IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS, VOL. 28, NO. 3, MAY/JUNE 2022

MOS Capacitor-Driven Silicon Modulators: A Mini Review and Comparative Analysis of Modulation Efficiency and Optical Loss

Wei-Che Hsu, Bokun Zhou, and Alan X. Wang, Senior Member, IEEE

Abstract—Metal-oxide-semiconductor (MOS) capacitor-driven silicon modulators have demonstrated exceeding performance in driving voltage, energy efficiency, and bandwidth through heterogeneous integration with other active materials, including III-V compound semiconductors, transparent conductive oxides (TCOs), and graphene. However, a proper quantitative comparison is lacking among various device structures due to the difference in material properties and device designs. In this article, we first briefly reviewed state-of-the-art MOS capacitor-driven silicon modulators. Following that, we modeled the modulation efficiency and optical loss of two most representative device structures by incorporating the optical properties such as plasma dispersion, band-filling, and bandgap shrinkage effects for several heterogeneously integrated materials. The comparative analysis based on simulated results shows that Ge, III-Vs, and TCOs all offer larger index modulation than Si, whereas graphene has the largest index modulation. At the end of this article, we discussed the strategy of heterogeneous integration based on design device and scalable fabrication.

Index Terms—Silicon photonics, heterogeneous integration, electro-optic modulators, III-V materials, transparent conductive oxides, graphene.

I. INTRODUCTION

METAL-OXIDE-SEMICONDUCTOR (MOS) capacitor, as one of the most pivotal electronic device structures, has gained growing usage in active photonic devices in recent years [1]–[3]. Particularly, MOS capacitor-driven silicon (Si) photonic devices have rapidly become one of the most promising building blocks for future integrated photonic devices due to their extraordinary performance in electro-optic (E-O) modulation and scalability in fabrication. Si photonic devices based on MOS-capacitor structures usually operate in the accumulation mode to achieve much larger refraction index modulation than the depletion mode of traditional reversed PN junctions. The refractive indices of the p-type Si (p-Si) waveguide and the n-type poly-Si gate are modified by the accumulation layers through the negative gate bias so that an optical phase shift is induced to the guided light. Furthermore, MOS-capacitor structures provide the feasibility of heterogeneous integration with other active materials, such as III-V compound semiconductors, transparent conductive oxides (TCOs), and graphene [4]–[7], which can overcome the limitation of the weak plasma dispersion effect in Si. From a materials perspective, III-V has a large carrier-induced refractive index change due to the strong plasma dispersion and band-filling effect [4]. In addition, the bonded III-V material is a single crystalline material with high electron mobility [4]. It has a lower material loss, and the resistance-capacitance (RC) delay is smaller than poly-Si or amorphous Si material at the same doping level. Likewise, the strong plasma dispersion effect of TCOs can achieve unity-order refractive index change [8]. As a two-dimensional (2-D) material, graphene has a very low density of states and high electron mobility owing to its electronic band structure, offering significant refractive index modulation [6], [9], [10].

Although various MOS capacitor-driven Si photonic devices have been demonstrated in the past two decades, there has been no proper comparison of these heterogeneous photonic devices in terms of modulation efficiency and optical loss due to several reasons. First, many of these devices were designed with different dimensions to be optimal for their own applications. Second, the gate materials vary significantly in properties such as carrier mobility and concentration because they are largely dependent on the crystallinity and doping level, which are determined by the fabrication process. For example, III-V materials (e.g., InGaAsP and InP), which can be crystalline and moderately doped \((10^{15} \text{ cm}^{-3})\), are grown by molecular beam epitaxy or metal-organic chemical vapour deposition and then transferred by wafer-bonding [4], [11]. TCOs such as Indium-Tin Oxide (ITO) and Titanium-doped Indium Oxide (TiO) for hybrid Si modulators are usually polycrystalline and heavily doped \((10^{19} \text{ cm}^{-3})\). They can be directly deposited through chemical vapor deposition, atomic layer deposition, or sputtering deposition [12]–[15]. Last but not least, the performance of these MOS capacitor-driven Si photonic devices strongly depends on the dielectric constant and thickness of gate insulators. Therefore, directly comparing the experimental results of reported MOS capacitor-driven Si photonic devices lacks a universal standard and baseline, which cannot reflect the advantages of these heterogeneously integrated gate materials. In this paper,
Fig. 1. Two representative active waveguides as fundamental building blocks for MOS capacitor-driven silicon modulators: (a) Configuration-1: Top-bonded design with crystalline films bonded on top of the Si waveguide. (b) Configuration-2: Conformal deposition design with polycrystalline films covering the top and side walls of the Si waveguide.

II. REVIEW OF MOS CAPACITOR-DRIVEN SI MODULATORS

A. Review of Device Structure

We briefly reviewed MOS capacitor-driven Si modulators with different gate materials and proposed two representative MOS device configurations with heterogeneously integrated gate materials. The materials properties for each configuration are chosen according to real device fabrication requirements. After that, a comparative analysis is conducted through numerical simulation, showing that III-V materials and TCOs have larger index modulation than Si although at the cost of higher loss; whereas graphene has the largest index modulation and low insertion loss due to Pauli blocking. At the end of this article, we discussed the strategy of heterogeneous integration based on device design and scalable fabrication. We conclude that conformal deposition design would benefit from better mode overlapping and the compatibility of sputtering deposition.

B. Review of Device Performance

This section reviewed the hybrid Si modulators, which were demonstrated through experiments, and the device performance materials, which are usually polycrystalline or amorphous with lower electron mobility [21], [43].

We summarized existing reports of MOS capacitor-driven Si modulators as shown in Fig. 2. Typically, III-V is top-bonded on the Si waveguide, while TCOs are usually integrated by conformal deposition such as RF-sputtering deposition. Hence, the proposed hybrid Si modulators with III-V and TCO are designed with Configuration-1 and Configuration-2, respectively. Poly-Si and graphene are grown by CVD, and they can be formed as either Configuration-1 or Configuration-2. In section IV, we will discuss the overlapping and device performance between these two configurations.
with different gate materials was summarized in Table I. These MOS capacitor-driven Si modulators operate in the accumulation mode through negative gate voltages to modulate the refractive indices of both the Si waveguide and gate materials through the plasma dispersion effect. There are two fundamental types of MOS capacitor-driven Si modulators: interference modulation based on Mach-Zehnder interferometer (MZI) and resonance modulation using microring or other types of resonators such as nanocavities. Fig. 3 shows the schematic of different modulators. Mach-Zehnder modulator (MZM) (Fig. 3(a)) relies on phase modulation induced by the accumulated carriers in one or two of its arms by the applied gate voltages. The phase difference from the two arms is translated into amplitude modulation of the light by optical interference when the two arms are combined into a single channel. From Table I, we can find that MZMs with Si as the gate materials have poor $V_pL$ in the early years due to the relatively weak plasma dispersion effect of Si and the large waveguide dimension, which induces small overlap factor between the optical mode and the thin accumulation layer of the gate materials. In contrast, other gate materials provide much better $V_pL$ below 1 V-cm so that the long device length is not required. Especially, the ITO-based MZM modulator achieves an extremely low $V_pL$ of 0.0095 V-cm that can reduce the device length to 1.4 $\mu$m. The other key factor in device performance is the insertion loss, predominated by the carrier concentration, mobility, and waveguide quality [44]. Note that the doping level of III-V is usually on the order of $1 \times 10^{17}$ to $1 \times 10^{18}$ cm$^{-3}$, whereas the doping levels of TCO and graphene are up to $\sim 10^{20}$ cm$^{-3}$. For semiconductor materials, high carrier concentration and low carrier mobility will induce high optical absorption. III-V materials generally have moderate carrier concentration and high mobility, which leads to relatively low optical loss. TCO materials such as ITO are degenerate semiconductors, and they intrinsically possess very high carrier concentration and low carrier mobility due to oxygen vacancies, which result in much higher optical loss. For graphene, the high optical loss is caused by the interband absorption. Hence, the MZM with ITO gate suffers high insertion loss, which is almost three orders of magnitude as high as other materials. In contrast to ITO, graphene has high mobility up to 1000 cm$^2$V$^{-1}$s$^{-1}$ and a single atomic layer thickness [45], which can help to reduce insertion loss.

Compared to interference modulation, resonance modulation using a microring resonator or nanocavity can potentially achieve much better energy efficiency and more compact size. For example, a microring modulator (MRM) (Fig. 3(b)) uses a small ring radius, and the resonant wavelength can be shifted by effective index changes of the waveguide [46]. The high energy efficiency of MRM comes from the high quality factor (Q-factor) and ultra-compact size of the microring resonator [47]. However, the resonant wavelength of a high Q-factor resonator is very sensitive to temperature variations or fabrication errors [48]. Thus, MRM commonly requires a thermal heater to control the resonant wavelength [49]. The E-O tuning efficiency in Table I depends on the resonant wavelength shift, which is mainly determined by the effective index modulation by the applied bias. Therefore, using pm/V can represent the E-O efficiency of the MRM without considering the size of the ring. Table I shows that there is no clear trend of the E-O tuning efficiency (pm/V) to the resonant wavelength regarding to the gate materials, which is different from the case on MZM devices. The reason is that the E-O modulation efficiency of MOS capacitor-driven Si modulators is strongly affected by the properties of the gate oxide since the accumulated carrier concentration is determined by the capacitance of the MOS capacitor [36], [50]. The gate oxide thickness listed in Table I is in the range of 5 to 20 nm in MRMs, and the oxides materials are also various in dielectric constants. For instance, the MRM can achieve a high E-O tuning efficiency of 130 pm/V with a thin oxide layer of 5 nm even if it uses a poly-Si as the gate material. The MOS capacitor-driven MRM provides a high E-O tuning efficiency to compensate for the temperature drift, and it only consumes $\sim$ pW/nm for the power, which is negligible compared to the traditional thermal tuning efficiency of mW/nm [28], [35]. Therefore, such a device eliminates the thermal controlling system and enables denser integration. While Q-factor represents the loss of the MRM, which is influenced by the properties of gate materials or the fabrication quality. For the MOS capacitor-driven Si MRMs, gate materials on the ring waveguide induce extra optical absorption compared with regular passive Si waveguide, which will result in lower Q-factors. Therefore, the property and thickness of gate materials have a significant impact on the Q-factor of MRMs. We need to point out that the Q-factor of the nanocavity resonator is less affected by the gate material compared to MRMs [50]. For the nanocavity, obtaining a high Q-factor is through the Gaussian-like attenuation profile in the device, which can be achieved by tapering air holes to fine-tune the mirror strength.
### TABLE I
MOS CAPACITOR-DRIVEN SILICON MODULATORS WITH DIFFERENT GATE MATERIALS

<table>
<thead>
<tr>
<th>Gate material</th>
<th>Gate method</th>
<th>Gate doping (cm⁻²)</th>
<th>Oxide thickness (nm)</th>
<th>V₄L (V·cm)</th>
<th>Insertion loss (dB/cm)</th>
<th>Active length (L) (μm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly-Si</td>
<td>CVD</td>
<td>1×10¹⁸</td>
<td>12</td>
<td>8.0</td>
<td>6.7</td>
<td>10000</td>
<td>[1]</td>
</tr>
<tr>
<td>ELO-Si</td>
<td>ELO</td>
<td>1×10¹⁸</td>
<td>10.5 (SiO₂+Si₃N₄)</td>
<td>3.3</td>
<td>10.1</td>
<td>3450</td>
<td>[18]</td>
</tr>
<tr>
<td>InGaAsP</td>
<td>Wafer bonding</td>
<td>1×10¹⁷</td>
<td>12+3 (SiO₂+Al₂O₃)</td>
<td>0.047</td>
<td>4.6</td>
<td>500</td>
<td>[23], [25]</td>
</tr>
<tr>
<td>InP</td>
<td>Wafer bonding</td>
<td>~10¹⁷</td>
<td>20 (Al₂O₃)</td>
<td>0.54</td>
<td>2.3</td>
<td>240</td>
<td>[26]</td>
</tr>
<tr>
<td>ITO</td>
<td>IBD</td>
<td>3.1×10²⁰</td>
<td>10 (SiO₂)</td>
<td>0.28</td>
<td>236</td>
<td>300</td>
<td>[10]</td>
</tr>
</tbody>
</table>

### Microring Resonator

<table>
<thead>
<tr>
<th>Gate material</th>
<th>Gate method</th>
<th>Gate doping (cm⁻²)</th>
<th>Oxide thickness (nm)</th>
<th>E-O tuning efficiency (pm/V)</th>
<th>Q-factor</th>
<th>Ring radius (R) (μm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly-Si</td>
<td></td>
<td>1×10¹⁸</td>
<td>5</td>
<td>130 (V₄L: 0.24 V·cm)</td>
<td>3500</td>
<td>5</td>
<td>[19]</td>
</tr>
<tr>
<td>InP</td>
<td>Wafer bonding</td>
<td>~10¹⁸</td>
<td>16 (HFO₂)</td>
<td>275</td>
<td>~1000</td>
<td>20</td>
<td>[27]</td>
</tr>
<tr>
<td>InGaAsP</td>
<td>Wafer bonding</td>
<td>1×10¹⁸</td>
<td>6 (Al₂O₃)</td>
<td>~115</td>
<td>3368</td>
<td>(Race-track) (Radius: 5, Active arm: 100)</td>
<td>[29]</td>
</tr>
<tr>
<td>ITO</td>
<td>Sputtering</td>
<td>~10¹⁹</td>
<td>16 (HFO₂)</td>
<td>216</td>
<td>1000</td>
<td>12</td>
<td>[33]</td>
</tr>
<tr>
<td>ITIO</td>
<td>Sputtering</td>
<td>2.63×10¹⁹</td>
<td>17 (HFO₂)</td>
<td>48.5</td>
<td>4800</td>
<td>6</td>
<td>[34]</td>
</tr>
<tr>
<td>ITIO</td>
<td>Sputtering</td>
<td>~10¹⁹</td>
<td>10 (HFO₂)</td>
<td>130</td>
<td>11900</td>
<td>6</td>
<td>[35]</td>
</tr>
<tr>
<td>Graphene</td>
<td>CVD</td>
<td>~10²⁰</td>
<td>20 (SiO₂)</td>
<td>~44 (blue shift) ~90 (red shift)</td>
<td>3875</td>
<td>10</td>
<td>[39]</td>
</tr>
</tbody>
</table>

### Nanocavity

<table>
<thead>
<tr>
<th>Gate material</th>
<th>Gate method</th>
<th>Gate doping (cm⁻²)</th>
<th>Oxide thickness (nm)</th>
<th>E-O tuning efficiency (pm/V)</th>
<th>Q-factor</th>
<th>Active length (L) (μm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITO</td>
<td>Sputtering</td>
<td>1×10²⁰</td>
<td>20 (SiO₂)</td>
<td>30</td>
<td>1000</td>
<td>0.375</td>
<td>[36]</td>
</tr>
<tr>
<td>In₂O₃</td>
<td>Sputtering</td>
<td>~10¹⁹</td>
<td>10 (HFO₂)</td>
<td>250</td>
<td>3700</td>
<td>0.350</td>
<td>[37]</td>
</tr>
<tr>
<td>ITO</td>
<td>Sputtering</td>
<td>2×10²⁰</td>
<td>16 (HFO₂)</td>
<td>71</td>
<td>5600</td>
<td>1</td>
<td>[39]</td>
</tr>
</tbody>
</table>

---

*a*: ELO: epitaxial lateral overgrowth.

*b*: IBD: ion beam deposition.

'c': The racetrack consists of two half rings with a radius of 5μm and two straight arms of 200 μm, and the active region only covers one arm with 100 μm.
Therefore, the Q-factor of the nanocavity is majorly determined by the air holes design. On the other hand, the Q-factor of microring is determined by the round-trip loss in the ring, including carrier absorption and bending loss. When the ring resonator has a large radius, the Q-factor of the ring is limited by the carrier absorption, which is strongly influenced by gate materials. When the radius is small, the Q-factor is majorly restrained by the bending loss and therefore a small radius microring resonator cannot achieve a high Q-factor.

Very recently, nanocavity modulators (Fig. 3(c)) as another type of resonance-based modulators have demonstrated even more compact size (at least 50× smaller than MRM [36]) and better energy efficiency. So far, only a few gate materials have been applied to MOS capacitor-driven Si nanocavity modulators. As discussed above, the E-O modulation efficiency benefits significantly from the large capacitance of the MOS capacitor; moreover, it also can be enhanced by the small active volume of the nanocavity [36]. As listed in Table I, the length of the nanocavity modulus in the active region is less than 1 μm, and it can achieve a high E-O tuning efficiency of 250 pm/V using a gate oxide with a high-κ material. By contrast, it is a minor influence on the Q-factor in nanocavity, and the Q-factor is majorly affected by the airhole designs and fabrication quality [36], [37]. The resonance-based modulators can achieve low energy consumption (CV^2/4), and such devices can reduce the energy consumption to the range of fJ/bit, which is much more efficient than the Si-gated MZM with the range of pJ/bit [29]. Nanocavity modulator can particularly achieve extremely high energy efficiency of 3 fJ/bit due to the ultra-compact size [37]. A more detailed comparison between nanocavity modulators and other types of resonance modulators can be found in our recently published article [50].

### III. Design and Simulation

#### A. Design of Two Representative MOS Capacitor-Driven Si Waveguides

Deriving from reported devices, we design two representative active waveguides as fundamental building blocks for MOS capacitor-driven Si modulators, which are illustrated in Fig. 1. Configuration-1 shows the top-bonded design using a p-Si [51] rib waveguide (250 nm in height, 450 nm in width, with a 50 nm thick partially etched slab) with a thin film stack on top. The 15 nm-thick HfO_2 gate insulator and 20nm-thick gate materials are grown on a separate Si substrate and then transferred onto the rib waveguide through wafer bonding. The index matching layer of 100 nm Si is used to maintain proper overlap between the mode profile and the active material due to the difference in the material refractive indices. Configuration-1 considers the case where epitaxially grown materials are maintaining excellent crystalline quality, offering larger carrier mobility. Configuration-2, which is a conformal deposition design, is built upon a p-Si rib waveguide (250 nm in height and 450 nm in width, surrounded by 50 nm thick partially etch slab) with conformally deposited thin film stacks. The 15 nm-thick gate insulator covers both the top and sidewalls of the rib waveguide. On top of that is a 20 nm-thick gate materials as listed in Table II, including n-Si [52], [53], Ge [54], InP [55], [56], ITO [14], ITiO [57], CdO [58] and Graphene [45]. The conformal deposition of gate materials increases the overlapping area with the optical mode profile. However, such design can only be implemented through vapor deposition, and gate materials are polycrystalline or amorphous with much smaller carrier mobility. In both configurations, Ni/Au is used to form Ohmic contact with the bottom Si and top gate material separately.

#### B. Materials Properties and Modeling Procedures

To simulate the carrier-induced refractive index change, materials properties of the bulk and the accumulation layers are both taken into consideration, as listed in Table II. For the bulk properties, carrier concentration and mobility are two critical variables, which are determined by fabrication processes. Carrier concentration can be effectively tuned by doping, while carrier mobility is decided by the crystallinity, depending on deposition techniques. Here, we assume that for Si [51], Ge, and InP materials, the doping level is at 1 × 10^{18} cm\(^{-3}\). For TCOs and graphene, the doping level is at 1 × 10^{20} cm\(^{-3}\) based on the typical setting from the literature [43]. Due to the wafer-bonding process [25], the materials used in the top-bonded configuration are considered to be crystalline. Therefore, the heterogeneously integrated gate materials maintain larger electron mobility shown as \(\mu_1\). In contrast, the materials used in conformal deposition are polycrystalline or amorphous, resulting in the smaller value of electron mobility shown as \(\mu_2\). Although efforts have been made in improving the crystallinity of TCOs, for example, by controlling the crystallinity of target and process pressure [59], [60], TCOs are mostly polycrystalline at best.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Relative permittivity (\epsilon_r)</th>
<th>Mobility (single crystalline) (\mu_1) (cm^2/V·s)</th>
<th>Mobility (polycrystalline) (\mu_2) (cm^2/V·s)</th>
<th>Effective mass (m^*)</th>
<th>Doping level (N_e) (cm(^{-3}))</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Si</td>
<td>11.8</td>
<td>500</td>
<td>100</td>
<td>0.36</td>
<td>1 × 10^{18}</td>
<td>[51]</td>
</tr>
<tr>
<td>n-Si</td>
<td>11.8</td>
<td>1600</td>
<td>100</td>
<td>0.26</td>
<td>1 × 10^{18}</td>
<td>[52], [53]</td>
</tr>
<tr>
<td>Ge</td>
<td>16.0</td>
<td>1000</td>
<td>200</td>
<td>0.12</td>
<td>1 × 10^{18}</td>
<td>[54]</td>
</tr>
<tr>
<td>InP</td>
<td>10.0</td>
<td>1600</td>
<td>100</td>
<td>0.08</td>
<td>1 × 10^{18}</td>
<td>[55], [56]</td>
</tr>
<tr>
<td>ITO</td>
<td>3.9</td>
<td>-</td>
<td>25</td>
<td>0.35</td>
<td>1 × 10^{10}</td>
<td>[14]</td>
</tr>
<tr>
<td>ITiO</td>
<td>3.9</td>
<td>-</td>
<td>70</td>
<td>0.31</td>
<td>1 × 10^{10}</td>
<td>[57]</td>
</tr>
<tr>
<td>CdO</td>
<td>5.3</td>
<td>-</td>
<td>205</td>
<td>0.24</td>
<td>1 × 10^{10}</td>
<td>[58]</td>
</tr>
<tr>
<td>Graphene</td>
<td>9.0</td>
<td>-</td>
<td>1000</td>
<td>-</td>
<td>1 × 10^{20}</td>
<td>[45]</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: University of Texas at Austin. Downloaded on September 21,2022 at 20:47:54 UTC from IEEE Xplore. Restrictions apply.
The mobility of TCOs is mostly determined by types of TCOs and deposition conditions. Here, we adopt different TCOs to represent their differences in mobility. Other material parameters can also be found in Table II. Si, Ge, and III-V intrinsically have larger relative permittivity and smaller effective mass. The carrier mobility is reduced due to moderate doping. TCOs have larger effective mass, smaller carrier mobility, but higher doping level, which result in higher free carrier absorption and scattering loss. In contrast, graphene has large carrier mobility, even when heavily doped [41], but the thickness is limited to a single atomic layer (0.37 nm) [61].

For active Si waveguides, the material properties in the accumulation layer play the most pivotal role for the modulator. We identified three key carrier effects: free-carrier absorption, band-filling, and bandgap shrinkage.

For TCOs, due to larger effective mass, the optical properties can be solely described by the well-established Drude Model [62]:

\[
\epsilon_r = \epsilon' + i\epsilon'' = \epsilon_{\infty} - \frac{\omega_p^2}{\omega^2 + \gamma^2} + i \frac{\omega_p^2 \gamma}{\omega(\omega^2 + \gamma^2)}
\]  

(1)

Here, \(\epsilon_{\infty}\) is the high-frequency dielectric constant. \(\omega_p\) is the plasma frequency, which is proportional to the carrier concentration \(N_c\) by \(\omega_p = \frac{qN_c^{\frac{3}{2}}}{m^*}\), where \(q\) is the charge of the electron. \(\epsilon_0\) is the permittivity of vacuum and \(m^*\) is the effective mass of the carrier. The plasma collision frequency \(\gamma\) is inversely proportional to the carrier mobility \(\mu\) by \(\gamma = \frac{\omega_p}{\mu q}\). The change of carrier concentration results in the change of both the real and imaginary parts of the permittivity. Due to the small high-frequency dielectric constant and large carrier concentration, the real part permittivity of ITO can be reduced to zero, and the frequency dielectric constant and large carrier concentration, of carrier concentration results in the change of both the real and imaginary parts of the permittivity, triggering the non-linear ENZ effect.

For III-V materials, however, both the band-filling effect and bandgap shrinkage play roles in interband absorption [34], [63]–[65]. On the one hand, due to the low density of states, the electron from the valance band requires greater energy to be optically excited to the conduction band after the lowest energy state is filled. On the other hand, when the concentration is large enough, the screening of electrons would lower the energy of the conduction band edge that is determined by free-carrier density. Overall, the absorption change caused by the band-filling and band-shrinkage effect at photon energy \(E\) is represented as [25]:

\[
\Delta \alpha = \begin{cases} 
0, & |E| < E_{g,eff} \\
\frac{C_{hh}}{E} \sqrt{E - E_{g,eff}} \left[ f_v(E_{hh}) - f_c(E_{hh}) \right] + \frac{C_{lh}}{E} \sqrt{E - E_{g,eff}} \left[ f_v(E_{lh}) - f_c(E_{lh}) \right], & E_{g,eff} \leq |E| \leq E_g \\
\frac{C_{hh}}{E} \sqrt{E - E_{g,eff}} \left[ f_v(E_{hh}) - f_c(E_{hh}) \right] + \frac{C_{lh}}{E} \sqrt{E - E_{g,eff}} \left[ f_v(E_{lh}) - f_c(E_{lh}) \right], & E > E_g 
\end{cases}
\]  

(2)

\[
\alpha_0 (E) = \frac{C_{hh}}{E} \sqrt{E - E_g} + \frac{C_{lh}}{E} \sqrt{E - E_g}
\]  

(3)

Here, \(C_{hh}\) and \(C_{lh}\) are constants corresponding to the density-of-states effective mass of heavy holes and light holes, \(f_v\) and \(f_c\) are Fermi-Dirac distribution for holes and electrons. \(E_{hh}, E_{lv}, E_{ch}, E_{cl}\) are the energy levels of electrons or holes. \(E_{g,eff}\) is the effective bandgap reduced by bandgap shrinkage. Using the carrier-induced absorption changes, the carrier-induced refractive index change can be calculated using the Kramers–Kronig relations [63]. As an example, we extrapolated the calculated optical properties of InP from [25], which shows that after considering the band-filling and band-gap shrinkage effect, the index change is more accurately depicted than only considering the plasma dispersion effect. The most deviation occurs when carrier concentration is between 1 \(\times\) 10\(^{17}\) \(cm^{-3}\) and 1 \(\times\) 10\(^{18}\) \(cm^{-3}\).

The optical properties of graphene can be modeled from its in-plane conductivity, considering the combined effect of intraband and interband transitions, given by the Kubo formula [45]:

\[
\sigma_p = \frac{i q^2 \omega}{\pi (2\pi h)^2} \times \left[ \frac{1}{(\omega - i 2\gamma)^2} \int_{-\infty}^{\infty} \left( \frac{\partial f_d (-\epsilon)}{\partial \epsilon} - \frac{\partial f_d (\epsilon)}{\partial \epsilon} \right) d\epsilon \right]
\]  

(4)

\[
\in_p = 1 - \frac{\sigma_p}{i \omega \epsilon_0 \Delta}
\]  

(5)

Here, \(f_d (\epsilon) = 1/[e^{(\epsilon-E_{\text{F}})/k_B T} + 1]\) is the Fermi-Dirac distribution, \(E_{\text{F}}\) is the Fermi level, \(h\) is the plank constant, and \(\Delta = 0.35\) nm is the thickness of graphene. The first and second terms represent the intra- and inter-band transitions. When the Fermi level reaches half of the photon energy, the permittivity of graphene reaches the maximum, as listed in Table II. Further increasing the Fermi level, the inter-band absorption will be the inter-band absorption will be prohibited due to the Pauli blocking [66], resulting in a dramatic decrease in the imaginary and real part of permittivity.

C. Workflow of Simulation

The workflow of the simulation consists of three steps. In the first step, the carrier distribution along the normal direction \(x\) to the MOS interface is simulated by Silvaco\textsuperscript{TM} under various applied voltage \(V\) using a 20-nm gate/15-nm HfO\textsubscript{2}/p-Si MOS capacitor configuration. The Quantum moment model calculates the carrier distribution of the accumulation layer in a MOS capacitor with the ultra-thin insulator more accurately [67]. Fig. 4 plots the carrier distribution in the normal direction to the MOS capacitor interface using different gate materials. The shaded area marks the region 5 nm away from the interface. On the side of p-Si, the hole concentration distribution in p-Si is almost identical when using various gate materials. The reason is that the total accumulated charge density is determined by the MOS capacitance density, which is a constant when using the same dielectric material and thickness. The spreading of the
hole distribution is decided by Debye length:

$$L_D = \sqrt{\frac{\varepsilon k_B T}{q^2 N_C}}$$  \hspace{1cm} (6)

$\varepsilon$ is the dielectric constant of the p-type (or n-type) material. $k_B$ is the Boltzmann’s constant. $T$ is the absolute temperature in kelvins. $q$ is the charge of the electron. $N_C$ is the doping concentration. Clearly, the Debye length for p-Si is a constant, regardless of the materials on the n-side. Fig. 4(b) plots the electron concentration distribution on the n-side, which strongly depends on the gate insulator thickness and dielectric constant. For n-Si, Ge, and InP, the bulk carrier concentration is set to be $1 \times 10^{18}$ cm$^{-3}$. At 0V, since InP has a much smaller effective mass, the peak carrier concentration is smaller and is distributed further from the interface. At higher voltage, the peak carrier concentration reaches above $1 \times 10^{20}$ cm$^{-3}$. Similarly, for ITO, ITiO, and CdO, the bulk carrier concentration is set to be $1 \times 10^{20}$ cm$^{-3}$. The difference in spreading can be estimated using Debye length. The dielectric constant of TCOs is only about half of the other materials due to larger effective mass and smaller mobility, but the doping concentration of TCOs is two orders of magnitude larger. Hence, the Debye length of TCO is much smaller than others. In fact, the most change of electron distribution in TCOs occurs at the first 1-nm accumulation layer.

Since the total accumulated charge density is constant and TCOs have the smallest Debye length, it is easier for the peak carrier concentration of TCOs to reach the ENZ concentration, which is approximately $6.5 \times 10^{20}$ cm$^{-3}$.

In the second step, the carrier distribution data obtained from the previous step is used to calculate the refractive index modulation according to the plasma dispersion effect and/or band-filling and band-shrinkage effects, as described in section III B. In the third step, the index modulation as a function of the applied voltage is simulated using Lumerical™ MODE at the wavelength of 1550 nm, with imported materials properties. Since graphene is a monolayer, the optical property is directly calculated and imported into the third part, not following the first and second steps. The simulation results are discussed in Section IV.

IV. SIMULATION RESULTS

A. Mode Profiles

Fig. 5 plots the simulated mode profile of the fundamental transverse electric (TE) mode of the active Si waveguides on the linear scale. Fig. 5(a)-(e) show the mode profiles at 0V bias for the two different waveguide configurations. In Configuration-1, the top bonded silicon layer serves as the index matching layer and shift the center of the optical waveguide mode profile toward the carrier accumulation layer, which can improve the overlap factor. Such effect is especially significant for n-Si, Ge, and InP, but less effective for TCOs and graphene. In Configuration-2, the optical field enhancement at the vertical waveguide sidewalls due to the discontinuity of the optical permittivity will also benefit the overlap factor. As can be seen in Fig. 5, the conformal deposition waveguides show strong optical field intensity inside the insulator and gate materials. For a quantitative analysis of the overlap between the optical mode and accumulation layers, the overlap factor is defined as:

$$\alpha = \frac{\iint_{\text{accumulation}} |E|^2 \, dA}{\iiint_{\text{all}} |E|^2 \, dA}$$  \hspace{1cm} (7)

Here the surface integral in the numerator is the electric field intensity in the accumulation layers. The calculated overlap factors are labeled in Fig. 5. We can see that both configurations offer similar overlap factors and the slight difference depends on the gate materials. In general, large refractive index gate materials will concentrate more optical field and induce higher overlap factors, which are more preferred in top-bonded design. As a comparison, TCOs will perform better in conformal deposition design due to the low refractive index. Since graphene is a single atomic layer (< 0.5nm) material, the overlap area compared to other gate materials is smaller and it works better in conformal deposition. We need to pay special attention to TCO materials. As illustrated in Fig. 4(b), when the applied bias is greater than ~8 V, the peak free carrier concentration of TCOs in the accumulation layer is above $5 \times 10^{20}$ cm$^{-3}$, which will turn TCOs into ENZ states, resulting in strong light-matter interactions. Fig. 5(f) compares the CdO-gated waveguides with Configuration-1 and Configuration-2, respectively. In contrast to Configuration-1, Configuration-2 clearly shows extremely high...
optical field intensity in the accumulation layers of the sidewall in the conformal deposition design at -10V bias. Such an ENZ effect will induce greater E-O modulation efficiency than regular waveguide modes.

B. Effective Index and Loss

Fig. 6(a) shows the modulation of the effective index ($\Delta n_{\text{eff}}$) for Configuration-1. When the optical mode profile does not change significantly in the waveguide upon E-O modulation, $\Delta n_{\text{eff}}$ is linearly proportional to the overlap factor, the capacitance of the device, and the applied bias [50]. MOS capacitor-driven Si modulators are generally operated in the accumulation mode, and therefore the device’s capacitance is a constant. As results, the effective index modulation of devices using Si, Ge, InP, and TCOs as gates are almost linear across the applied bias range. However, this assumption is not true at ENZ modulation, in which the optical mode profile is significantly altered by the ENZ material.

In Fig. 6, Ge and graphene have much large modulation than n-Si, while InP offers the most significant index modulation. TCO materials do not perform well in top-bonded design due to the low overlap factors. Fig. 6(b) shows the optical loss of the waveguides under different biases. As n-Si and InP have moderate doping levels with high mobility, such MOS capacitor-driven waveguides demonstrate very low optical loss. In contrast to n-Si and InP, Ge has high optical loss due to the high absorption coefficient at the telecom wavelength range [68]. Besides, TCOs also have a large optical loss due to the two orders of magnitude higher doping levels and the lower mobility. For graphene, the initial -2V of bias is used to achieve Pauli blocking. After that, the dramatic decrease in both the real and imaginary parts is anticipated, resulting in the lower loss at higher voltages.

As illustrated in Fig. 7(a), Configuration-2 provides slightly larger overlap between the optical mode and the accumulation layers, and the refractive index modulation is slightly improved compared with Configuration-1. However, the difference also depends on the gate materials and device parameter optimization. In this configuration, it seems that InP and Ge have better EO modulation efficiency than TCO materials due to the ultra-small effective mass, which will result in higher plasma dispersion effect. However, this conclusion is not very convincing since there is little research about the plasma dispersion effect.
Fig. 7. Simulated (a) modulation of effective index and (b) insertion loss, for conformal deposition design (Configuration-2), showing the non-linear change of TCOs. Note: n-Si, Ge and InP are poly-crystalline by considering the fabrication process.

TABLE III
SIMULATED DEVICE PERFORMANCE

<table>
<thead>
<tr>
<th>Materials</th>
<th>Top-bonded (Configuration-1)</th>
<th>Conformal Deposition (Configuration-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( V_L ) (V/cm)</td>
<td>Insertion loss (dB/cm)</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>n-Si</td>
<td>0.210</td>
<td>12.1</td>
</tr>
<tr>
<td>Ge</td>
<td>0.199</td>
<td>166</td>
</tr>
<tr>
<td>InP</td>
<td>0.0783</td>
<td>6.06</td>
</tr>
<tr>
<td>ITO</td>
<td>0.242</td>
<td>352</td>
</tr>
<tr>
<td>TiO</td>
<td>0.228</td>
<td>167</td>
</tr>
<tr>
<td>CdO</td>
<td>0.199</td>
<td>101</td>
</tr>
<tr>
<td>Graphene</td>
<td>0.126</td>
<td>196</td>
</tr>
</tbody>
</table>

effect of polycrystalline Ge and III-V compounds. Also there are no experimental articles of Ge and III-V gated modulators using Configuration-2. When the applied bias is above \(-8\) V, the non-linear index change of TCOs is shown because TCOs in the accumulation layers are turned into the ENZ state, offering both enhanced index modulation and increased optical loss. Fig. 7(b) shows higher optical loss than Configuration-1 (Fig. 6(b)) due to the lower mobility of the polycrystalline gate materials used in Configuration-2. Moreover, the \( V_L \) in the linear region is calculated by \( \Delta n_{eff} \) from Figs. 6(a) and 7(a). The insertion loss is based on the simulation data of Figs. 6(b) and 7(b) at 0V, which are calculated from the imaginary part of the effective index. The \( V_L \) and insertion loss at 0V for each material and configuration are summarized in Table III. A more comparative analysis of these materials in device performance will be discussed in Section IV-C.

C. Strategy of Heterogeneous Integration

The performance of different MOS capacitor-driven waveguides is listed in Table III. From the perspective of materials compared to n-Si, Table III shows that Ge and TCOs offer the improved \( V_L \) although at the cost of higher loss than n-Si. In contrast, III-V can provide more efficient \( V_L \) with a low loss. Graphene also gives good \( V_L \) and suppresses insertion loss once reached Pauli blocking, which will drop below 75.5 dB/cm. Because of the improvement in \( V_L \), the MOS capacitor-driven waveguides can be heterogeneously integrated with other gate materials instead of n-Si.

From the waveguide configuration perspective, the top-bonded design (Configuration-1) with III-V can provide the lowest insertion loss with a good \( V_L \) by bonding single-crystalline gate materials but requires a larger footprint. In contrast, the conformal deposition design (Configuration-2) can provide a compact size to enable denser integration. It offers similar or slightly better E-O modulation efficiency compared with Configuration-1, but with significant improvement for TCOs and graphene, which reduces \( V_L \) by at least \( 1.5 \times \). However, the insertion loss for Configuration-2 is also much higher due to the polycrystalline state of gate materials.

From the perspective of devices, although TCOs in the ENZ states have the potential for dramatic refractive index modulation, it is not suitable for the phase or resonance modulators due to the high insertion loss. However, this ENZ effect can be applied in electro-absorption (EA) modulators [69]–[71]. Typically, MZM needs a relatively long active length, so the low \( V_L \) with the low or moderate loss materials are preferred. Hence, Configuration-1 with III-V gate materials and Configuration-2 with high mobility TCO (e.g., CdO) are good candidates. Similarly, for the MRM, the low \( V_L \) reduces the modulation voltage to shift the resonant wavelength. Besides, the low or moderate loss materials, such as III-V, CdO, and graphene, maintains the high Q-factor of MRM. Therefore, MRM can achieve much lower energy consumption with the compact size of the microring resonator. As for the nanocavity modulator, it is proper to earn a small feature size by Configuration-2. It also benefits from those gate materials with low \( V_L \) to improve the index modulation. Moreover, the Q-factor of the nanocavity modulator has a minor impact on the gate materials, so the high Q-factor can be achieved by optimizing air hole design and improving fabrication quality. These advantages discussed above render the nanocavity modulator much higher energy efficiency and less sensitive to the gate materials, realizing the atto-joule/bit energy efficiency [50].

From the perspective of fabrication in the actual application, however, the most established deposition technique of the III-V combines epitaxial growth and wafer-bonding, limiting the morphology of the fabricated device. Hence, the III-V gated Si modulator will be limited to Configuration-1. TCOs are the most feasible choice for Configuration-2 with conformal deposition techniques such as sputtering deposition, which is
highly mature in the semiconductor industry and ready for mass production [72]. Graphene has the largest index modulation with a low insertion loss. Still, scaling existing graphene production techniques to the semiconductor industry without compromising its properties is a current challenge [73].

V. CONCLUSION

In this article, we first briefly reviewed the state-of-the-art MOS capacitor-driven Si modulators. We derived two most representative waveguide configurations, representing high-mobility crystalline materials in top-bonded design and low-mobility polycrystalline films in conformal deposition design. We modeled E-O modulators based on the two proposed waveguide configurations by incorporating optical properties such as plasma dispersion, band-filling, and bandgap shrinkage effects for several heterogeneously integrated materials reported in recent literature. The properties of these materials were chosen based on actual device fabrication requirements. The comparative analysis based on simulated results shows that the gate materials of Ge, III-V, TCOs, and graphene all offer better effective index modulation than Si. However, the optical loss induced by these heterogeneously integrated materials must be carefully considered. Based on the quantitative simulation results, we discussed the strategy of heterogeneous integration from different perspectives, including the feasibility of large-scale device fabrication. In conclusion, it requires comprehensive consideration and holistic optimization when choosing the waveguide configuration and gate materials to achieve both high EO modulation efficiency and low optical loss.

REFERENCES


et al. received the B.S. degree from Tsinghua University, Beijing, China, in 2000, the M.S. degree in 2002 in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include a modulator and photonic circuits based on TCO gated microring.

Wei-Chie Hsu received the B.S. degree in physics from National Taiwan Normal University, Taiwan, in 2015, and the M.S. degree in 2020 in materials science from Oregon State University, Corvallis, OR, USA, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include a modulator and photonic circuits based on TCO gated microring.

Bokun Zhou received the B.S. degree in physics from Northwest University, Xi’an, China, in 2015, and the M.S. degree in 2020 in materials science from Oregon State University, Corvallis, OR, USA, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include a modulator and photonic circuits based on TCO gated microring.

Xiaoxiao Wang (Senior Member, IEEE) received the B.S. degree from Tsinghua University, Beijing, China, in 2000, the M.S. degree in 2002 in electrical and computer engineering from the Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China, in 2003, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2006. From January 2007 to August 2011, he was with Omega Optics Inc., Austin, Texas, where he was the Chief Research Scientist for multiple SBIR/STTR Projects. Since September 2011, he has been with the School of Electrical Engineering and Computer Science, Oregon State University. He has authored or coauthored more than 100 journal and conference publications, including twelve invited and plenary presentations. He holds seven issued or pending U.S. patents. He is a Senior Member of SPIE and OSA.