Research Article Vol. 31, No. 14 / 3 Jul 2023 / *Optics Express* 22481

Optics EXPRESS

High-speed polarization-independent plasmonic modulator on a silicon waveguide

NIKHIL DHINGRA,[1] **HAMID MEHRVAR,**[2]
AND PIERRE BERINI[3,4,5,*]

[1] *Lanemap Photonics, New Delhi 110031, India*
[2] *Huawei Technologies Canada, Canada, Kanata, ON K2K 3J1, Canada*
[3] *School of Electrical Engineering and Comp. Sci., University of Ottawa, Ottawa, ON K1N 6N5, Canada*
[4] *Department of Physics, University of Ottawa, Ottawa, ON K1N 6N5, Canada*
[5] *NEXQT Institute, University of Ottawa, 25 Templeton St., Ottawa ON, K1N 6N5, Canada*
[*] *berini@eecs.uottawa.ca*

Abstract: The electrical bandwidth of an electro-optic modulator plays a vital role in determining the throughput of an optical communications link. We propose a broadband plasmonic electro-optic modulator operating at telecommunications wavelengths ($\lambda_0 \sim 1550\,\text{nm}$), based on free carrier dispersion in indium tin oxide (ITO). The ITO is driven through its epsilon-near-zero point within the accumulation layers of metal-oxide-semiconductor (MOS) structures. The MOS structures are integrated into a pair of coupled metal-insulator-metal (MIM) waveguides aligned on a planarized silicon waveguide. The coupled MIM waveguides support symmetric and asymmetric plasmonic supermodes, excited adiabatically using mode transformation tapers, by the fundamental TM_0 and TE_0 modes of the underlying silicon waveguide, respectively, such that the modulator can operate in either mode as selected by the input polarisation to the silicon waveguide. The modulator has an active section 1.5 to 2 μm long, enabling the modulator to operate as a lumped element to bandwidths exceeding 200 GHz (3 dB electrical, RC-limited). The modulators produce an extinction ratio in the range of 3.5 to 6 dB, and an insertion loss in the range of 4 to 7.5 dB including input/output mode conversion losses. The AC drive voltage is ±1.75 V. The devices comprise only inorganic materials and are realisable using standard deposition, etching and nanolithography techniques.

1. Introduction

The ongoing increase in demand for data requires modulators to be high speed and efficient [1]. Silicon modulators offer good performance while being amenable to well-established CMOS fabrication technologies [2–4]. However, in recent years, a variety of materials have shown improvement in different aspects of modulator performance, such as organic polymers [5,6], phase change materials [7], graphene [8], and indium tin oxide [9–12]. Plasmonic modulator structures allow light confinement beyond the diffraction limit, thus allowing extremely compact modulators [5,13,14]. They offer additional advantages such as small capacitances and resistances, and consequently, small associated RC time constants which enable lumped-element operation over extremely high electrical bandwidths.

Inorganic electro-optic materials offering strong index modulation and compatibility with silicon photonics represent ideal choices for plasmonic modulator design. One such material is indium tin oxide (ITO), which recently has shown potential in several photonic devices including modulators [9–12] and optical phased arrays [15–17]. In a metal-oxide-semiconductor (MOS) structure, the charge density within the ITO can be modulated by an applied bias, which in turn modulates the complex refractive index of ITO. Depending on the polarity and the magnitude of the applied voltage, the device can be driven into accumulation or depletion near the oxide-ITO interface. Given the low background permittivity of ITO, the epsilon-near-zero (ENZ) regime

#489902 https://doi.org/10.1364/OE.489902
Journal © 2023 Received 17 Mar 2023; revised 29 May 2023; accepted 30 May 2023; published 21 Jun 2023

can be accessed under strong accumulation. As the carrier concentration crosses the ENZ point, ITO undergoes a transition from dielectric-like to metal-like behaviour resulting in a very large refractive index modulation at the telecommunication wavelengths. This effect can be combined with the strong light confinement produced by plasmonic waveguides, leading to deep optical modulation over a large electrical bandwidth in a compact footprint.

In most of the previously reported designs of ITO modulators based on plasmonic structures (*cf.* [8,10,11]), a hybrid plasmonic waveguide (HPW) configuration was used to manage the confinement-attenuation trade-off [18], implemented as a MOS structure. A design based on coupled hybrid plasmonic waveguides (CHPW) has also been proposed for modulator applications [19], as another approach to manage the confinement-attenuation trade-off. These designs use direct butt coupling from an input silicon waveguide, and do not include mode transformation sections to optimise coupling to the HPW mode. Due to mode mismatch, direct butt coupling from the input silicon waveguide produces insertion loss and may result in power coupling to other forward propagating modes such that the power output depends on the interference between the excited modes of the HPW modulator section. Also, this configuration gives rise to additional semiconductor resistance in the cathode contacts, which in turn limit the electrical bandwidth of the design.

We propose a design for an ITO based plasmonic modulator where the carrier density is modulated in metal-oxide-semiconductor (MOS) structures backed by a bottom metal film, which collectively forms a vertical plasmonic metal-insulator-metal (MIM) waveguide [18]. The MIM waveguide strongly confined fields in the thin HfO_2-ITO regions, producing a strong overlap between the modal field and the ITO layer which is perturbed into the ENZ region, and thus requires a smaller length of the active region to achieve a given extinction ratio. A smaller active area enables the capacitances to be very small, leading to a very high electrical bandwidth. A pair of coupled MIM waveguides form the active section of the modulator, which operates with coupled plasmonic modes. Input and output taper sections are designed for efficient excitation of the relevant mode of the active section by the fundamental TE or TM mode of an underlying Si waveguide. The input and output tapered sections are kept electrically isolated from the active section by a gap in the anode metal film at the junction between them, such that they do not contribute to the device capacitance and do not limit the electrical bandwidth of the design.

We first give an overview of our design, then present charge transport results within the ITO layer to obtain the spatial charge density as a function of applied voltage. Subsequently, we transform the spatial charge density profiles into corresponding refractive index profiles and carry out modal analyses of the active section to determine the optical performance. We then discuss how the relevant modes of the active section of the device can be excited by an appropriately designed taper. We also calculate the parasitic capacitance and resistance, and estimate the electrical bandwidth of the design. Finally, we conclude with brief remarks.

2. Device concept

Our concept exploits an active section of length L_{active}, and input and output mode conversion tapers of length L_T, as shown in Fig. 1. The active section is responsible for modulating the MIM modes propagating through, whereas the input and output tapered sections transform adiabatically the modes of the underlying Si waveguide into corresponding modes of the active section, and vice versa. The modes supported by the active section are TM supermodes of the MIM waveguide pair, having symmetric or asymmetric lateral field distributions, evolving along the input taper from the fundamental TM or TE mode (respectively) of the Si waveguide. The same taper design can enable adiabatic transformation of the TE or TM mode of the Si waveguide, such that the modulator can operate with either, depending on which one is launched into the input Si waveguide. The same taper design is applied at the output of the active section. The computed mode profiles plotted on Fig. 1 show the transformations.

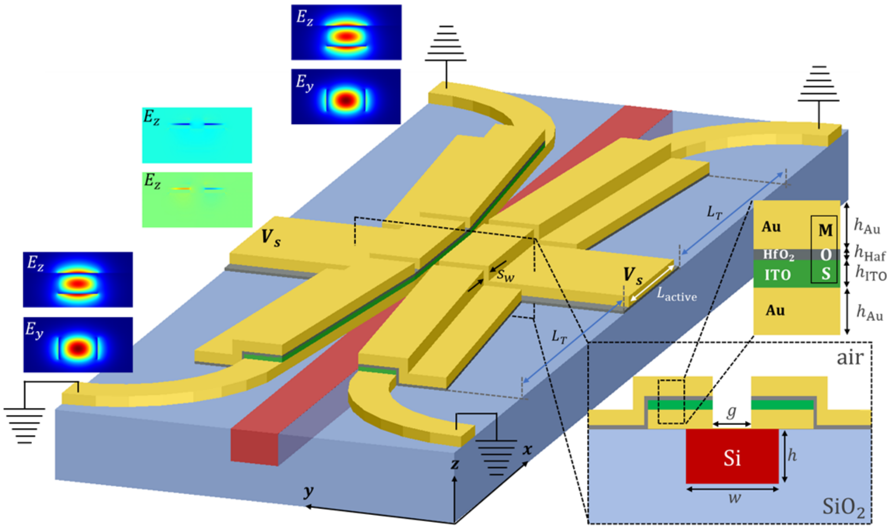

Fig. 1. Schematic of the modulator along with expanded cross-sectional views of the active section. The structure is covered by a 20 nm thick hafnia passivation layer (not shown) which also infills the slits (s_w) created to isolate electrically the tapered sections from the active section.

The active section of the design is composed of a planarized silicon waveguide below two metal-oxide-semiconductor-metal stacks, separated from each other by a gap g. The top Au anode along with the hafnia (HfO_2) and ITO films form a MOS structure. The Au cathode below the ITO serves three purposes: (*i*) forming MIM plasmonic waveguides producing strong confinement in the region between the two Au films to increase optical overlap with the accumulation region in the ITO; (*ii*) forming a low-loss Ohmic contact to the ITO; and (*iii*) enabling low-resistance electrical ground (cathode) contacts by extending this layer to the sides (rather than the ITO region). Anode contacts can be formed by extending the top Au/ HfO_2 layers to their respective sides. The anode and cathode contacts on each side of the device can be thickened into signal (V_s) and ground pads for contacting to a broadband electrical probe.

The applied voltage between the anode and cathode results in a vertical charge distribution profile within the ITO region. This leads to a corresponding change in the complex refractive index in the ITO, and the complex effective index of the propagating MIM modes. Depending upon the change in the real or imaginary part of the effective index, the propagating modes undergo a modulation of phase or absorption as a function of applied voltage. The modulation efficiency depends primarily on the interaction between the modal fields and the perturbed semiconductor region, which justifies our preference for MIM structures.

The active section is electrically isolated from the tapered sections by slits of width s_w in the top metal layer at the junction of the active section and each tapered section. These slits are required to eliminate parasitic capacitance and resistance associated with the tapers, thereby enabling a broad electrical bandwidth. We assume a 20 nm thick hafnia passivation layer on top of the device (not shown in Fig. 1), which also infills the slits of width s_w.

Finally, an embedded planarized Si waveguide is selected for three reasons: (*i*) to provide a flat planar surface facilitating fabrication of the modulator stack; (*ii*) to minimise the parasitic series resistance associated with the anode and cathode contacts, and (*iii*) to enable effective mode conversion tapers to adiabatically transform the modes of the underlying Si waveguide into corresponding plasmonic modes of the active section.

Modelling this design requires a variety of simulation techniques. For the modal analysis, we use the vectorial finite difference eigenmode (FDE) technique. We use the eigenmode expansion (EME) method to model the adiabatic mode conversion tapers. To calculate the insertion loss associated with the gap region between the active and tapered sections, we use the finite difference time domain (FDTD) method as the absence of guided mode in the gap region makes it difficult to model using the EME method. For charge transport simulation, we find self-consistent solutions to Poisson's and the drift-diffusion equations. All of these simulation techniques are available as different modules in Ansys Lumerical [20]. FDE and EME solvers are part of the MODE module whereas FDTD is a standalone module. Charge transport solver is available in the Multiphysics module.

Fabrication could proceed by applying nanofabrication techniques implemented using wafer processing tools [9,15,17,19]. For instance, embedded Si waveguides could be fabricated by etching a SOI wafer, depositing a SiO_2 cladding, and planarizing using chemical-mechanical polishing. The MIM stack could be formed by depositing and lifting-off the bottom Au layer (cathode) and the thin ITO layer. The HfO_2 layer could be formed by atomic layer deposition, and the top Au layer (anode) also by deposition and lift-off. The gap region of width g could be formed by focussed ion beam milling as well as the isolation slits (S_w).

3. Results and discussion

3.1. Electrostatic modeling

The detailed cross-sectional view of the active section of our proposed design is shown in Fig. 2. As the driving electric field is applied between the anode (top Au electrode) and cathode (bottom Au electrode), the carrier concentration within the ITO region starts to vary. Depending upon the applied voltage and its relation to the work function difference between the anode material and ITO ($\phi_{\mathrm{Au}} - \phi_{\mathrm{ITO}}$) there may be accumulation or depletion of the majority carriers (electrons) within the ITO near the HfO_2-ITO interface. For an applied voltage $V_s > V_{\mathrm{fb}}$, where $V_{\mathrm{fb}} = (\phi_{\mathrm{Au}} - \phi_{\mathrm{ITO}})/e$ is the flat-band voltage, free electrons accumulate near the hafnia-ITO interface, whereas for $V_s < V_{\mathrm{fb}}$, the ITO is depleted near the HfO_2-ITO interface (bottom right inset of Fig. 2). Inversion is not typically observed in ITO given its large bandgap energy.

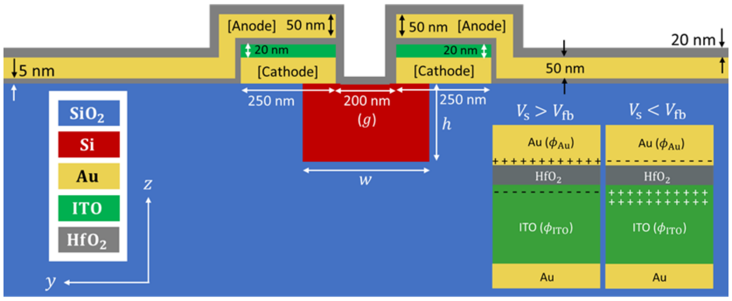

Fig. 2. Expanded cross-sectional view of the active section of the modulator. The bottom right inset illustrates two regimes of operation: accumulation $V_s > V_{\mathrm{fb}}$ (left), and depletion $V_s < V_{\mathrm{fb}}$ (right).

The 2D charge distribution over the device cross-section was computed by finding self-consistent solutions to the Poisson and drift-diffusion equations [20]. The length of the active section is expected to be larger (µm's) than its width (100's nm), so computationally expensive 3D charge simulations are not required.

ITO was modelled as an n-doped semiconductor of bandgap energy $E_g = 2.8$ eV, DC permittivity $\varepsilon_{\mathrm{DC}}^{\mathrm{ITO}} = 9.3$, electron affinity $\chi_s = 4.8$ eV, effective electron mass $m_n^* = 0.35 m_e$, and effective hole mass $m_p^*=m_e$, where m_e is the mass of a free electron [15]. In addition, we also assume an unperturbed carrier density of 10^{19} cm^{-3}. Corresponding to this value of carrier concentration, the Fermi level of ITO lies within the conduction band, so ITO behaves as a degenerate semiconductor. The work function of ITO (ϕ_{ITO}) with unperturbed carrier density of 10^{19} cm^{-3} is obtained as 4.766 eV. The work function of gold is taken as $\phi_{\text{Au}} = 5.1$ eV. HfO_2 was selected as the insulating layer because its DC relative permittivity is high, taken as 25 [21,22], such that the required charge perturbation in ITO can be achieved at a voltage much smaller than required for other dielectrics of the same thickness but lower permittivity, such as SiO_2 or Al_2O_3. The breakdown field in HfO_2 strongly depends on the deposition process. For HfO_2, we assume a breakdown field of 6.4 MV/cm [16], which corresponds to a breakdown voltage of $V_b = 3.2$ V for a 5 nm thick film. The underlying silicon waveguide has a negligible effect on the charge distribution in ITO, so it is not included in the charge simulations. The thicknesses of the HfO_2 and ITO films (h_{Haf}, h_{ITO}) are chosen as 5 and 20 nm, respectively. The cathode and anode metal films are assumed to be 50 nm thick. The width of the cathode (w_{cat}) is chosen as 250 nm as below this value the MIM structure does not support the plasmonic supermodes of interest.

We simulate the structure to obtain the spatial carrier profile within the ITO as a function of applied voltage from - 4 V to + 4 V. We observe two different regimes: accumulation for V_s>0.334 V, and depletion for V_s<0.334 V, where $(\phi_{\text{Au}} - \phi_{\text{ITO}})/e = 0.334$ V is the flat-band voltage. In Fig. 3, we plot the 2D electron profiles within the ITO for −2 and 3 V, which are in the depletion and accumulation regimes, respectively. As the structure is symmetric along the horizontal (y), we show the charge distribution in the right half of the structure only.

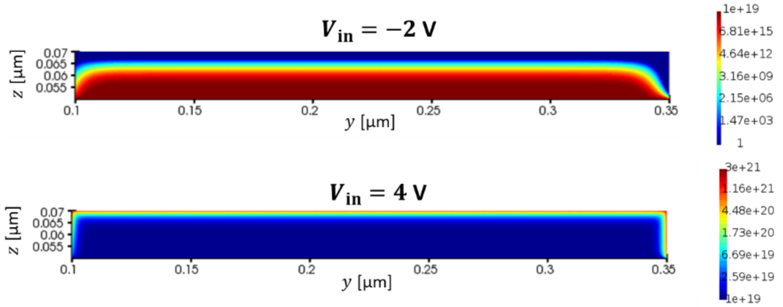

Fig. 3. 2D electron concentration distributions in the ITO over the right half of the device for $V_s = -2$ V (depletion, top) and 4 V (accumulation, bottom). The colorbar is in units of cm^{-3}.

As the width of the MOS stack is larger than the thickness of the ITO, the 1D carrier density profile in the z-direction does not vary much along the width of the ITO film except near the corners. For better visualization of the charge profile as a function of applied voltage, we plot the 1D carrier density along the thickness of ITO at the horizontal centre of one of the MOS stacks. Figure 4(a) shows the 1D carrier density distribution along the ITO thickness in the depletion regime V_s<V_{fb}, revealing that the depletion width extends deeper into the ITO as the applied voltage drops further below the flat-band voltage.

In Fig. 4(b), we plot the 1D charge profile in the accumulation regime V_s>V_{fb}, along with an expanded view of the distribution for better visualization. The charge density profile decays exponentially away from the HfO_2-ITO interface, following classical drift-diffusion theory (the accuracy of the profile could be improved by using the more computationally intensive Schrödinger-Poisson theory [23]). The charge density is observed to increase near the HfO_2-ITO interface as the voltage increases above the flat-band voltage. The horizontal dashed line in Fig. 4(b) represents the epsilon-near-zero (ENZ) carrier concentration at the free-space operating

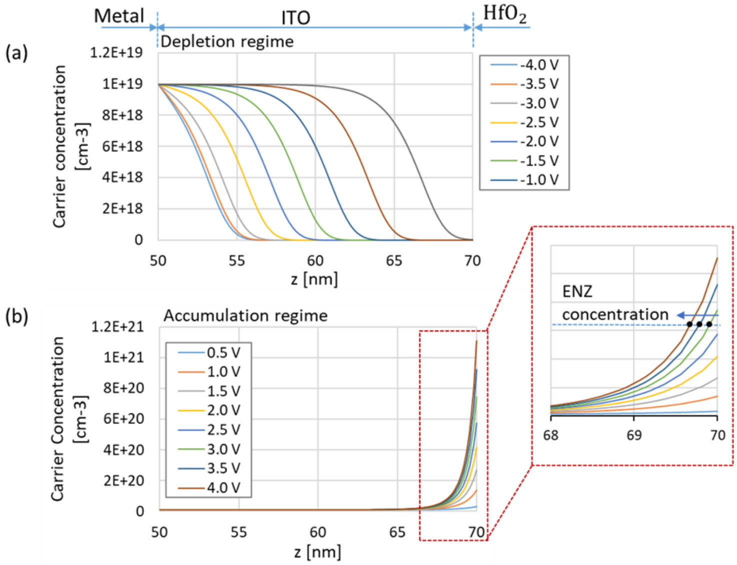

Fig. 4. 1D electron concentration distributions along the ITO thickness for different applied voltages in (a) the depletion and (b) accumulation regimes.

wavelength of $\lambda_0 = 1550$ nm (our design wavelength), where the real part of permittivity becomes zero (the ENZ point and its relation to the carrier concentration is discussed in greater detail in the next section). As the voltage increases, the ENZ point shifts away from the HfO_2-ITO interface into the bulk ITO. On one side of the ENZ point the carrier concentration is greater than the ENZ concentration and ITO appears metallic, whereas on the other side it appears dielectric.

Next, we plot the potential distribution inside the HfO_2 and ITO for different values of applied voltage, as shown in Fig. 5. The maximum voltage that can be applied is limited by the breakdown field of HfO_2, and is determined by the potential drop across this layer. For example, at an applied voltage of 4 V, the voltage drop across HfO_2 is ~2.5 V, which is lower than the breakdown voltage of 3.2 V for 5 nm thick HfO_2. Thus, to evaluate the performance of our design, we keep the maximum applied voltage to +4 V to keep the voltage drop across HfO_2 lower than the breakdown voltage. The values of potential plotted in Fig. 5 are referenced to the intrinsic Fermi level.

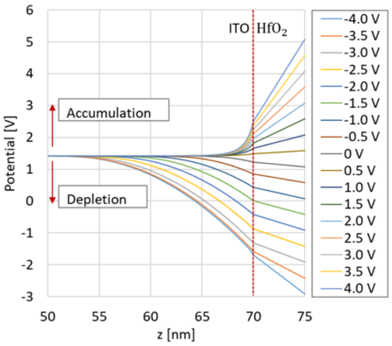

Fig. 5. Potential distribution within the ITO and HfO_2 layers for different applied voltages.

In the next section, we use the 2D electron profiles obtained in this section at different values of applied voltage (*cf.* Figure 3) to obtain corresponding refractive index profiles which are then used in the optical mode calculations.

3.2. Modal analysis of the active section

To transform the charge distribution in the ITO film obtained via the electrostatic simulations to corresponding relative complex permittivity profiles, we use the Drude equation [16]:

$$\varepsilon_{\mathrm{r}}(\omega, y, z) = \varepsilon_{\infty} - \frac{\omega_{\mathrm{p}}^2(y, z)}{\omega^2 + i\gamma\omega} \tag{1}$$

with parameters representative of ITO: $\varepsilon_{\infty} = 4.2345$ and $\gamma = 1.7588 \times 10^{14}$ rad/s. The distribution of the plasma frequency is given by $\omega_{\mathrm{p}}(y, z) = N(y, z)\mathrm{e}^2/\varepsilon_0 m_{\mathrm{n}}^*$, where m_{n}^*=0.35m_{e} is the effective mass of electrons in ITO. $N(y, z)$ is the electron density distribution obtained from the electrostatic modelling (*cf.* Fig. 3), e is the electron charge, and ε_0 is the vacuum permittivity. At the wavelength of 1550 nm and a carrier concentration of 6.5×10^{20} cm^{-3}, the real part of epsilon becomes zero. As the carrier concentration crosses this value, the real part of epsilon changes sign, so the material behaviour changes from dielectric to metallic. The permittivity of gold is also represented using the Drude model with $\varepsilon_{\infty} = 8.4156$, $\gamma = 4.8257 \times 10^{13}$ rad/s and $\omega_{\mathrm{p}} = 1.4117 \times 10^{16}$ rad/s [24]. HfO_2 is taken as a lossless dielectric of refractive index 2.0709 [25]. For SiO_2 and Si, the refractive indices of 1.4440 and 3.4752 were obtained at 1550 nm from a polynomial fit to the Palik data [26].

We perform modal analyses of the full device cross-section (Fig. 2) based on a vectorial finite difference eigenmode (FDE) technique [20]. When g is large, the system of Fig. 2 presents three uncoupled waveguides (two MIM and one Si), supporting distinct MIM and Si waveguide modes. As g decreases the system becomes coupled, the modes hybridise and evolve into coupled modes. We investigate the modes for the case where g is small, corresponding to the active modes of the modulator. We present results for two cases of Si core dimensions $w \times h$: 500 nm × 300 nm (Design I) and 350 nm × 500 nm (Design II).

The fundamental modes of the active section are the TM symmetric ($\mathrm{TM_s}$) and TM antisymmetric ($\mathrm{TM_a}$) supermodes, both strongly confined within the MIM structures, but distributed symmetrically and asymmetrically laterally along y, over the gap $g = 200$ nm separating the left and right MIMs, as plotted in Fig. 6(a) for Design I. In Fig. 6(a), we show the dominant transverse electric field component (E_z) of these modes for an applied voltage of 0.25 V, which is near the flat-band voltage. The field is strongly confined within the MIM structures, overlapping essentially perfectly with the HfO_2 and ITO regions, thus allowing strong interaction with the perturbed ITO layer, ultimately resulting in strong modulation.

In Fig. 6(b), we show the modal profile of the $\mathrm{TM_s}$ mode near the hafnia-ITO interface, for applied voltages of 2.5, 3.0, 3.5, and 4.0 V in the accumulation regime. Highly localised enhancement of the E_z component is observed at the ENZ location compared to the flat-band case of Fig. 6(a). Since the ENZ point shifts away from the HfO_2-ITO interface into the bulk ITO with an increase in V_{s} (*cf.* Fig. 4(b)), the peak of E_z follows. (There is no ENZ point for $V_s = 3.0$ V, but the real part of permittivity is small and close zero, thereby still producing significant field localisation and enhancement).

Next, we plot the variation in the complex effective indices of these modes ($\mathrm{TM_s}$ and $\mathrm{TM_a}$) as a function of applied voltage for Designs I and II, in Figs. 7(a) and 7(b), respectively. The variation in the imaginary part of the complex effective index is smaller for values of V_{s} in the depletion regime (V_s<0.334 V). As the applied voltage drops below 0.334 V, the imaginary part becomes almost constant, as the depletion region in ITO reaches its maximum depth. In the accumulation regime (V_s>0.334 V), the imaginary part increases by more than 4 × over the voltage range investigated for both modes in both designs. In the accumulation regime, the most

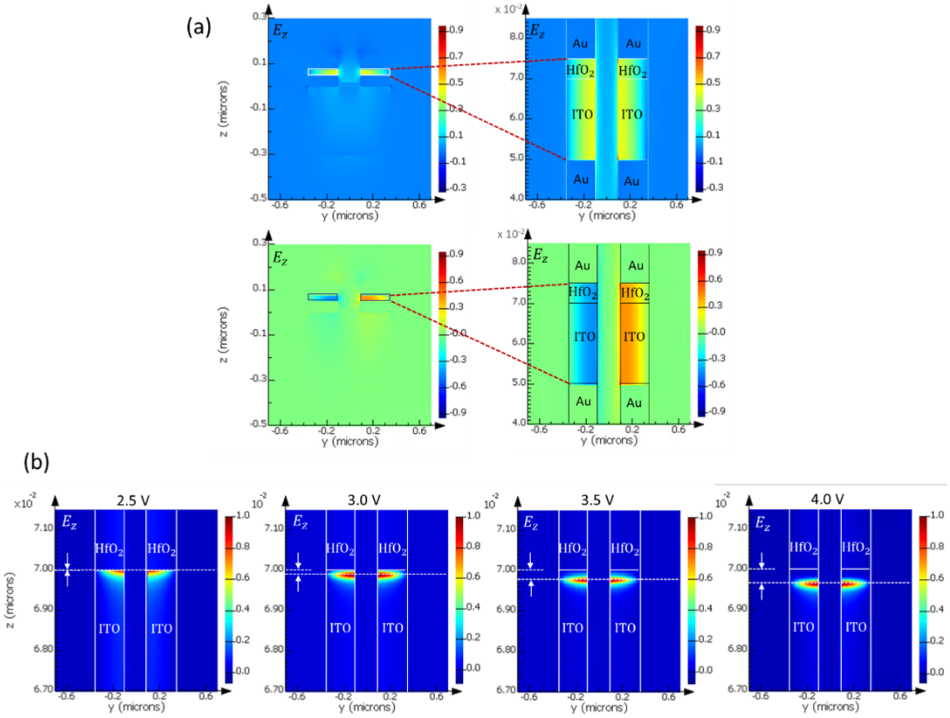

Fig. 6. (a) Modal profiles of the TM$_s$ (top) and TM$_a$ (bottom) modes of the active section for $V_s = 0.25$ V, and (b) modal profile of the TM$_s$ mode for $V_s = 2.5$, 3.0, 3.5, and 4.0 V in the accumulation regime, plotted in expanded view to show the shift in the position of max$|E_z|$.

significant change in the imaginary part occurs between 2 and 3 V due to the emergence of the ENZ point in this voltage range.

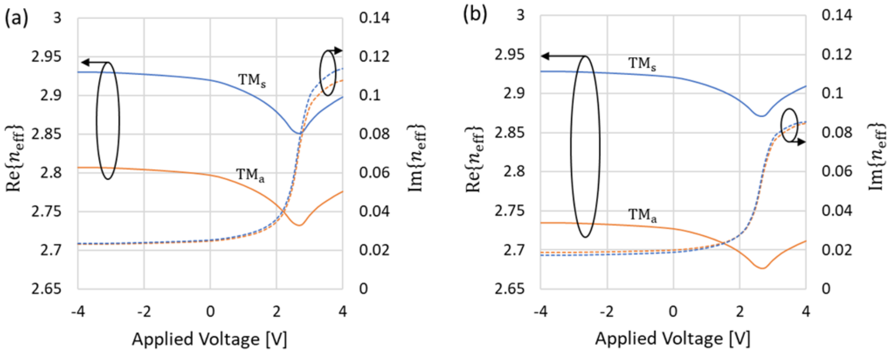

Fig. 7. Complex effective indices of the TM$_s$ and TM$_a$ modes of the active section as a function of applied voltage for (a) Design I, and (b) Design II.

For Design I, the complex effective indices of the TM$_s$ mode are $2.9142 - j0.02646$ and $2.8981 - j0.1140$ for applied voltages of 0.5 and 4.0 V, resulting in an insertion loss $\mathrm{IL}_{\mathrm{Active}} = 0.9$ dB/µm at 0.5 V, and an extinction ratio of $ER = 3.1$ dB/µm. The corresponding complex effective indices of the TM$_s$ mode of Design II are $2.9168 - j0.01958$ and $2.9095 - j0.08536$ resulting in $\text{IL}_{\text{Active}} = 0.7$ dB/μm and ER = 2.3 dB/μm. Design I produces a higher extinction ratio and insertion loss due to the slightly stronger confinement of the modal field in the MIM structures. The AC drive voltage required to achieve this performance in both designs is $V_s = \pm 1.75$ V, about a DC bias voltage of 2.25 V.

3.3. Modal evolution and hybridization

In plasmonic ITO modulators, the efficient and selective excitation of the plasmonic mode of the active section is an important design consideration. Improper excitation results in issues such as a high insertion loss due to loss of power into other modes including radiative modes, low extinction ratio due to forward propagating unmodulated modes, or interference between multiple modes excited at the input of the active section.

To achieve efficient and selective excitation of the modes in the active section, we design an adiabatic taper which transforms the TM_0 and TE_0 modes of the input silicon waveguide into the TM_s and TM_a plasmonic supermodes of the active section, respectively. To show how modal transformation occurs, we plot the real part of the effective index of the modes as a function of g for Designs I and II in Fig. 8. As we decrease the gap from $g = 1200$ nm, the active region couples to the Si waveguide as the two MIM stacks approach, and the modes of the silicon waveguide transform into the corresponding plasmonic supermodes of the active section. As observed from Fig. 8, the largest change in $\text{Re}\{n_{\text{eff}}\}$ occurs for $g<w$ ($w = 500$ nm for Design I, $w = 350$ nm for Design II), indicating that most of the transformation occurs for $g<w$. For $g = 1200$ nm $\gg w$, the effective indices of the modes are those of the corresponding TM_0/TE_0 modes of the input silicon waveguide. On the left side of the plots, we show the modal profiles of the TM_s and TM_a modes at $g = 200$ nm, and on the right, we show the modal profiles of the corresponding TM_0 and TE_0 modes at $g = 1200$ nm.

It is intuitively clear that the TM_0 mode of the Si waveguide should transform into the TM_s supermode of the MIM pair, because the E_z field is the dominant transverse field component of both modes and it is distributed symmetrically along the lateral dimension. However, that the TE_0 mode of the Si waveguide transforms into the TM_a mode of the MIM pair is surprising given that the dominant transverse field component of the former is E_y and orthogonal to that of the later which is E_z. We recall that modes in rectangular dielectric waveguides are not strictly TE or TM as in a dielectric slab - all field components are present in all modes, including the E_z component of the TE_0 mode. Indeed, this field component is asymmetrically distributed along the lateral dimension, as shown in Fig. 8, and although weak, enables transformation to the TM_a mode of the MIM pair. Thus, the mode pairs (TM_0, TM_s) and (TE_0, TM_a) evolve smoothly with g, implying that a taper can be designed to adiabatically transform them.

The taper is created by decreasing the gap g between the two MIM stacks, linearly from 1000 nm at the input of the taper, to 200 nm in the active section, over a length L_T which should be chosen to ensure adiabatic mode transformation. To evaluate the taper performance and optimise its design, we use the eigenmode expansion (EME) technique [20].

In Fig. 9 we plot the power in the $\text{TM}_\text{s}/\text{TM}_\text{a}$ plasmonic supermodes at the output of the tapered section due to excitation at the input by the TM_0/TE_0 modes of the silicon waveguide, as a function of the taper length L_T. The curves show a maximum at a specific value of L_T. At short lengths, loss is due to power coupling into other modes (including radiative) due to non-adiabatic modal evolution, and at long lengths propagation loss dominates. There is always a trade-off between them, and thus an optimum length exists minimising the overall loss. The powers were normalised such that the output power plotted in Fig. 9 can be interpreted as the transmittance of the tapers for the corresponding mode pairs (TM_0, TM_s) and (TE_0, TM_a). For Design I, the taper transmittance maximizes at about 0.6 for $L_T = 6$ μm for both mode pairs. For Design II, the taper transmittance maximizes slightly above 0.7 for $L_T = 11$ μm for mode pair (TE_0, TM_a), and at 0.73 for $L_T = 15$ μm for mode pair (TM_0, TM_s). In the case of Design I, the same taper

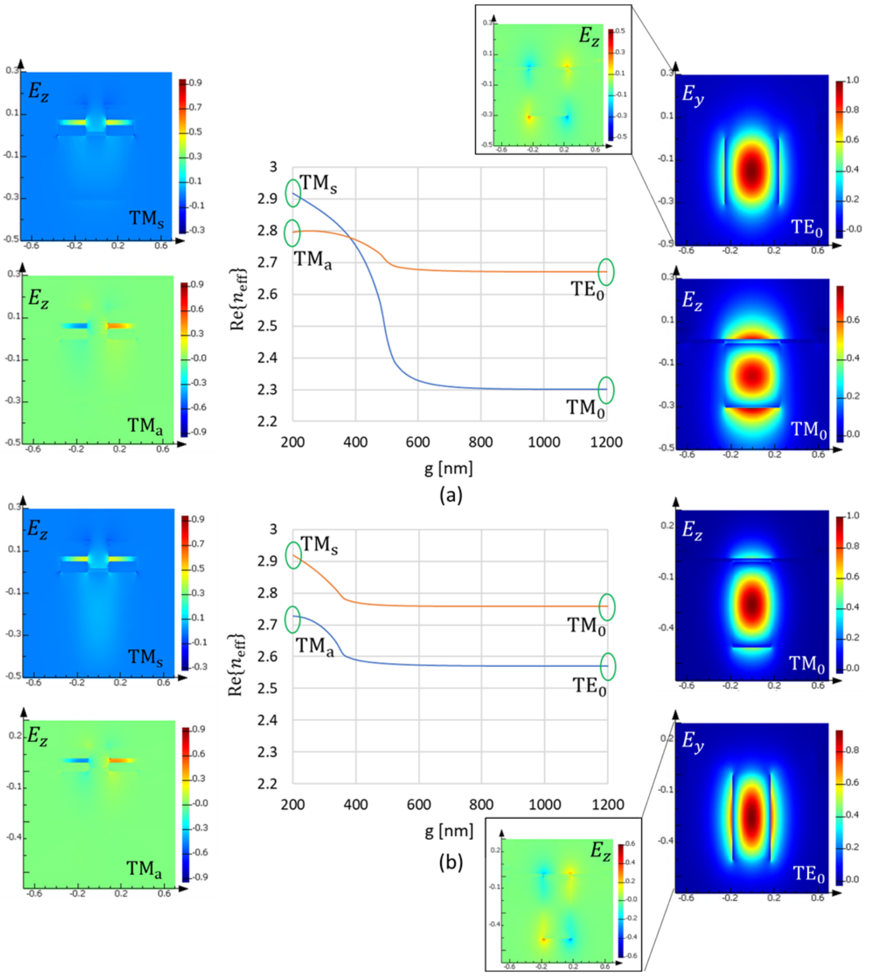

Fig. 8. Real part of mode effective index as a function of g for (a) Design I and (b) Design II. The corresponding field distributions are shown on the left for g = 200 nm and on the right for g = 1200 nm (the E_z field component is also plotted on the right for the TE$_0$ mode).

could be used to operate optimally with both mode pairs, whereas for Design II, a good taper length would be L_T = 11 µm, yielding an equal transmittance of about 0.7 for both mode pairs.

In Fig. 10(a), we show evolution of the (TM$_0$, TM$_s$) mode pair through a tapered section of length L_T = 15 µm, and in Fig. 10(b) of the (TE$_0$, TM$_a$) mode pair for L_T = 11 µm (both cases for Design II). We plot the field profiles of the dominant components on two planes: one passing through the vertical centre of the silicon core (left panels), and one passing through the vertical centre of the ITO (right panels). From these plots we confirm that the TM$_0$ mode input at g = 1000 nm gets transformed into the TM$_s$ mode at g = 200 nm, whereas the TE$_0$ mode input at g = 1000 nm gets transformed into the TM$_a$ mode at g = 200 nm. We also notice the expected enhancement in the field (the colour bars can be directly compared), as the TM$_{s,a}$ modes are strongly confined within the HfO$_2$ and ITO regions of the MIM structures.

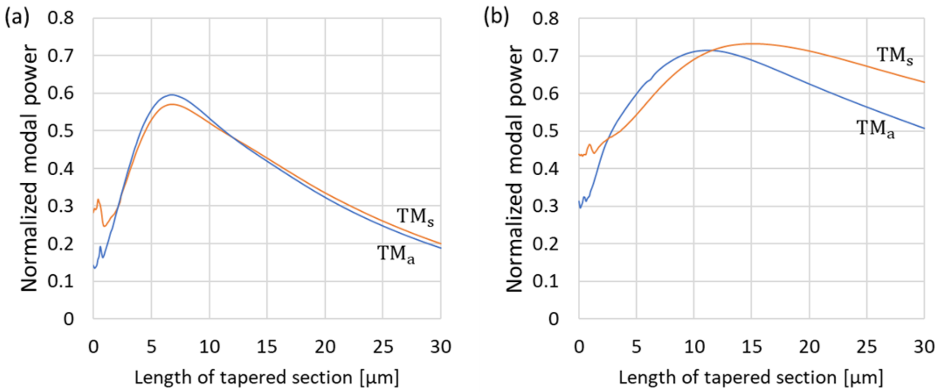

Fig. 9. Normalised output power (transmittance) in the $\mathrm{TM_s}$ and $\mathrm{TM_a}$ plasmonic supermodes at $g = 200$ nm *vs.* the length of the tapered section L_T, due to excitation at the input ($g = 1000$ nm) by the $\mathrm{TM_0}$ and $\mathrm{TE_0}$ modes of the silicon waveguide. (a) Design I, and (b) Design II.

3.4. Electrical isolation of the active section

The tapered sections at the input and output of the active section are strictly passive and must be isolated electrically to maximise the modulator bandwidth. Specifically, it is undesirable for the tapered sections to contribute capacitance or series resistance to the modulator.

To electrically isolate the active section, we create slits of width s_w in the top metal film comprising the MIM sections, perpendicular to the direction of propagation, at the junction of the taper and the active section, as shown in Fig. 1. These slits electrically isolate the active section, as they ensure that the drive voltage (V_s) is applied thereon only. However, there is an

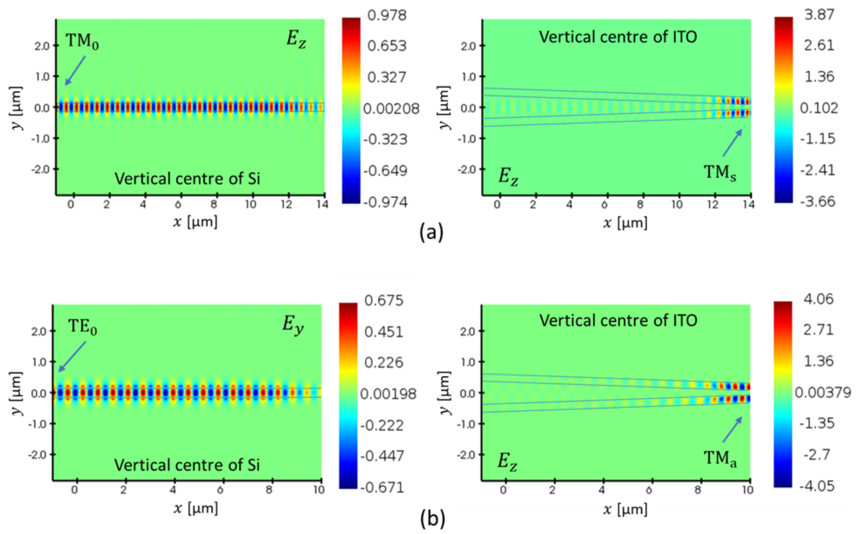

Fig. 10. Electric field profiles of the dominant component on two planes, one passing through the vertical centre of the silicon core (left panels), and one passing through the vertical centre of the ITO (right panels), showing propagation through tapered sections for Design II, where g changes from 1000 nm to 200 nm. (a) Input $\mathrm{TM_0}$ mode transforms adiabatically into the output $\mathrm{TM_s}$ mode for $L_\mathrm{T} = 15$ µm. (b) Input $\mathrm{TE_0}$ mode transforms adiabatically into $\mathrm{TM_a}$ mode for $L_T = 11$ µm.

insertion loss (IL_{slit}) associated with this discontinuity in the propagation direction. In Figs. 11(a) and 11(b), we plot IL_{slit} computed using the FDTD technique as a function of slit width (s_w) for Designs I and II. As s_w is reduced, IL_{slit} is also reduced. For s_w = 10 nm in Design I, IL_{slit} = 0.37 dB for both TM_s and TM_a supermodes. For Design II, IL_{slit} = 0.24 dB for the TM_s mode and 0.33 dB for the TM_a mode. IL_{slit} accounts for a pair of slits through both MIM structures along the same plane y-z plane. Slits are needed at input and output planes of the active section, so IL_{slit} must be doubled when computing the full insertion loss of the device.

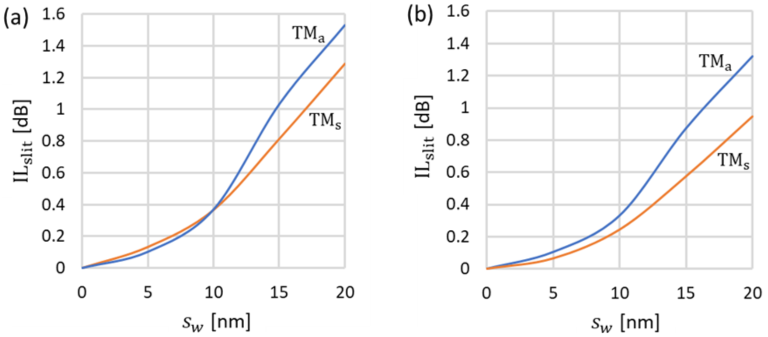

Fig. 11. Insertion loss of isolation slits, IL_{slit}, *vs.* s_w for (a) Design I and (b) Design II.

3.5. Electrical bandwidth

As the active device length is very small, the active section of the modulator can be modelled as the lumped RC network shown as the inset to Fig. 12. The 3-dB electrical bandwidth associated with this model is given by [27]:

$$\mathrm{BW} = \frac{1}{2\pi\left[R + \left(\frac{1}{R_S} + \frac{1}{R_L}\right)^{-1}\right]C} \tag{2}$$

where R is the parasitic series resistance from ground to signal through the modulator assuming that the insulator (HfO_2) is removed. The source and load resistances are represented by R_S and R_L, and set by the system designer. The capacitance of the modulator is represented by C, which varies as a function of applied voltage and is maximum under strong accumulation:

$$C = \varepsilon_0 \varepsilon_{r,ox} \frac{A}{h_{ox}} \tag{3}$$

where $\varepsilon_{r,ox}$ is the relative permittivity of the oxide (HfO_2), A is the area of both MIM structures in the active section, and h_{ox} represents the thickness of the oxide layer (HfO_2).

We performed 2D electrostatic modelling (*cf.* Sub-section 3.1) to obtain the total charge per unit length in the simulation domain comprising both MIM structures, as a function of applied voltage, which we plot in Fig. 12. We then used these results to calculate the capacitance per unit length of the modulator, as the derivative of the total charge with respect to applied voltage. The capacitance per unit length in the strong accumulation regime is 18 fF/µm, which matches well with the oxide capacitance calculated using the parallel plate capacitor model of Eq. (3). The capacitance (C) of the modulator is then simply obtained by multiplying this capacitance per unit length by the active length (L_{active}).

To calculate the parasitic series resistance, we set the thickness of the oxide layer to zero and calculate via 2D electrostatic modelling the current density flowing through the modulator

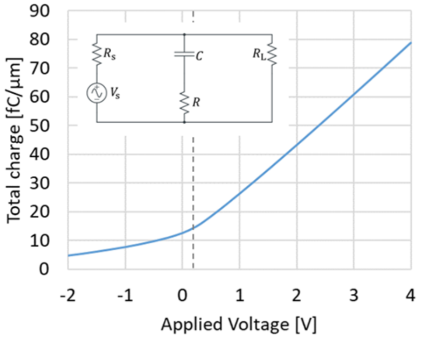

Fig. 12. Total charge per unit length of the modulator active section *vs.* applied voltage. The inset gives a sketch of the lumped-element equivalent circuit of the modulator.

incorporating both MIM structures for a specific applied voltage. Using this data, we obtain the series resistance per unit length of the modulator as 5.4 Ω·µm. The series resistance is dominated by the resistance through the ITO layer, which decreases with modulator length as it is inversely proportional to the modulator area A. The series resistance of the modulator R is then obtained by dividing the resistance per unit length by the active length (L_{active}).

To estimate the 3-dB electrical bandwidth, we assume a source of impedance $R_S = 50\ \Omega$ and a load resistance R_L of 50 Ω, resulting in a simplified expression for the bandwidth (Eq. (2)): $\text{BW} = [2\pi(R+25)C]^{-1}$. Using the calculated values of capacitance and resistance, we obtain the corresponding bandwidths tabulated in Table 1 for various design options.

Table 1. Summary of modulator results.

Si core, $w \times h$ [nm × nm]	L_{active} [µm]	Mode Pair	$\text{IL}_{\text{active}}$ [dB]	IL_{slit} [dB]	IL_{taper} [dB]	IL_{total} [dB]	ER [dB]	BW [GHz]
Design I (500 × 300)	1.5	$\text{TM}_0, \text{TM}_\text{s}$	1.40	0.37	2.44	7.02	4.62	206
		$\text{TE}_0, \text{TM}_\text{a}$	1.35	0.37	2.25	6.59	4.35	206
	2.0	$\text{TM}_0, \text{TM}_\text{s}$	1.86	0.37	2.44	7.48	6.16	160
		$\text{TE}_0, \text{TM}_\text{a}$	1.80	0.37	2.25	7.04	5.80	160
Design II (350 × 500)	1.5	$\text{TM}_0, \text{TM}_\text{s}$	1.03	0.24	1.35	4.21	3.47	206
		$\text{TE}_0, \text{TM}_\text{a}$	1.09	0.33	1.46	4.67	3.82	206
	2.0	$\text{TM}_0, \text{TM}_\text{s}$	1.38	0.24	1.35	4.56	4.63	160
		$\text{TE}_0, \text{TM}_\text{a}$	1.46	0.33	1.46	5.04	4.51	160

3.6. Summary of designs

We summarize in Table 1 several important modulator characteristics, IL, ER, and BW, for Designs I and II, and for both modes pairs of operation. The total insertion loss (dB) of the design is: $\text{IL}_{\text{total}} = 2 \times \text{IL}_{\text{taper}} + \text{IL}_{\text{active}} + 2 \times \text{IL}_{\text{slit}}$ where IL_{taper} (dB) is the insertion loss of each tapered section at the input and output, $\text{IL}_{\text{active}}$ (dB) is the insertion loss of the active section, and IL_{slit} (dB) is the insertion loss due to one slit pair at the junction of a tapered section and the active section.

The total insertion loss of Design I ($\text{IL}_{\text{total}} \sim 7$ dB) is larger than for Design II ($\text{IL}_{\text{total}} \sim 4.5$ dB), mostly because the insertion loss of the tapered sections (IL_{taper}) is larger in the former. The insertion loss of the modulator is dominated by the taper losses, especially as two tapers are needed per modulator, one at the input of the active section and another at the output. However, Design I provides a larger ER compared to Design II.

Two cases of length for the active section are considered to illustrate the trade-off between ER and BW. For $L_{\text{active}} = 1.5$ µm, the capacitance and series resistance of the modulator are $C = 27$ fF and $R = 3.6\ \Omega$ yielding an electrical BW of 206 GHz - in this case, Design I yields ER ~ 4.5 dB for operation with both mode pairs. For $L_{\text{active}} = 2$ µm, the capacitance and series resistance of the modulator are $C = 36$ fF and $R = 2.7\ \Omega$ yielding an electrical BW of 160 GHz - in this case, Design I yields ER ~ 6 dB for operation with both mode pairs. The overall length of the designs, including the active length and both tapers, range from 13 to 24 µm.

We note from Table 1 that the performance of mode pair ($\text{TM}_0, \text{TM}_\text{s}$) is very similar to the performance of mode pair ($\text{TE}_0, \text{TM}_\text{a}$) within the same design. This is due to the facts that the plasmonic supermodes (TM_s, TM_a) behave similarly in the active section in terms of $\text{Im}\{n_{eff}\}$ (*cf.* Figure 7), and the modes of the silicon waveguide (TM_0 and TE_0) transform similarly into the corresponding plasmonic supermodes (*cf.* Figures 8 - 10). Thus, relative to the mode of the input silicon waveguide (TM_0 or TE_0) the modulator performance is polarization independent.

In Table 2, we compare our design with other plasmonic modulators including different designs of plasmonic ITO modulators. We also include GeSi [28], Ge [29] and graphene [30] based electro-absorption modulators. For comparison, we use specifications of importance to assess the performance of modulators, such as $\text{IL}_{\text{Active}}$, L_{Active}, ER, electrical bandwidth, drive voltage, and supported polarizations.

Table 2. Comparison with other modulators[a]

Material [Reference]	$\text{IL}_{\text{Active}}$ [dB]	ER [dB]	BW [GHz]	L_{Active} [µm]	Drive [V]	Configuration	Polarization
ITO [9]	2.24	3	1.1	1.4	20	HPW + MZI	Single (TM)
ITO [10]	1	5	<50	5	5	HPW + EA	Single (TM)
NLP [13]	>6	>15	>200	~10	12	MIM + MZI	Single (TE)
ITO [12]	2	2	NA	4	4	EA	Single (TM)
GeSi [28]	10.6	5.2	56	40	4	EA	Single (TE)
Ge [29]	6.2	3.6	36	40	3	EA	Single (TE)
Graphene [30]	2.2	16	29	120	10	EA	Single (TE)
ITO [This work]	1.4	4.5	206	1.5	3.5	MIM + EA	Both (TE and TM)

[a] ITO: Indium Tin, Oxide; NLP: Nonlinear Polymer; HPW: Hybrid Plasmonic Waveguide; MIM: metal insulator metal; EA: Electro-Absorption; MZI: Mach-Zehnder Interferometer.

4. Conclusions

We propose a broadband plasmonic electro-optic modulator operating at $\lambda_0 = 1550$ nm, exploiting the free carrier dispersion effect in ITO to access its ENZ point within the accumulation layers of MOS structures integrated into a pair of coupled MIM waveguides aligned on a planarized Si waveguide. The coupled MIM waveguides support symmetric and asymmetric plasmonic supermodes, excited adiabatically by the TM_0 and TE_0 modes of the underlying Si waveguide, respectively.

Through a combination of strong index modulation of ITO near its ENZ point, and a strong interaction of the propagating plasmonic supermodes with the perturbed ITO, the active section of the modulator can be extremely compact, of length between 1.5 and 2.0 µm, such that it may be driven as a simple capacitive lumped element, with a broad electrical bandwidth ranging between 160 and 200 GHz, and an extinction ratio in the range of 3.5 to 6 dB. The modulator section is isolated electrically from passive mode conversion tapers at the input and output which adiabatically transform the input TM_0 and TE_0 modes of the Si waveguide to the symmetric and asymmetric plasmonic supermodes, TM_s and TM_a, of the modulator. The total insertion loss of the modulators, including the mode conversion sections, ranges from 4 to 7.5 dB. The AC drive voltage required is ±1.75 V about a DC bias of 2.25 V. The modulators make use of inorganic materials only and are essentially polarization independent.

Funding. Huawei Technologies Canada.

Acknowledgment. Fruitful discussions with Dominic Goodwill, formerly of Huawei Technologies Canada, and with Eric Bernier of Huawei Technologies Canada are gratefully acknowledged.

Disclosures. The authors declare no conflicts of interest.

Data availability. All data underlying the results or needed to evaluate the conclusions of the paper are present in the paper.

References

1. A. Rahim, A. Hermans, B. Wohlfeil, D. Petousi, B. Kuyken, D. Van Thourhout, and R. G. Baets, “Taking silicon photonics modulators to a higher performance level: state-of-the-art and a review of new technologies,” Adv. Photonics **3**(02), 024003 (2021).
2. J. Witzens, “High-speed silicon photonics modulators,” Proc. IEEE **106**(12), 2158–2182 (2018).
3. D. Thomson, A. Zilkie, J. E. Bowers, T. Komljenovic, G. T. Reed, L. Vivien, D. Marris-Morini, E. Cassan, L. Virot, J. M. Fédéli, and J. M. Hartmann, “Roadmap on silicon photonics,” J. Opt. **18**(7), 073003 (2016).
4. G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. Thomson, “Silicon optical modulators,” Nat. Photonics **4**(8), 518–526 (2010).
5. A. Melikyan, L. Alloatti, A. Muslija, D. Hillerkuss, P. C. Schindler, J. Li, R. Palmer, D. Korn, S. Muehlbrandt, D. Van Thourhout, B. Chen, R. Dinu, M. Sommer, C. Koos, M. Kohl, W. Freude, and J. Leuthold, “High-speed plasmonic phase modulators,” Nat. Photonics **8**(3), 229–233 (2014).
6. S. Wolf, *Silicon-organic hybrid (SOH) electro-optic modulators for high-speed and power-efficient communications*, KIT Scientific Publishing, Vol. 23, 2018.
7. P. Guo, A. M. Sarangan, and I. Agha, “A review of germanium-antimony-telluride phase change materials for non-volatile memories and optical modulators,” Appl. Sci. **9**(3), 530 (2019).
8. M. Liu, X. Yin, E. Ulin-Avila, B. Geng, T. Zentgraf, L. Ju, F. Wang, and X. Zhang, “A graphene-based broadband optical modulator,” Nature **474**(7349), 64–67 (2011).
9. R. Amin, R. Maiti, Y. Gui, C. Suer, M. Miscuglio, E. Heidari, R.T. Chen, H. Dalir, and V. J. Sorger, “Sub-wavelength GHz-fast broadband ITO Mach–Zehnder modulator on silicon photonics,” Optica **7**(4), 333–335 (2020).
10. R. Amin, J. K. George, S. Sun, T. Ferreira de Lima, A. N. Tait, J. B. Khurgin, M. Miscuglio, B. J. Shastri, P. R. Prucnal, T. El-Ghazawi, and V. J. Sorger, “ITO-based electro-absorption modulator for photonic neural activation function,” APL Mater. **7**(8), 081112 (2019).
11. R. Amin, R. Maiti, Y. Gui, C. Suer, M. Miscuglio, E. Heidari, J. B. Khurgin, R. T. Chen, H. Dalir, and V. J. Sorger, “Heterogeneously integrated ITO plasmonic Mach–Zehnder interferometric modulator on SOI,” Sci. Rep. **11**(1), 1287 (2021).
12. M. Tahersima, Z. Ma, Y. Gui, S. Sun, H. Wang, R. Amin, H. Dalir, R. Chen, M. Miscuglio, and V. Sorger, “Coupling-enhanced dual ITO layer electro-absorption modulator in silicon photonics,” Nanophotonics **8**(9), 1559–1566 (2019).
13. M. Ayata, Y. Fedoryshyn, W. Heni, B. Baeuerle, A. Josten, M. Zahner, U. Koch, Y. Salamin, C. Hoessbacher, C. Haffner, and D. L. Elder, “High-speed plasmonic modulator in a single metal layer,” Science **358**(6363), 630–632 (2017).
14. C. Hössbacher, A. Josten, B. Baeuerle, Y. Fedoryshyn, H. Hettrich, Y. Salamin, W. Heni, C. Haffner, C. Kaiser, R. Schmid, and D. L. Elder, “Plasmonic modulator with> 170 GHz bandwidth demonstrated at 100 GBd NRZ,” Opt. Express **25**(3), 1762–1768 (2017).
15. G. K. Shirmanesh, R. Sokhoyan, R.A. Pala, and H. A. Atwater, “Dual gated active metasurface at 1550 nm with wide (>300∘) phase tunability,” Nano Lett. **18**(5), 2957–2963 (2018).
16. A. Calà Lesina, D. Goodwill, E. Bernier, L. Ramunno, and P. Berini, “Tunable Plasmonic Metasurfaces for Optical Phased Arrays,” IEEE J. Select. Topics Quantum Electron. **27**(1), 1–16 (2021).
17. P. Berini, “Optical Beam Steering Using Tunable Metasurfaces,” ACS Photonics **9**(7), 2204–2218 (2022).
18. P. Berini, “Figures of merit for surface plasmon waveguides,” Opt. Express **14**(26), 13030–13042 (2006).
19. C. C. Lin, P. Chang, Y. Su, and A. S. Helmy, “Monolithic plasmonic waveguide architecture for passive and active optical circuits,” Nano Lett. **20**(5), 2950–2957 (2020).
20. Ansys Inc., https://www.lumerical.com/ansys/.
21. G. D. Wilk, R. M. Wallace, and J. M. Anthony, “High-κ gate dielectrics: Current status and materials properties considerations,” J. Appl. Phys. **89**(10), 5243–5275 (2001).
22. J. Robertson, “High dielectric constant gate oxides for metal oxide Si transistors,” Rep. Prog. Phys. **69**(2), 327–396 (2006).
23. M. Shabaninezhad, L. Ramunno, and P. Berini, “Tunable plasmonics on epsilon-near-zero materials: the case for a quantum carrier model,” Opt. Express **30**(26), 46501–46519 (2022).

24. K. M. McPeak, S. V. Jayanti, S. J. P. Kress, S. Meyer, S. Iotti, A. Rossinelli, and D. J. Norris, "Plasmonic films can easily be better: Rules and recipes," ACS Photonics **2**(3), 326–333 (2015).
25. D. L. Wood, K. Nassau, T. Y. Kometani, and D. L. Nash, "Optical properties of cubic hafnia stabilized with yttria," Appl. Opt. **29**(4), 604–607 (1990).
26. E. D. Palik, *Handbook of Optical Constants of Solids*, (Academic Press: Orlando, 1985).
27. A. Olivieri, C. Chen, S. Hassan, E. Lisicka-Skrzek, R. N. Tait, and P. Berini, "Plasmonic nanostructured metal-oxide-semiconductor reflection modulators," Nano Lett. **15**(4), 2304–2311 (2015).
28. L. Mastronardi, M. Banakar, A. Z. Khokhar, N. Hattasan, T. Rutirawut, T. Dominguez Bucio, K. M. Grabska, C. Littlejohns, A. Bazin, G. Mashanovich, and F. Y. Gardes, "High-speed Si/GeSi hetero-structure electro absorption modulator," Opt. Express **26**(6), 6663–6673 (2018).
29. Z. Liu, X. Li, C. Niu, J. Zheng, C. Xue, Y. Zuo, and B. Cheng, "56 Gbps high-speed Ge electro-absorption modulator," Photonics Res. **8**(10), 1648–1652 (2020).
30. M. A. Giambra, V. Sorianello, V. Miseikis, S. Marconi, A. Montanaro, P. Galli, S. Pezzini, C. Coletti, and M. Romagnoli, "High-speed double layer graphene electro-absorption modulator on SOI waveguide," Opt. Express **27**(15), 20145–20155 (2019).