<td>1e18</td>
<td>1e18</td>
<td>1000</td>
<td>2 (at 3V)</td>
<td>2.3</td>
<td>0.5</td>
<td>10</td>
</tr>
<tr>
<td>[32]</td>
<td>500</td>
<td>220</td>
<td>150</td>
<td>2e18</td>
<td>2e18</td>
<td>1000</td>
<td>1.2 (at 3V)</td>
<td>4.5</td>
<td>–</td>
<td>40</td>
</tr>
</tbody>
</table>

*Excludes passive waveguide loss. *Intrinsic gap of 80 nm between p and n regions is used to reduce capacitance.
light in time causing a limitation on the device speed as highlighted in the slow wave device reported by Brimont et al. [59]. RF propagation loss should be minimised. High frequency components experience greater attenuation than low frequency components meaning that the electrodes will have a limited bandwidth for which signals can be transmitted. A longer electrode will have a lower bandwidth and this means it can be difficult to achieve a large extinction ratio together with high speed due to electrode based bandwidth limitations.

Previously we have only considered devices with the pn junction fixed equidistant from the rib edges, but clearly the position of this pn junction is another parameter which can be varied. Some research groups have looked at offsetting the pn junction as shown in Figure 3 and have stated that an improvement in performance can result [23, 33, 34, 36]. Table 2 summarises the design parameters used and performance of these phase modulators.

It can be seen that usually the offset is such that the majority of the rib is formed of p-type material. This is because a larger refractive index change and lower absorption loss can be achieved with p-type than with n-type material as described in the expressions produced by Soref and Bennett [60]. Interestingly You et al. used a negative offset [23] giving the explanation that electron depletion leads to a faster electrical response than hole depletion does and therefore a faster device should result.

Since only the electrically varying depletion region within the rib waveguide contributes to the modulation of the refractive index whereas the entire doped region contributes to the optical loss some research groups have considered how the doping is localised in the waveguide with the aim of reducing optical absorption losses whilst maintaining phase efficiency and device speed. For example, in the device of Marris-Morini et al. a p-type doped slit was positioned in the core of the optical mode with the remainder of the waveguide rib left largely undoped [22]. In this way optical absorption losses due to the interaction with doping around the edges of the waveguide which does not deplete during biasing is avoided. A cross-sectional depiction of this device is shown in the top of Figure 4. A loss of 1 dB/mm occurred in this case with a phase efficiency of 3.5 V.cm. The authors were also able to demonstrate 40 Gbit/s modulation [43]. In a similar concept Tu et al. [35] used a doping compensation method to reduce the concentration of free carriers in the upper sides of the rib (as depicted in the bottom image of Figure 4). An optical loss of <1 dB/mm was reported with a phase efficiency of around 2.67 V.cm. Modulation up to 50 Gbit/s was reported.

Another consideration in the design of the device is the accuracy with which the fabrication tools can position the pn junction within the rib waveguide. A study by Gardes et al. into the extent which misalignment of the pn junction could affect device performance showed that the phase efficiency could be reduced by more than 40% with just 50 nm misalignment [36] as shown in Figure 5.

To overcome this issue and to therefore create a design which is robust to fabrication variations and therefore results in a high yield, a fabrication process using self-aligned pn junction formation has been devised [37, 38]. The process used by Thomson et al. is shown in Figure 6. Firstly, the active areas of the wafer are implanted with

**Table 2** Comparison of horizontal pn junction modulators with the pn junction positioned offset from the centre of the waveguide rib.

<table>
<thead>
<tr>
<th>References</th>
<th>WG width (nm)</th>
<th>WG height (nm)</th>
<th>Slab height (nm)</th>
<th>P conc. (cm⁻³)</th>
<th>N conc. (cm⁻³)</th>
<th>Offset towards n+ side (nm)</th>
<th>P+,N+ to WG separation (nm)</th>
<th>Efficiency (V.cm)</th>
<th>Loss (dB/mm)</th>
<th>Capacitance (fF/μm)</th>
<th>Speed (Gb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[23]</td>
<td>600</td>
<td>600</td>
<td>200</td>
<td>2e18</td>
<td>3e17</td>
<td>Negative</td>
<td>–</td>
<td>14.3</td>
<td>–</td>
<td>–</td>
<td>12.5</td>
</tr>
<tr>
<td>[36]</td>
<td>300</td>
<td>200</td>
<td>50</td>
<td>2e17</td>
<td>6e17</td>
<td>75</td>
<td>1000</td>
<td>11 (at 3V)</td>
<td>–</td>
<td>–</td>
<td>19²</td>
</tr>
<tr>
<td>[33]</td>
<td>450</td>
<td>340</td>
<td>80</td>
<td>2e17</td>
<td>2e17</td>
<td>50</td>
<td>1000</td>
<td>1.8</td>
<td>1.6</td>
<td>–</td>
<td>60</td>
</tr>
<tr>
<td>[34]</td>
<td>220</td>
<td>600</td>
<td>60</td>
<td>1e17</td>
<td>1e17</td>
<td>50⁴</td>
<td>500</td>
<td>1.5</td>
<td>1.5</td>
<td>–</td>
<td>12</td>
</tr>
</tbody>
</table>

¹This is a bandwidth measurement of 19 GHz rather than a data rate measurement. ²In this case the edge of the n-type region is offset 50 nm however the edge of the p-type is not offset, therefore there is a 50 nm intrinsic region.
boron to form background p-type doping. Next the waveguide design is etched into a silicon dioxide hard mask layer which has been deposited onto the wafer surface. This hard mask is used to etch the waveguide design into the silicon overlayer and then together with a photoresist layer it is used to guide the phosphorus implant which produces the n-type regions. Either the hard mask or the photoresist layer are sufficiently thick to prevent penetration of the phosphorus ions into the silicon overlayer and therefore the edge of the photoresist window can be aligned with low accuracy on top of the waveguide with the resultant pn junction positioned precisely at the edge of the waveguide. The remainder of the process can be performed using standard CMOS like processing steps. Using this simple process the phase modulator with cross section as shown in Figure 7 has been produced.

As the pn junction is positioned at the edge of the waveguide the doping concentration of the n-type region (1.5e18 cm⁻³) is made much higher than that of the p-type region (3e17 cm⁻³) so that the depletion extends mainly into the waveguide during biasing. This approach allows for a large phase shift with voltage with a peak efficiency figure of 2.3 V.cm demonstrated so far [39]. In a different variant of the device modulation at 40 Gbit/s with 10 dB extinction ratio [40] and 50 Gbit/s with a 3 dB extinction ratio [41] has been demonstrated from 3.5 mm long and 1 mm long phase shifters of efficiency ~2.7 V.cm, respectively. The optical loss in the phase modulator is 4 dB/mm excluding passive waveguide loss, with the largest contribution coming from the interaction with the highly doped regions. It was later shown that by increasing the separation of these regions from the waveguide by 100 nm the optical loss could be reduced to 1.1 dB/mm with little impact on the device speed and efficiency [39]. By angling the implantation in the final step of the self-aligned process as shown in top image of Figure 8 the junction can

![Figure 4](image-url)  
**Figure 4** Cross-sectional diagram of device by Marris-Morini et al. (top) [22] and Tu et al. (bottom) [35].

![Figure 5](image-url)  
**Figure 5** Theoretical study of effective index change against pn junction alignment error [36].
be accurately positioned inside the waveguide. Gardes et al. used this variant of the process to produce the phase modulator shown in the bottom of Figure 8 [38].

The aim of this device was to achieve polarisation independence from a device formed with a self-aligned process. The different sections of n-type doping allowed more degrees of freedom to balance the phase efficiencies for the fundamental TE and TM modes. Modulation at 10 Gbit/s and 40 Gbit/s was demonstrated with the same extinction ratio for both polarisations (7.3 dB and...
6.5 dB respectively). Recently other research groups have also used the same self-aligned angled ion implantation approach [42]. Furthermore it was also later reported that the device of Marris-Morini et al. could also be formed using a self-aligned process [43]. The device of Gardes et al. is a slightly different category of device which combines different pn junction types, in this case the vertical and horizontal pn junctions. This further category of device will be revisited later in this paper. The interdigitated/interleaved category of devices can also be free of alignment errors. A diagram of a typical device of this type was shown in Figure 2.

In this case the cross sectional diagram alone does not reveal the design of the junction however in the plan view the sections of p-type and n-type doping along the waveguide can be seen. Theoretically very high modulation efficiencies can be achieved with this type of device since the relative density of depletion region in the waveguide is higher than in the other two categories of device if the period of the p-type and n-type regions is small. With this increase in depletion volume comes an increase in capacitance per unit length which can be detrimental in terms of device speed and power consumption. As mentioned above, tolerance to alignment errors is possible with this structure. It can be achieved if the overlap of the doped regions with the waveguide is larger than the optical mode plus the alignment accuracy of the tools used to fabricate the device. The negative impact of increasing the overlap is again the increased capacitance per unit length. The first of such devices was proposed in 2009 by Li et al. [44]. Recently there has also been some experimental demonstrations of modulation from interleaved devices as summarised in Table 3 [31, 45–47].

A further approach as briefly mentioned above is a junction design which encompasses a combination of vertical, horizontal and interleaved sections. A device with a combination of horizontal and interleaved junctions can be achieved if a negative overlap with the waveguide is used. In this case a horizontal junction exists on the sides of the interleaved doped regions. This approach may enable optimising a trade-off of efficiency and capacitance; however it will now be sensitive to alignment errors in the fabrication. Another way to achieve a combination of horizontal and interleaved junctions is to have a zigzag shape running along the waveguide as demonstrated in a device by Xiao et al. [48] and is depicted in Figure 9.

Their device was based in rib waveguides of height 340 nm, width 500 nm and slab height 80 nm. The doping concentrations of the p-type and n-type regions was 2e17 cm⁻³, whilst the separation between the highly doped regions and the rib edge was 500 nm. With this approach the authors reported a phase efficiency of 1.7 V.cm with modulation up to 44 Gbit/s. The capacitance of the phase modulator was approximately 0.7 fF/μm, however the optical loss due to the phase modulator was not reported.

As mentioned when discussing the device of Gardes et al. [38] as shown in Figure 8 (bottom) it is also possible to combine vertical and horizontal junction type devices. Such an approach has also been proposed by Spector et al. [49] and demonstrated recently by Liow et al. [42]. The device of Liow et al. was similar to that shown in Figure 8 (bottom) without the n-type region on the right hand rib edge. The authors motivation for this design was to produce a device with a vertical junction without requiring a top contact. The authors compared a horizontal junction device and this device achieving phase efficiencies of 2.56 V.cm and 1.12 V.cm, respectively, which demonstrates the advantage of this approach.

Resenberg et al. recently presented a device that combines vertical, horizontal and interleaved junctions as depicted in Figure 10 [50]. The doping densities of the p and n-type regions was 2e18 cm⁻³, the period of the interleaving 560 nm and the width of the p-type region on the left edge of the waveguide 100 nm. The target thickness of the p-type region at the top of the rib was 110 nm. With this device the authors achieved an impressive phase efficiency of 0.84 V.cm for -2 V reverse bias. A capacitance of 0.65 fF/μm was reported with an optical loss of 3.5 dB/mm. Optical eye diagrams up to 40 Gbit/s were presented.

### Table 3: Comparison of interleaved/interdigitated pn junction modulators.

<table>
<thead>
<tr>
<th>References</th>
<th>WG width (nm)</th>
<th>WG height (nm)</th>
<th>Slab height (nm)</th>
<th>P conc. (cm⁻³)</th>
<th>N conc. (cm⁻³)</th>
<th>PN period (nm)</th>
<th>Overlap (nm)</th>
<th>P+,N+ to WG separation (nm)</th>
<th>Efficiency (V.cm)</th>
<th>Loss (dB/mm)</th>
<th>Capacitance (fF/μm)</th>
<th>Speed (Gb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[45]</td>
<td>450</td>
<td>340</td>
<td>80</td>
<td>2e17</td>
<td>2e17</td>
<td>600</td>
<td>0</td>
<td>1000</td>
<td>1.7 (at 3V)</td>
<td>1</td>
<td>0.65</td>
<td>44</td>
</tr>
<tr>
<td>[46]</td>
<td>500</td>
<td>340</td>
<td>80</td>
<td>2e17</td>
<td>2e17</td>
<td>600</td>
<td>100</td>
<td>500</td>
<td>1.4 (at 3V)</td>
<td>1.7</td>
<td>0.34</td>
<td>25</td>
</tr>
<tr>
<td>[47]</td>
<td>420</td>
<td>390</td>
<td>100</td>
<td>4e17</td>
<td>1e18</td>
<td>700*</td>
<td>0</td>
<td>–</td>
<td>3 (at 3V)</td>
<td>–</td>
<td>0.2</td>
<td>10</td>
</tr>
<tr>
<td>[31]</td>
<td>500</td>
<td>220</td>
<td>150</td>
<td>2e18</td>
<td>2e18</td>
<td>500</td>
<td>550</td>
<td>1000</td>
<td>0.8 (at 4V)</td>
<td>3.5</td>
<td>1.6</td>
<td>10</td>
</tr>
</tbody>
</table>

*P region is 400 nm, N region is 300 nm.
4 Current research trends

It can be seen that the performance of phase modulators has improved dramatically in recent years. Devices with operation in excess of 40 Gbit/s [25, 33, 41, 45, 48], phase efficiencies below 1 V.cm [50], power consumption down to 2 fJ/bit [51] and loss at 1 dB/mm and below [35, 43, 45] have been demonstrated although not with all these parameters together. Continual improvements in the device performance are still regularly being reported. Whilst work is on-going in improving the performance of the phase modulator element, research is also active on some emerging research areas and challenges for silicon optical modulators. A few examples of that work are included here.

4.1 Silicon based carrier depletion optical modulators for longer haul applications

Silicon photonics is mainly seen as a technology for short reach links. This is due to demanding performance requirements and the high performance of components formed in more traditional photonic materials such as LiNbO₃ which are used for longer reach links. As the performance of the silicon optical modulator is ever improved, application in systems with longer transmission distances could be feasible in the future. For example commercial modulators used for longer haul links report extinction ratios in excess of 13 dB (http://www.photline.com/filer/get/14/Review-ModulSpecs.pdf; http://dev.oclaro.com/product_pages/Powerbit_F10-0.php). Recently there have been demonstrations of extinction ratios in excess of 10 dB from silicon optical modulators [61, 62] with 18 dB demonstrated in one case at a data rate of 10 Gbit/s [39]. Several groups have now looked at using silicon modulators for long haul applications considering additional performance metrics which are not important for short reach links such as the modulation chirp [34, 39, 50, 62–64]. Plasma dispersion based silicon optical modulators could be problematic in terms of the chirp they produce since a change in both real and imaginary parts of the refractive index occur with changes in free carrier density. In LiNbO₃, modulators zero chirp Mach-Zehnder modulators can be achieved by balancing the optical power in the two waveguide arms and applying synchronised but complementary signals of the same amplitude in either arm (dual-drive operation). Since the phase movements in either arm cancel and the optical power from either arm is the same no movement of the output phase occurs. The accompanying modulation of the optical power in the silicon phase modulator...
means that it is not possible to balance the optical power in either arm during modulation and therefore a change in the MZI output phase with time will occur (chirp). The chirp from dual-drive silicon Mach-Zehnder modulators has been studied by different groups to assess if its impact is significant. In our work [39] we confirm that some degree of chirp is produced in the dual-drive case however when the pulse propagation over fibre is compared to that of a totally un-chirped pulse its effect is negligible. This has also been confirmed by Goi et al [62] who showed a low-dispersion-penalty over transmission lengths up to 80 km with a performance comparable to that of a commercial lithium-niobate modulator. Recently the use of more complex modulation formats from silicon optical modulators has also been the subject of investigation. For example modulation using DPSK [65], QPSK [66], DQPSK [66], PDM-QPSK [67] and PAM-16 [68] has been demonstrated from silicon based optical modulators.

### 4.2 Plasma dispersion based modulators for MID-IR

In recent years there has been a growing interest in the use of group-IV materials for mid-infrared wavelength photonics [69], motivated by possible applications in telecommunications in the 2.0–2.5 µm range, and at longer wavelengths in bio-chemical sensing, gas sensing and related applications. Research groups have begun to investigate modulation in silicon in the mid-IR.

For NIR wavelengths the Soref and Bennett equations [60] are used for predicting the size of the free-carrier plasma dispersion effect in silicon, but as the effect is wavelength dependent those equations cannot be used in the mid-IR. Nedeljkovic et al. [70] have calculated predictions for the change in refractive index and change in absorption coefficient for a given change in charge carrier concentration at wavelengths in the 1.3–14 µm range, using the same semi-empirical approach used by Soref and Bennett. The change in absorption coefficient was calculated from the absorption spectra of heavily doped silicon wafers found in literature, and the change in refractive index was derived from the change in absorption spectra by way of the Kramers-Kronig relations. They found that the size of the effect is approximately proportional to \( \lambda^2 \), as described by the Drude-Lorentz model. They predict that for modulators operating at \( \lambda > 4 \) µm it would be more effective to depart from NIR modulator design where phase modulators used in conjunction with a passive structure to convert to intensity modulation, and to use electro-absorption modulators, where modulation is provided by the change in absorption coefficient directly, such as simple charge-carrier injection into a straight waveguide.

Researchers at IBM have demonstrated the first SOI modulator operating at 2165 nm [71], which is based on carrier-injection in a MZI, with 1 mm long p-i-n diode phase shifters in each arm. Using pre-emphasis they achieved bitrates up to 3 Gbps, an insertion loss of 9.0 dB and a V_{RL} figure of merit of 0.012 V.cm. The rib waveguides were similar to those common in NIR devices (width×height =500×220 nm, with 170 nm etch depth). The authors do not compare their results to the modulation predictions in [70].

### 4.3 Silicon based carrier depletion modulator for short reach links

As discussed in Section 4.1 the target application areas for silicon photonics are in shorter reach links where the performance requirements are relaxed from those required for traditional long haul photonic links. The power consumption becomes a key concern especially as the link reach is decreased and the density of links increases. In recent years producing low power consumption silicon optical modulators has been a focus area for many groups who have followed different routes. As was discussed in section 2 the length of a typical Mach-Zehnder based carrier depletion modulator requires travelling wave electrodes to be used with a 50 Ohm termination. Any power not consumed by the modulator will be dissipated in the termination and therefore only the drive voltage (and termination impedance) dictate the power consumption of the device. If it is not practical to employ a DC block between the device and the termination, additional DC power consumption should be avoided by operating around zero volts with an RF drive amplitude which is low enough not to cause the device to go into the carrier injection regime. This approach has been followed by Ding et al. where a power consumption of 146 fJ/bit was achieved [32]. The key in these devices is to use highly efficient phase modulators in order to produce a sufficient modulation depth with a lower drive voltage. Employing a slow light waveguide in which to produce the phase modulator is one possibility to achieve a high efficiency [59, 72]. Dispersion engineering can be used to mitigate against the problem of wavelength and temperature sensitivity which these devices usually suffer from to some extent [72].

A more popular technique is to use a resonant structure in which the phase modulator is fabricated, for
example a ring or a disk resonator. These devices produce intensity modulation by shifting the resonance wavelength on and off the input wavelength through operation of the phase modulator. Since the device can be driven as a lumped element, no termination is required and therefore power dissipation in the termination is avoided. Several examples of low power and high speed resonant modulator have been demonstrated [24, 48, 50, 51, 73, 74]. The sensitivity of resonant based devices to fabrication variations and temperature changes require them to be actively tuned to fix the resonance wavelength as required. Athermal techniques can be used to avoid the resonance wavelength shifting with temperature [75], however this prevents thermal tuning being used to counter fabrication variations. Active tuning can be performed by incorporating a heater as demonstrated by Li et al. [24] or by adding a controllable DC level to the RF drive voltage if the modulator is efficient enough and the tuning range required is sufficiently small [76]. An important consideration is the power consumption of the tuning element itself and therefore the efficiency of the different heater architectures possible [77], and methods to improve the efficiency [78], have attracted increasing research interest recently.

4.4 Integration of silicon photonic modulators and detectors

One of the key cited advantages of silicon photonics is the potential to co-fabricate electronic and photonic elements on the same piece of silicon [79]. Such an approach is technologically elegant and also has the potential for optimised performance since any degradation introduced by the electrical connections between different chips should be eliminated. One example of this approach is that followed by Luxtera in the fabrication of their Active optical cables (http://www.luxtera.com/). The drawback to such an approach is the high per-area cost of high performance electronic processing and the relative large size of the photonics elements. There is also the complexity of combining the processes for the photonic and electronic components without affecting device performance and the consideration of the required substrate in each case.

A more cost effective and less complex strategy could be to separately fabricate photonic and electronic chips and then to integrate them by a wire bonding or flip chip bonding approach. The photonic chip can therefore be fabricated in a lower specification fabrication facility at lower cost.

Wire bonded integration is often referred to as an approach limited to low speed applications. The parasitic inductance introduced by the bonding wire and parasitic capacitance introduced by the bonding pads may reduce the bandwidth of the link and cause an impedance discontinuity. Within the UK Silicon Photonics programme the wire bonded approach for integrated silicon photonic transmitters and receivers operating up to 10 Gbit/s has been investigated.

Figure 11 shows a optical microscope image of the integrated transmitter module which has been assembled. The image shows a CMOS modulator driver and a silicon optical MZI modulator. The modulator incorporates 3.5 mm long carrier depletion based phase modulators in either arm. Coplanar waveguide (CPW) electrodes are used to co-propagate the 10 Gbit/s electrical drive signal with the light in the optical waveguide. At the end of phase shifters the CPW electrodes are tapered to facilitate wire bonding. The dual drive CMOS driver was fabricated using the IBM-8RF 130 nm process [80]. A receiver has also been constructed consisting of a trans-impedance amplifier wire bonded with a germanium optical detector. The TIA was fabricated using the 180 nm TSMC technology [81].
If the lengths of the wire bond are minimised then higher speed operation should also be possible; indeed it has been shown elsewhere that wire bond integration of silicon photonics can operate at 25 Gbit/s [82]. It should be noted in this case the limitation in the speed was set by the electronic elements and the detector which were designed to operate up to 10 Gbit/s.

5 Summary

We have presented and provided published examples of three different configurations of carrier depletion phase modulator in silicon, as well as devices which are a mixture of these configurations. Each category of the device contains several degrees of freedom in the design which can be used to tailor a profile of performance metrics as required for a specific application. There are however some distinct advantages and disadvantages to each approach which are summarised herein.

Horizontal and interleaved junction devices can be electrically contacted via the slab regions on either side of a conventional rib waveguide, whereas the vertical junction device requires contact to the top of the waveguide. This means that the use of a conventional rib waveguide is not possible in the case of a vertical junction device and that a growth process is required in the fabrication which is more complex than the implantation approaches used to form the other two configurations of device.

Interleaved devices with high p and n type doping concentrations can theoretically be fully depleted if the period of the interleaving can be made small enough. This means that highly efficient devices can be made with low absorption loss. In highly doped vertical and horizontal junction devices, carriers far from the junction remain undepleted and therefore cause high optical losses limiting the achievable efficiency. Doping localisation methods can be used to overcome this to some extent at the expense of fabrication complexity.

The intrinsic bandwidth of the phase modulator element is dictated by the access resistance and junction capacitance. The power consumption of the device can also be influenced by the junction capacitance depending on how it is driven. The device capacitance which causes efficient modulation (or the capacitance-efficiency product) is theoretically higher (or worse) in interleaved devices since regions at the edges of the waveguide where low optical power is propagating, undergo the same depletion as at the core. Doping localisation methods can again be used to counteract this to some extent. In each
category of device the access resistance can be kept low by keeping the contacts in close proximity to the waveguide whilst not introducing significant overlap with the optical mode which would result in high optical losses. The doping levels in the slab regions which connect to the device should also be kept sufficiently high in order to avoid high access resistance.

Vertical and interleaved junction devices have a further advantage of being inherently alignment tolerant which simplifies fabrication; however self-aligned processes can be used to produce horizontal devices with relaxed alignment demands.

Devices which mix the different junction types can bring about further advantages, for example, by combining an interleaved and horizontal junction device in way such that the pn junction is corrugated along the centre of the waveguide, the efficiency of the horizontal device can be improved.

Some excellent device performance has been demonstrated in the literature in recent years, with operation in excess of 40 Gbit/s, phase efficiencies below 1 V.cm, power consumption down to 2 fJ/bit and loss at 1 dB/mm and below demonstrated (although not simultaneously), confirming the carrier depletion optical modulator as an increasingly promising approach. Improving the performance of the carrier depletion modulator is an ongoing challenge. A device which combines high efficiency, low optical loss and high speed simultaneously is still sought after. At the same time significant research effort is devoted to the integration of the carrier depletion silicon optical modulator with electronics and other photonic components.

There are several approaches and challenges for integrating electronics with photonics. Front-end integration where the electronics and photonics are formed side by side on the same wafer offers the prospect of the best performance, however the cost and complexity are high and integration density low. A back-end approach where the photonic components are formed using process temperature low enough to not damage the pre-fabricated electronics can be used, however, the performance of the photonic components formed in this way is generally inferior. A back-end wafer bonding approach can be used to affix pre-processed electronics and photonics wafers, however the performance of the resultant electrical connections between the two wafers and the yield of the bonding process require consideration. Flip chip and wire bonding can also be used, however, the parasitic capacitance and inductance introduced by these approaches could degrade and limit the performance to some extent. A 10 Gbit/s wire bond integrated silicon photonic transmitter and receiver have been presented in this work. Furthermore a 25 Gbit/s wire bond integrated photonic link has been demonstrated in the literature.

Carrier depletion based silicon optical modulators remain a strong candidate for use in the growing number of short reach photonic applications and possibly even some longer reach links. Novel design features are continuously being introduced which are yielding improvements in the device performance. With the significant worldwide resources now devoted to research in this area it is clear that there is more to come from the carrier depletion silicon optical modulator into the future.

Acknowledgements: Authors acknowledge funding from the EPSRC in the UK to support the UK Silicon Photonics project.

Received May 30, 2013; accepted November 19, 2013; previously published online December 12, 2013

References


