

Research article

Rubab Amin, Can Suer, Zhizhen Ma, Ibrahim Sarpkaya, Jacob B. Khurgin, Ritesh Agarwal and Volker J. Sorger*

Active material, optical mode and cavity impact on nanoscale electro-optic modulation performance

<https://doi.org/10.1515/nanoph-2017-0072>

Received July 7, 2017; revised August 6, 2017; accepted August 25, 2017

Abstract: Electro-optic modulation is a key function in optical data communication and possible future optical compute engines. The performance of modulators intricately depends on the interaction between the actively modulated material and the propagating waveguide mode. While a variety of high-performance modulators have been demonstrated, no comprehensive picture of what factors are most responsible for high performance has emerged so far. Here we report the first systematic and comprehensive analytical and computational investigation for high-performance compact on-chip electro-optic modulators by considering emerging active materials, model considerations and cavity feedback at the nanoscale. We discover that the delicate interplay between the material characteristics and the optical mode properties plays a key role in defining the modulator performance. Based on physical tradeoffs between index modulation, loss, optical confinement factors and slow-light effects, we find that there exist combinations of bias, material and optical mode that yield efficient phase or amplitude modulation with acceptable insertion loss. Furthermore, we show how material properties in the epsilon near zero regime enable reduction of length by as much as by 15 times. Lastly, we introduce and apply a cavity-based

electro-optic modulator figure of merit, $\Delta\lambda/\Delta\alpha$, relating obtainable resonance tuning via phase shifting relative to the incurred losses due to the fundamental Kramers-Kronig relations suggesting optimized device operating regions with optimized modulation-to-loss tradeoffs. This work paves the way for a holistic design rule of electro-optic modulators for high-density on-chip integration.

Keywords: ITO; modulation; cavity; electro-optic; plasmonic.

1 Introduction

Photonic integrated circuits (PICs) are an important enabling technology in today's communication networks and in future chip-to-chip and potentially on-chip interconnects for information processing. The development of PIC has been historically hindered by the inherently weak interaction of light and matter, which requires fundamental size (footprint), speed, and power (S²P) limits for photonic-based information processing technologies. Among the active PIC components, the sources (lasers) and detectors have seen rapid improvement in their performance in the last decade utilizing high efficiency and low noise. But the third key active PIC component, light modulators, still remains far from achieving S²P characteristics.

While for a long time there has been no shortage of speculations about achieving all-optical control of signals in PICs, in the foreseeable future the electrical control of optical signals or, in short, electro-optic (EO) conversion, remains the only practical option [1]. Efficient light-matter interaction (LMI) within the active material is key to any optical modulation, and strong modulation requires active material that exhibits adequate interaction with the light [2, 3]. EO modulation can be enabled by either changing the real part (n) of the modal refractive index leading to phase shifting-based interferometer-like devices termed EO modulators (EOM) or by modulating the imaginary part (κ) of the modal index of linear electro-absorptive modulators (EAMs). In both types, the fundamental complex

*Corresponding author: Volker J. Sorger, Department of Electrical and Computer Engineering, George Washington University, 800 22nd St., Science and Engineering Hall, Washington, DC 20052, USA, e-mail: sorger@gwu.edu

Rubab Amin, Can Suer, Zhizhen Ma and Ibrahim Sarpkaya: Department of Electrical and Computer Engineering, George Washington University, 800 22nd St., Science and Engineering Hall, Washington, DC 20052, USA

Jacob B. Khurgin: Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA

Ritesh Agarwal: Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

index of refraction is altered electrically inside the active material, which in turn modifies the propagation constant of the mode inside the respective waveguide.

Electrically driven modulators are classified by the physical mechanism responsible for the change of the complex index. One can distinguish between current-driven and voltage-driven modulators; in current-driven modulators, the change in index is achieved when carriers are injected into (or removed from) the active region thus enabling (disabling) the optical transitions and therefore adding (subtracting) to the oscillator strength. Possible design options based on silicon, indium tin oxide (ITO) and graphene modulators considered and compared in this work are all classified as being current-driven. In contrast, for voltage-driven EOMs no current flows in/out of the active region, and the change in index is evoked by the energy level shifts and oscillator strength change caused by the electric field (e.g. Stark effect). The widely used lithium niobate (LiNbO_3) modulators based on Pockels effect and III-V semiconductor-based quantum-confined Stark effect modulators are two examples of voltage-driven EOMs [4–6]. Incidentally, the performance characteristics of both current-driven and voltage-driven modulators are essentially similar, except that the latter is bounded by limits set to avoid electrical breakdown in the active device region. This may be challenging, since realizing high fields can be difficult to achieve in regions that are only a few hundred nanometers thick. Given the present state of technology, almost all novel miniaturized modulators that have appeared on the scene over the last few years have been current-driven and are the subject and focus of this work [7–29].

The challenge facing EOMs is fundamental in nature; the Kramers-Kronig relations dictate that changing the real part of the complex index independent from simultaneously altering the imaginary part is impossible, unless one operates far from the resonance with any optical transition as is the case when the photon energy is deeply inside the bandgap of modulator material. Yet, both real and imaginary index changes are enhanced near the resonance and any attempt to increase the efficiency of modulation inevitably leads to the increased insertion loss (defined as a loss in the on (light-through) state). The index tuning-to-loss ratio is therefore a fundamental figure of merit (FOM) for phase shifting modulators and gets progressively worse near the resonance. Hence, most EOMs such as LiNbO_3 and polymers operate far from the resonance and naturally require long propagation lengths to incur sufficient (π) phase change [30, 31]. Furthermore, EOMs, as opposed to EAMs, always require some form of interferometric scheme [Mach-Zehnder (MZ), ring, etc] to

convert phase into amplitude modulation further increasing their footprint. Despite all these shortcomings, EOMs are indispensable when it comes to operating in advanced high bandwidth (i.e. speed) efficiency modulation formats like QPSK, QAM and others, but the advantage of these formats is significant only in relatively long-haul communication links, while here we concentrate on much shorter range interconnects, possibly for on-board on-chip links. The spurious phase modulation produced by the change in index causes frequency chirp, which is detrimental in long-haul links but remains relatively benign over the short distances considered here (<cm). In contrast, performance of miniature EAMs is not affected significantly by the change in the real part of refractive index that accompanies modulation of the absorption (loss).

1.1 Recent advances and developments

The highest FOM in terms of ability to achieve full switching with a minimum insertion loss indisputably belongs to the conventional EOMs based on LiNbO_3 and to a lesser extent to polymer-based devices. However, here low insertion loss is achieved at the cost of large footprint. To reduce the length, one can exploit two complimentary strategies of increasing the applied electric field by using the plasmonic modes with sub-wavelength confinement [30, 31] or using materials in which a large index change can be attained by using plasma dispersion. Both strategies result in increased insertion loss, but in many on-chip applications it can be tolerated as the necessary price to pay for the reduced footprint.

Of the materials relying upon the index change due to plasma dispersion, silicon is the most prominent candidate [8, 24, 25], but the index change in silicon is rather small [32]. Similar to silicon, transparent conducting oxides (TCOs), such as ITO, are able to tune their index via the carrier-dependent Drude model (i.e. they can be best described as current-driven modulators). Yet ITO holds a number of significant advantages over Si with respect to the index modulation. Firstly, the concentration of carriers in ITO can exceed concentration of carriers in Si by at least an order of magnitude; the concentration of indium atoms in ITO reaches a few percent, which is significantly above attainable donor concentration in Si. Secondly, the effect of changing the carrier concentration in ITO on its refractive index is more dramatic than in silicon (Figure 1C and D). This can be attributed to the higher bandgap and consequently lower refractive index of ITO compared to that of silicon. If the change of the carrier concentration δN_c (e.g. due to an applied voltage bias) causes a change

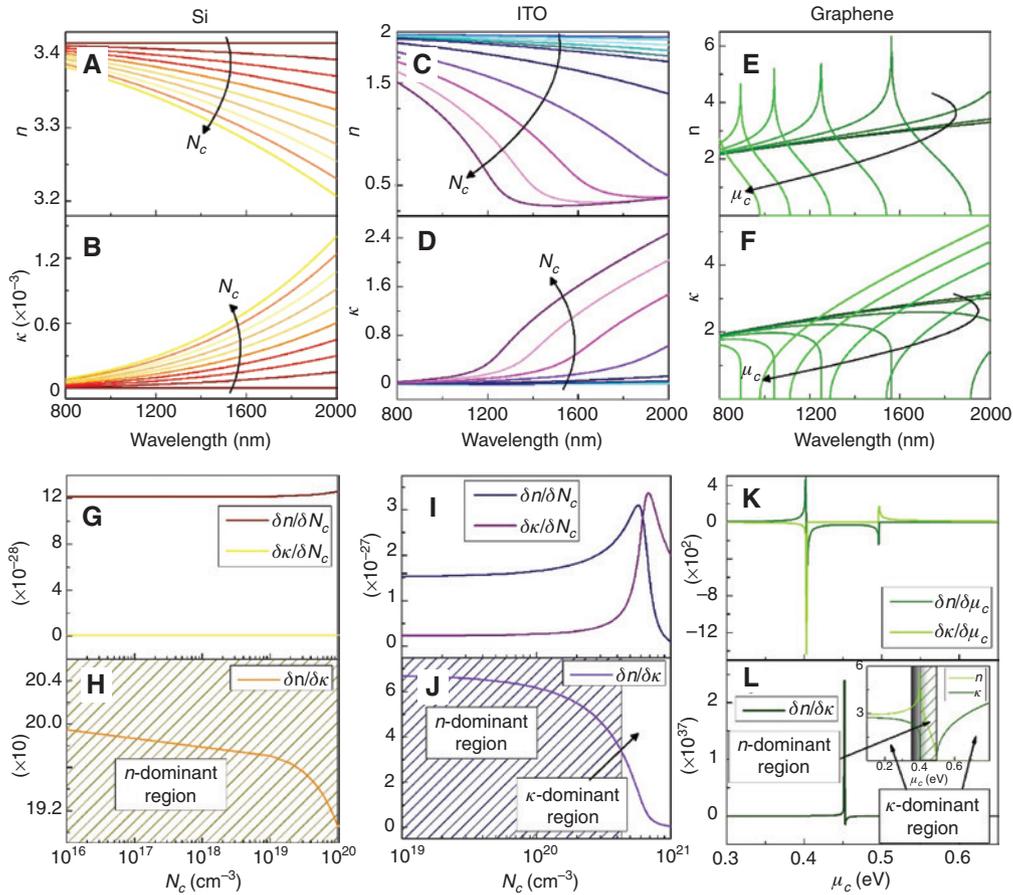


Figure 1: Material refractive indices and extinction coefficients vs. wavelength dispersion by varying carrier concentration, N_c for Si and ITO, and chemical potential, μ_c for graphene (A–F).

$\delta n/\delta N_c$, $\delta \kappa/\delta N_c$ and $\delta n/\delta \mu_c$ vs. carrier concentration, N_c (for Si & ITO) and chemical potential, μ_c (for graphene) at $\lambda = 1550$ nm (G–L). The n - k vs. chemical potential, μ_c for graphene is shown as an inset of the graphene $\delta n/\delta \mu_c$ plot (L).

in the relative permittivity (dielectric constant) $\delta \epsilon$, the corresponding change in the refractive index can be written as $\delta n = \delta \epsilon^{1/2} \sim \delta \epsilon / 2\epsilon^{1/2}$; hence, the refractive index change is greatly enhanced when the permittivity ϵ is small. The strongest modulation of both real and imaginary parts of refractive index occurs in the vicinity of $\epsilon \sim 0$, or “epsilon-near-zero” (ENZ) condition. Low “background” dielectric constant of ITO allows the calculation of ENZ carrier concentration at telecommunication wavelength of 1550 nm using experimentally obtainable parameters ($\sim 6 \times 10^{20} \text{ cm}^{-3}$) [28]. Thus, a stronger modulation per unity carrier change is expected for ITO compared to silicon with a unity order change in refractive index under electrical gating, as has been recently demonstrated [10, 11, 27]. The effect of ENZ will be explored further below (Figure 1I and J).

The accumulated charge-induced index change in ITO and other TCOs is based on the plasma (Drude) dispersion of free carriers and hence is an “off-resonance” effect. On the other hand, the index (and absorption) change can

also be induced by Pauli blocking of the interband (band-to-band) transitions in many materials such as found in gated graphene [12–14]. The carrier accumulation in graphene causes an upward (or downward if the carriers are holes) shift of the chemical potential and ensures a strong resonant change of both refractive index and extinction coefficient. This resonant change enables graphene to be suitable for both EO and electro-absorptive (EA) operations (Figure 1E, F, K, and L) up to the near infrared (IR) region of the spectrum [8, 9]. But the atomically thin thickness of single layer graphene produces a challenge for modulation, because of the miniscule optical confinement factor, Γ , and the requirement for the electric field of the propagating wave to be in plane. The latter restricts the use of plasmonic modes where field lines always connect normal to the metal surface. Thus, placing graphene onto a “regular” flat surface plasmon polariton (SPP) waveguide will not introduce a high field overlap [33]. To take full advantage of the strong modulation potential of n

and κ in graphene, it is critical to provide a suitable mode structure with high confinement factor, further discussed below.

A properly designed mode can provide not only a high overlap between the active material and the optical field (i.e. confinement factor Γ approaching 100%), but also an additional enhancement of the index (or extinction factor) change can be achieved by slow-light effects [34]. Such slow-light effect essentially increases the interaction time between the photons and active medium by a factor of n_g/n_{eff} where n_{eff} is the effective index of the mode and $n_g = \omega^{-1} \partial \omega / \partial n_{\text{eff}}$ is the group index. It is important to note that slow-light enhancement is conceptually similar to the enhancement achieved in the optical resonators where the photon-active medium interaction time is also enhanced but hence is subject to bandwidth limitations due to enhanced dispersion. Slow-light enhancements can be introduced by waveguide dispersion engineering [34], or matter-photon like modes (polaritonic), such as SPP modes. The latter are enabled by introducing the metal into the waveguide which can provide both high confinement factors and elevated group indices compared to photonic modes but are inherently lossy [1]. The latter is less critical for EAMs but can severely limit obtainable extinction ratios (ER, i.e. signal modulation depth) in EOMs. Thus, we do expect and indeed find an optimum LMI length for EOMs using plasmonics. Furthermore, the SPP modes can be hybridized with the dielectric waveguide modes creating the hybrid plasmonic modes where the losses can be reduced, evidently at the price of reduced confinement and group index.

Hence, the question arises: what are the details regarding index-to-loss in EO modulation, and do optimum device design parameters exist? In general, the modulator performance depends greatly on the underlying waveguide determining coupling and propagation losses, the confinement factor of the active material with the optical mode, the strength of the optical index change being altered [10, 35] and subsequent impacts on energy efficiency, modulation speed, footprint and optical power penalty [12, 36–39]. While previous works focused on addressing these in an ad hoc manner [10, 15, 16], here we show the first systematic approach for a selected set of active materials (silicon, ITO, and graphene), optical waveguide modes (bulk, slot, and hybrid-plasmon) and a cavity [Fabry-Pérot (FP)].

In the following discussion the nomenclature regarding the states of operation for the modulators relate to the light transmission characteristics through the device, rather than the applied voltage bias. As such the modulator's on state implies high light transmission, whereas the

off state signifies the lossy state corresponding to a drop in transmission. The discussion beyond Figure 1 is limited to a single operating wavelength of 1550 nm.

2 Steps in tuning

2.1 Material tuning properties

With the aim of evaluating the EO modulation potential of the active material in mind, we consider the tuning of the complex refractive index for our chosen materials, i.e. silicon, ITO, and graphene. The silicon Drude model is used to calculate the refractive index and extinction coefficient dispersion relations with respect to various wavelengths for a range of permitted carrier concentrations from 10^{16} to 10^{20} cm^{-3} (Figure 1A and B). Given our fixed operating wavelength at 1550 nm ($\hbar\omega = 0.8$ eV), corresponding to the telecom C band, the derivatives are taken for the real and imaginary components of the complex refractive index. Starting with silicon, from Figure 1A and B it is apparent that the real part of the index, n , changes notably with carrier concentration. However, while the relative change of the imaginary part of the index, κ , appears significant, the small absolute value of the off state (high carrier concentration) does not make silicon a high performing EAM material. This shows that silicon may be used for its EO properties, but it is not well suited for EA operations. This becomes evident when plotting $\delta n / \delta \kappa$; for higher carrier concentrations the slope kinks downwards for carrier concentrations above 10^{19} cm^{-3} (Figure 1H), which indicates the material becoming lossy but not κ -dominant or favorable for EA operations yet. Also, the gradual slope in $\delta n / \delta \kappa$ suggests weak EO modulation before this transition region making the entire carrier concentrations range n -dominant, which is analogous to favoring EO operations. Note that the material can be termed n - or κ -dominant, based on the carrier concentration regions where the change in one exceeds the change in the other corresponding to tuning.

The Drude model characterizes the ITO material accurately within our specific wavelengths of interest (near IR regime). Modulation is an effect of change in the refractive index and the broadening of the corresponding resonances inside an optical cavity (see below) with respect to wavelength which can be related back to the loss in the material absorption itself, corresponding to the extinction coefficient, κ . When ITO is packaged as one electrode of an electrical capacitor, applying the voltage puts the capacitor into the three known states of accumulation, depletion,

or inversion. For example, in accumulation, free carriers are accumulated in the interface of the ITO and the oxide, thus changing the carrier concentration. From the Drude model, explicit relations for both the index and extinction coefficient as functions of the carrier concentration can be obtained. The optical property of the active material therefore changes dramatically, resulting in strong optical modulation effects (Figure 1C and D). Derivatives of the real and imaginary parts of the complex index with respect to the carrier concentration are taken, from whom the respective derivative $\delta n/\delta \kappa$ can be calculated (Figure 1I). It is interesting to note that the ratios $\delta n/\delta N_c$ and $\delta \kappa/\delta N_c$ intersect each other close to the ENZ region at a carrier concentration of $6.2 \times 10^{20} \text{ cm}^{-3}$, which is close to that reported first in ref [28]. The relative change in $\delta \kappa/\delta N_c$ is more than the relative change in the $\delta n/\delta N_c$ after a certain point in the carrier concentration prior to ENZ, which suggests that the effect of n change with modulation is stronger to the left of this point and the effect of κ change is stronger to the right. This can be observed in the graph for the derivatives with respect to the carrier concentration for ITO (Figure 1J), which suggests that the material is n -dominant to the left of this point, and κ -dominant to the right. Thus, ITO can be used for the EO case in the n -dominant regime or for the EA cases in the κ -dominant regime depending on the carrier (bias) condition.

The graphene dispersion relation can be obtained via the Kubo formula (supplementary information, Eq. (2), Figure 1E and F). The value of n and κ for graphene are calculated with varying wavelength and chemical potential, μ_c . We note that the chemical potential for graphene used in the Kubo formula moves up and down with the carrier concentration, and hence it is used here as the voltage-dependent index tuning parameter for modulation. Keeping the operating wavelength at 1550 nm, we calculate the complex index dispersion and subsequent derivatives (Figure 1K and L). As mentioned above, the photons with energies larger than $2\mu_c$ experience strong absorption, while the photons with smaller energies do not; therefore, when the chemical potential moves across $\mu_c = \hbar\omega/2 \approx 0.4 \text{ eV}$ the extinction coefficient undergoes rapid downward change as evidenced by the sharp negative peak of $\delta \kappa/\delta \mu_c \sim -\delta(\mu_c - 0.4)$, where $\delta(x)$ is the delta function, in Figure 1K. According to Kramers-Kronig transform, the real part of the refractive index also experiences rapid change in the vicinity of $\mu_c \approx 0.4 \text{ eV}$ as $\delta n/\delta \mu_c \sim -1/(\mu_c - 0.4)$. In practice, large change in extinction coefficient occurs in the range of μ_c values near 0.35–0.45 eV as the Pauli blocking effect can be smeared for $\sim 0.1 \text{ eV}$ at room temperature (Figure 1L, inset, shaded region). In our calculations, we take an approximation

for the Kubo formula from Hanson where the interband term is approximated at cryogenic temperature, i.e. 0 K [40]. The modulation technique we investigate here involves interband transitions, whereas intraband free carrier absorption can rise even after the Pauli blocking range of $\sim 0.1 \text{ eV}$ and can be approximated by the Drude model there. Depending on these interband transitions of Pauli blocking and temperature-dependent smearing of the transition, n dominance is present for μ_c ranges past the transition until the free carrier absorptions start around 0.5 eV due to very low κ values there and abruptly changing n values. κ dominance is on either side of the n -dominant region. For the subsequent modal and cavity discussions below we only consider the κ dominance on the left side of the n -dominant region, because either κ -dominant region is suitable for EA purposes and the n -dominant region is of interest in EO modulation. We note that there are certain differences in the characterization of graphene using the Kubo model, different approximations and a discussion on different results are included in the supplementary information.

2.2 Modal tuning properties

Having established these index-character regimes as a function of voltage tuning, we are next interested in the possible modal impact of such modulation. We study three different mode structures for each of the three active materials introduced above (Figure 2), while our aim is to explore modulator-suitable material/mode combinations for both EA and EO modulation mechanisms. The target is to increase the LMIs towards ultra-compact modulators while preserving ER, i.e. modulation depth, and we consider plasmonics as a spatial mode compression tool towards increasing the LMI and compare two distinct plasmonic modes with a bulk-case for comparison. The two plasmonic modes analyzed are the slot waveguide in a metal-insulator-metal (MIM) configuration [21–23, 41–43] and a hybrid plasmonic polariton (HPP) design in a metal-insulator-semiconductor (MIS) configuration. In order to understand the LMI enhancement effect from modal compression, we compare each active material with a bulk case where the entire waveguide consists of the active material only. The resulting design space is a 3×3 matrix, where we track the modal field distribution (Figure 2) as well as effective index for each waveguide case (Figure 4).

All mode structures are chosen on top of a SiO_2 substrate, thus providing a leveled playing field. The bulk Si waveguide is the classical silicon on insulator (SOI) rib waveguide. Starting our discussion with the active

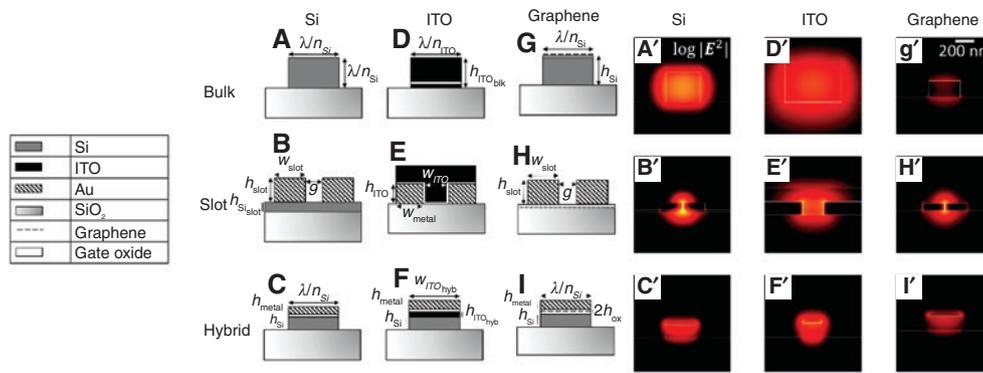


Figure 2: Schematic of the mode structures and FEM simulated mode profiles for all the structures at their respective starting point from the material dispersion at $\lambda = 1550$ nm.

The relevant parameters are $\lambda/n_{\text{Si}} = 451$ nm, $\lambda/n_{\text{ITO}} = 800$ nm, $h_{\text{Si}} = h_{\text{ITO}} = 200$ nm, $h_{\text{ITO}_{\text{bik}}} = 600$ nm, $h_{\text{slot}} = 100$ nm, $w_{\text{slot}} = 300$ nm, $h_{\text{Si}_{\text{slot}}} = 30$ nm, $g = 20$ nm, $w_{\text{metal}} = 550$ nm, $w_{\text{ITO}} = 300$ nm, $h_{\text{metal}} = 20$ nm, $h_{\text{ITO}_{\text{hyb}}} = 10$ nm, and $w_{\text{ITO}_{\text{hyb}}} = 250$ nm. The simulated results are shown in log scale due to their largely varying electric field strengths. All gate oxides in this work have thickness $h_{\text{ox}} = 5$ nm to ensure similar electrostatics.

material Si case, the height and width of the waveguide were chosen to coincide with the diffraction-limited dimension λ/n towards maximizing packing density. For all three Si cases we assume that it is possible for the entire Si portion to change index with applied bias. This is possible as the method of modulation here is bulk carrier injection as previously demonstrated [24, 25, 32, 44, 45]. It is noticeable that this mode is not strongly confined to the physical cross-sectional area of the waveguide mode due to modal field leakage (Figure 2A'). A similar mode profile can be seen from Figure 2D' for the case of bulk ITO structure, which is a waveguide made of ITO on top of a SiO_2 substrate (Figure 2D). Index modulation for bulk ITO can in principle be done via thermo-refractive effects, but the effect is rather slow (\sim ms); here the carrier concentration in ITO increases with annealing temperature [46, 47]. A second mechanism for index tuning of ITO is a capacitive carrier modulation-based mechanism discussed above [3, 10, 14, 15, 21, 26, 27] Thus, for biasing purposes, a 5 nm gating oxide layer is included (for every mode considered for electrostatic consistency). Here the accumulated carriers shift the plasma dispersion via the Drude model. In praxis, a $1/e$ decay length of about 5 nm was measured before [21], and high index modulation has been experimentally verified over $1/e^2$ (10 nm) thick films [27, 28]. As such, the modulation mechanism for ITO considered is based on capacitive-gated carrier accumulation rather than injection as for the Si case. However, similar to the Si bulk case, the bulk ITO waveguide is also fairly leaky and unconfined due to the relatively small index contrast with the surrounding air (Figure 2D'). Only the thin accumulation layer of about 10 nm changes optical properties with tuning, while the remainder of the ITO

does not change with applied bias. The bulk graphene structure consists of placing a single layer graphene on a gate oxide on top of a Si waveguide, thus forming an electrical capacitor. The Si waveguide was chosen to have a height of 200 nm, which supports the second-order transverse magnetic (TM) mode resulting in an improved modal overlap (i.e. in-plane electric field) with the active graphene sheet [12]. The silicon width is taken as diffraction-limited λ/n similar to the bulk Si case. It is important to point out that all the waveguides requiring a gate oxide in this work have been designed with a fixed oxide thickness such that we can compare them in a similar standard. That is, to ensure the same electrostatic potential during the comparison, all gate oxides in this work are 5 nm in thickness. The Si slot is inspired by the dielectrically loaded surface plasmon polariton structure; a 30 nm Si layer on top of the substrate is chosen to enhance the plasmonic modal interaction with the Si material [23]. This narrow thickness allows the necessary index contrast between the Si and SiO_2 substrate to squeeze the light in the plasmonic gap. Two metal pads are placed to facilitate the plasmonic mode in the gap of 20 nm. From simulation results, the gap mode is chosen to ensure high ER. We note that the mode is rather lossy, and the field intensity in the slot is about 232 times more compared to the bulk case despite the structure being a Si slot. This is due to the light being higher confined in the gap inside the slot between the metal walls compared to that of the active Si. The ITO slot follows the PlasMOSstor design [22]. The gap between the metal pads is chosen as a 5 nm Al_2O_3 layer for gating and subsequently topped with ITO. The slot gap is taken to be 300 nm to preserve the photonic mode within the gap when ITO is not in the lossy state (modulator on

state). Due to the fact that carrier accumulation can only be around 10 nm, the width of the slot beyond 20 nm will have little impact on confinement since only activated ITO contributes to the confinement factor with tuning. Our results show a highly confined structure (Figure 2E'). Also, the considerably larger dimension reduces the fabrication complexity. The graphene slot consists of placing a single layer of graphene on top of the SiO₂ substrate separated by a gating oxide of 5 nm. Then two metal pads form the slot structure. Here the gap is also 20 nm similar to the Si slot, which we have found previously to deliver high modulation performance [14]. We note that broader gap dimensions lead to higher order modes, lower optical confinement, and hence lower ER. This value (20 nm) can be understood from two aspects both relating to the fact that metallic confinement beyond 20 nm is not favorable: (a) the skin depth of plasmons at telecom wavelengths is about 20–30 nm, and (b) the Purcell factor reduces dramatically beyond 10 nm small plasmonic cavities due to high losses and field leakage [48]. Our results indeed confirm a modal confinement to the gap and a high field strength (peak $|\mathbf{E}|^2$), which is 4600 times higher compared to the bulk case (Figure 2H'). Finally, for confinement of light in the transverse direction and to obtain a dielectrically aided plasmonic mode for better modulation control, we choose the hybrid mode structures [15, 29, 49, 50]. The hybrid Si mode comprises a metal and an oxide layer on top of a Si waveguide that is 200 nm thick. The results show a confined mode with most of the field being inside the oxide. The hybrid ITO mode is similar to the hybrid Si mode, with an added 10 nm layer of ITO between the oxide and Si layers [28]. The hybrid ITO structure is essentially a metal oxide semiconductor (MOS) structure with a 10 nm ITO layer inserted between the oxide and Si interfaces. With the metal on top and Si on bottom, it creates an optical capacitor. The gap (oxide+ITO) provides the means to store electromagnetic energy, leading to sub-wavelength optical guiding [50]. The strong energy confinement in the gap arises from the continuity of the displacement field at the material interfaces, which leads to a strong normal electric-field component in the gap [51]. The dielectric discontinuity at the semiconductor-oxide interface produces a polarization charge that interacts with the plasma oscillations of the metal-oxide interface; that is, the gap region has an effective optical capacitance. The hybrid graphene mode comprises a metal layer on top of a 10 nm oxide layer, and the graphene single layer is sandwiched inside the oxide. These are stacked on top of a Si waveguide with 200 nm thickness [10]. Both the hybrid ITO and graphene structures show reasonably high confined modes and field enhancements of 46 and 63 times

compared to their corresponding bulk cases, respectively. It is worthy to mention that the slot and HPP modes are comparably lossy without accounting for the material loss to contribute as a byproduct of tuning. As such, one intuitively would expect these to be suitable for EAM devices. However, there are also regions where the EOMs (real part index tuning) via phase shifting outperform bulk cases despite the high losses, as discussed below. It should also be noted that all the metal used for the modal simulations is gold (Au), which has a reasonably low ohmic loss at near IR wavelengths.

Key to the modulator performance such as ER is the amount of index change that can be obtained upon biasing the device. EO modulation or phase modulation, i.e. the effective change of the k -vector of the light, $\delta k = \delta \omega_0 \frac{\partial k}{\partial \omega} = \frac{\delta n}{n} \omega_0 \frac{\partial k}{\partial \omega} = \frac{\delta n}{n} k_0 c \frac{\partial k}{\partial \omega} = \delta n k_0 \frac{n_g}{n}$, where n_g is the group index in the corresponding structure. The phase change then becomes $\Delta \phi = \frac{2\pi}{\lambda} \Delta n_{\text{eff}} L = \epsilon_r \Gamma \delta n k_0 \frac{n_g}{n} L$, where Γ is the optical confinement factor and $\epsilon_r = \frac{\epsilon_r}{\iint_s \epsilon_{\text{eff}} ds}$. The index change inside the modulator (Δn_{eff}) is then given by $\Delta n_{\text{eff}} = \epsilon_r \Gamma \delta n \frac{n_g}{n}$. For discrete modulation states this relationship can be expressed as the ratio of the altered active material index relative to its initial condition ($\Delta n_{\text{mat}}/n_{\text{mat}}$) multiplied by its modal confinement factor (Γ), relative modal permittivity enhancement (ϵ_r) and effective group index (n_g), i.e. [31],

$$ER \propto \Delta n_{\text{eff}} = \epsilon_r \Gamma \frac{\Delta n_{\text{mat}}}{n_{\text{mat}}} n_g \quad (1)$$

We note that the group index n_g is inside the cross-sectional structure corresponding to propagation in the longitudinal direction given by $n_g = n_{\text{eff}} - \lambda \frac{\partial n_{\text{eff}}}{\partial \lambda}$. Equation (1) applies for isotropic index materials, i.e. silicon and ITO-based structures in this work. Due to the unique EO nature of graphene and anisotropy of the indices (tensor), this simple equation insufficiently describes modulation performances in the graphene-based structures. The graphene propagating energy index and group index need to be represented by directional tensor terms and solved for each component, which is, however, beyond the scope of this work and should be explored by researchers in the field. Here, we follow a similar approach for the graphene-based modes to the bulk cases in order to associate modulation effects relating to the modal illumination pattern and effective index change. The material index ratio change from Eq. (1) was previously discussed

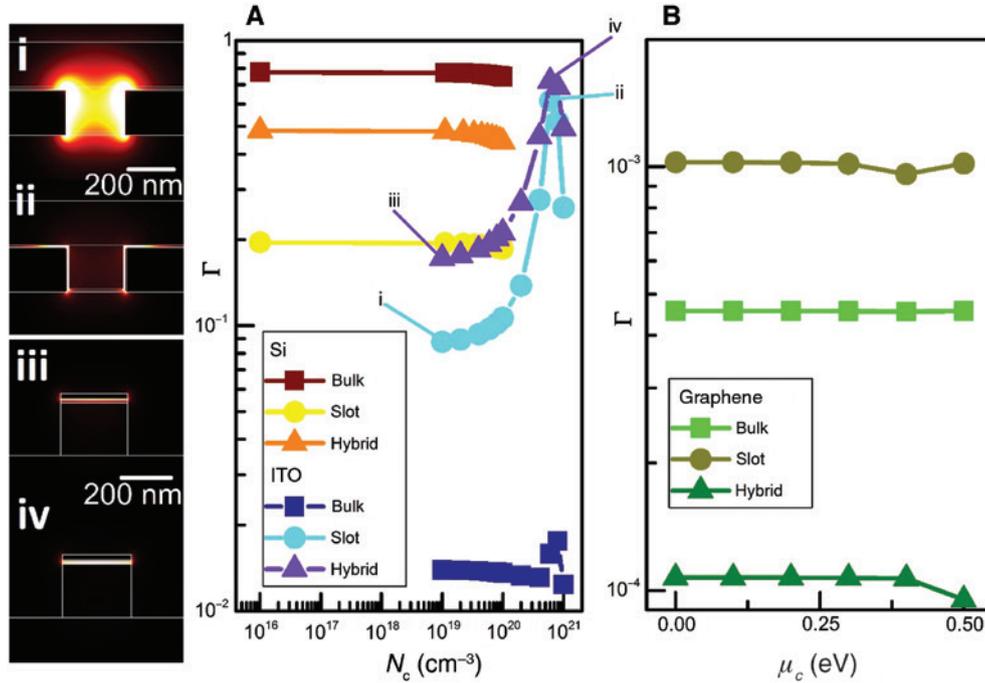


Figure 3: Confinement factors of the modes from Figure 2 with tuning.

Confinement factors corresponding to the Si and ITO modes vs. carrier concentration (A) and confinement factors corresponding to the graphene modes vs. chemical potential (B). (i, ii) ITO slot at 10^{19} cm $^{-3}$ and 6×10^{20} cm $^{-3}$, (iii, iv) ITO hybrid at 10^{19} cm $^{-3}$ and 6×10^{20} cm $^{-3}$, respectively.

in Figure 1. Next, we focus on the confinement factor first, then discuss obtainable effective index changes that govern the modulator operation. Our modulator bias analysis is based on selecting discrete bias points for all nine waveguides to obtain their corresponding effective indices (Figures 3 and 4, and see supplementary information).

2.2.1 Confinement factor and index tuning

For waveguide modes containing diffraction-limited mode (bulk) the modal confinement factor is given by the spatial field (\mathbf{E}) ratio of the modal overlap of the active region, i.e. the material whose index is electrically being altered, relative to the size of the entire mode:

$$\Gamma_{\text{Bulk}} = \frac{\iint_{\text{Bulk}} |\mathbf{E}|^2 dS}{\iint_S |\mathbf{E}|^2 dS} \quad (2)$$

A more comprehensive approach to define the modal confinement factor would be to define it with respect to the Poynting vector ($|\mathbf{E} \times \mathbf{H}|$) corresponding to the power flow in the entire structure. For a more straightforward and comparable approach, here we define the modal confinement in terms of the overall electric field intensity ($|\mathbf{E}|^2$)

[50, 52]. Similarly, two-dimensional (2D) materials show their unique EO tunability when the electric field is in the lattice plane (i.e. in-plane), but not perpendicular to it (out-of-plane), due to the low polarizability in this direction [14]. Hence, the confinement factor to characterize the light-2D material interaction becomes

$$\Gamma_{2\text{-D}} = \frac{\iint_{2\text{-D}} |\mathbf{E}_{\text{in}}|^2 dS}{\iint_S |\mathbf{E}|^2 dS} \quad (3)$$

where \mathbf{E}_{in} denotes the in-plane electric field, and \mathbf{E} is the overall electric field. The latter is a critical requirement for 2D materials, since their polarizability vanishes for out-of-plane fields. Thus, simply placing a 2D material on a plasmonic waveguide surface will not lead to increased overlap factors despite the high field concentration near that metal-dielectric interface. In fact, this is a fundamental challenge of integrating 2D materials with plasmonics. A possible waveguide design, however, to overcome this bottleneck is a MIM-like slot waveguide with a graphene layer parallel to the slot field (Figure 2H). Starting with the silicon as the active material, the bulk mode has an overlap factor approaching unity (80%) for the modulator on state at low carrier concentrations (Figure 3A). However, upon plasma dispersive biasing, the overlap

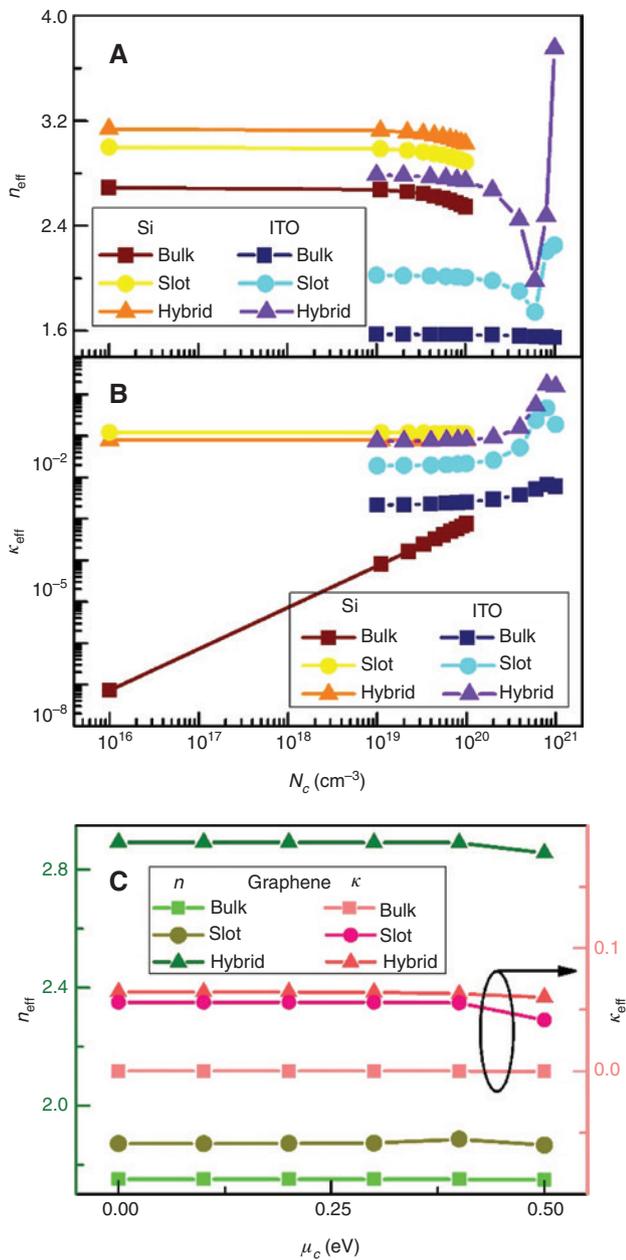


Figure 4: Simulated results for effective indices for all the modes in Figure 2 for different applied bias.

(A) Real part of the effective refractive index for Si and ITO vs. carrier concentration, (B) imaginary part of the effective index, i.e. effective extinction coefficient for Si and ITO vs. carrier concentration, and (C) Real and imaginary parts of the effective index for graphene vs. chemical potential.

factor slightly decreases since the mode becomes more lossy. The two plasmonic modes, hybrid and slot (Figure 2B and C), worsen the overlap since the silicon can only sit near the highest fields of both modes but not directly inside the slot. The latter could be different, for instance, if the slot would be filled with silicon, but this is not part of

this study. The Si slot has the lowest confinement of about 20% since the field is mainly concentrated inside the slot gap. All three silicon-based waveguides have a relatively flat confinement factor with tuning, since the modal distribution does not experience any significant change.

Both the confinement factor and material index change impact the effective index with applied bias (Figure 4A). The applied bias changes the carrier concentration by sweeping in more free carriers in Si and ITO, thus changing the optical response, and by increasing the chemical potential level in graphene. The relationship between the carrier concentration and chemical potential is about a factor of 10, i.e. 0.1 eV \sim 1 V, but depends on electrostatics (supplementary information, Section 2). The change in effective n is not dramatic, but κ_{eff} varies rather rapidly with the carrier concentration for the bulk Si case. However, even with the increase in κ_{eff} , it is noticeable that the mode is relatively lossless compared to the other modes studied as Si material inherently exhibits little loss. The change in the effective index and effective extinction coefficient follows the material model from previous sections, since this bulk mode is material homogeneous. This is further evident from the confinement factor of the mode; while we find a high modal confinement factor (i.e. high overlap), its value does not change significantly with increasing carrier concentration (Figure 3A). The slot and hybrid structures for Si exhibit a higher amount of loss than the bulk Si. The changes in the real part of the effective indices with modulation, in these cases, are stronger than the bulk, as expected from higher confined structures. Although the effective extinction coefficients for the slot and hybrid Si are relatively flat (Figure 4B), they do in fact decrease with carrier concentration near 10^{19} – 10^{20} cm^{-3} . However, this does not violate the Kramers-Kronig relations because the more tuning is in effect, the more light is projected out from the active Si, which becomes increasingly lossy with tuning. This effect can be seen from Figure 3; the slot and hybrid Si confinement factors decrease towards the right end. This means that the light leaks more into the oxide and the gap for the hybrid and slot cases, respectively.

Unlike the silicon case, the three ITO-based waveguides experience a rather strong index change with bias. This is mainly driven by the drastic change in the confinement factor. Starting with the slot waveguide, the field is fairly loosely confined within the 10 nm thin ITO, which is in close proximity from the metal slot arms, only separated by the 5 nm gate oxide (i, Figure 3A). This is due to the index change in the activated ITO, which forms sufficient contrast with both the bulk ITO and metal on either side, effectively creating a MOS capacitor facilitating

electromagnetic charge storage. This effect occurs for up to near the ENZ point in carrier concentration leading to highest modal overlap with the ITO (ii, Figure 3A). Beyond the ENZ point, the mode starts to become more unbounded, and confinement decreases as the effective index increases and the field tends to be squeezed into the lower index oxide again. In (ii), it is noticeable that almost all the light in the structure is compressed inside the activated ITO accumulation layer of about 10 nm thickness. A similar trend is observed for the hybrid ITO case. However, the field confinement of the simple hybrid waveguide (iii, Figure 3A) offers a higher overlap for the on state. This, however, is necessarily desired since the ITO has usually a higher loss than the silicon (typical SOI) part of the hybrid design. Hence, lower overlap factor for the modulator on state would be beneficial to reduce the insertion loss.

For the hybrid ITO case, (iii) is the starting point in the range of carrier concentration for ITO. The most confined state is shown in (iv), which has a carrier concentration level near ENZ. Most light is confined in ITO in this state; thus, the LMI in ITO is at its maximum level here. By means of the effective optical capacitance in the gap (oxide + ITO), arising from the continuity of the displacement field at the material interfaces, the light is strongly confined in the gap region in the hybrid structure [53]. At the start of the carrier concentration range, ITO index is higher than that of the oxide and light is prone to occupy spaces with lower index in such modes – it means that the light tends to be squeezed predominantly into the oxide layer (Figure 3A, iii), resulting in a lower confinement factor. With tuning, as the carrier concentration increases, the material index of ITO drops to values lower than that of the oxide. This causes a change in the illumination profile of the mode, and more light is therefore confined in the ITO layer (Figure 3A, iv). As a result, the confinement factor increases significantly near ENZ, which is a positive effect for the modulator. For carrier concentrations beyond ENZ, similar to the slot ITO case, more light is confined into the oxide layer rather than the ITO since the oxide layer again has a lower index, and thus the confinement decreases. Hence, the slot ITO structure can be thought of as two hybrid modes rotated vertically and then mirrored sideways. The mode structure sideways consists of material interfaces of metal, oxide, activated ITO and bulk ITO, respectively; and the same combination mirrored. The activated ITO, i.e. accumulation layer of about 10 nm from the oxide interface is, in effect, similar to the ITO layer in the hybrid case. The rest of the ITO material remains in its bulk state and can be compared to the Si in the hybrid case. In this structure, similarly the oxide and activated ITO confines more light at the starting carrier concentration. But

the un-modulated ITO confines considerable amount of light as well because of the mirror symmetry in this mode (Figure 3A, i). The asymmetry in the hybrid mode causes the light to be squeezed into the gap creating an optical capacitance able to store the electromagnetic energy. But in the slot structure, the symmetry allows the light to be more spread out in the bulk (un-modulated) ITO. This effect also is increased due to the fact that the bulk (un-modulated) ITO index is considerably lower than that of bulk Si. The bulk (un-modulated) ITO index being closer to the oxide index also decreases the forming of the optical capacitance effect. With increasing carrier concentration, i.e. tuning, as the activated ITO index becomes comparable to that of the oxide and lower, light is more confined into the activated ITO giving rise to the confinement factor (Figure 3A, ii). This effect also can be observed increasingly more near ENZ in the carrier concentration. Similar to the hybrid case, the confinement drops down for carrier concentrations beyond the ENZ region.

For bulk ITO, the effective n change is very low as tuning only changes the accumulation layer of only about 10 nm (Figure 4A). This is because the carrier accumulation is only about 10 nm into the bulk at the ITO-oxide interface; the rest of the ITO remains in the bulk state. The minimal confinement for the bulk ITO mode also reflects this effect. The confinement increases marginally towards the end of the carrier concentration range, which accounts for the small increase in the effective extinction coefficient in that range as the material becomes lossy. The ITO slot provides better results in terms of both the effective index change and effective extinction coefficient change. Figure 3A shows a higher confinement for this structure as well. The highest confined structure, in the ITO case, is the hybrid mode (Figure 3A), which reflects a considerable change with tuning in both real and imaginary parts of the effective index. Both the slot and hybrid structures for ITO are quite lossy to begin with, as both of them are plasmonic, and with tuning they become even more and more lossy as the material becomes κ -dominant. Traditionally with tuning, the effective index drops, and the effective extinction coefficient rises due to the Kramers-Kronig relations. But with the ITO slot and hybrid cases we see the opposite in the limit of higher carrier concentration corresponding to near and beyond ENZ. A closer look into these reveals that the mode, in a way, “flips” near the ENZ region. Towards the left of the ENZ region, the mode is more confined in the ITO and follows the material traits; i.e. the mode becomes lossy with the material becoming lossy. Near and beyond the ENZ region the mode is leaking toward the oxide layer for hybrid case and toward the un-modulated bulk ITO for slot case as ITO becomes metallic

across ENZ region; thus, towards the end of the carrier concentration range the confinement factor decreases.

The graphene structures experience lowest change in effective indices with tuning, for both the real and imaginary parts, which is expected as graphene used here is a single-layer 2D material with a naturally low light confinement into the graphene layer, due to its miniscule 0.35 nm thickness (Figure 3B). This essentially separates graphene from the rest of the bulk materials in this study; the unique properties of graphene and the drastic change in refractive index between μ_c values in the Pauli blocking range with an infinitesimally small extinction coefficient make it a formidable opponent as it enables operations in the nanoscale. The real and imaginary parts of the effective index of the bulk graphene mode remain fairly the same and follow some minor changes with tuning. The effective n of the slot graphene structure increases a small amount and then decreases with tuning, following the material trait. The effective κ stays almost constant and then decreases in the n -dominant region of the material due to the fact that κ of the material itself here is infinitesimally small, near zero. The confinement factor also follows the same trend; i.e. confinement increases as graphene becomes lossless. The hybrid graphene case is contrary to the slot, where the effective index takes a dip towards the end of the chemical potential range. The effective extinction coefficient also takes a minor dip in this region. This can be explained by the confinement factor, which is less as the chemical potential reaches the n -dominant region of the material (Figure 3B). The drastic change in the index amounts to only small changes in the confinement factor because of the single layer of the material and the fact that only in-plane components of the associated electric field correspond to it.

2.3 Assessments based on material and modal analysis

The higher losses of deeply confined waveguide modes such as found in plasmonic waveguide design must be accompanied by strongly index-changing active materials for high modulator ERs. That is, combining a low-index altering material such as silicon with plasmonic modes does not improve ER performance. This is because the parasitic loss incurred upon plasma dispersion tuning of silicon is yet orders of magnitude lower compared to the modal loss of plasmonic waveguides. The high modal overlap for silicon bulk waveguides results in almost similar index changes compared to plasmonic-silicon-based modes with lower overlap factors than bulk. For

example, the modal overlap differs by a factor of 4 for the silicon bulk relative to the plasmon slot, and it is compensated by the field enhancement more than 200 times in the plasmonic mode compared to that of the bulk mode. The strong index modulation of active TCO materials in combination with increased modal overlap upon plasmonic field squeezing offers high modulation index differences beneficial for EO modulation. However, the electrical biasing scheme and subsequent index changing region of TCOs make them weak candidates due to low overlap factors when diffraction-limited modes are used. For 2D materials such as graphene, the combination of such a material with diffraction-limited modes (bulk case) does not allow for strong modulation despite high material index potential due to the miniscule field overlap. Furthermore, plasmonic modes need to be designed with care since their TM character (field normal to the metal surface) hinders in-plane field components inside the 2D material. If this is not observed, the only modulation performance enhancement relevant to the modulator is the plasmonic field enhancement upon mode squeezing that offers any in-plane component such as roughness or edges. On the other hand, slot waveguides do provide the correct field polarization (in-plane) for 2D materials and allow for overlap factors with 2D materials approaching 1% [14]. This is possible when the plasmonic slot gap and metal height is each decreased in dimensions to tens of nanometers.

2.4 Tuning properties in cavities

Next, we are interested in exploring to enhance the electrical field further inside an EOM by deploying an optical cavity while trading in spectral bandwidth. We consider the 3×3 material-mode combinations inside a FP cavity model to study the resulting modulation performance. The effect of tuning is cavity dependent as the change in resonance can lead to corresponding losses in turn, relating back to the Kramers-Kronig relationships from the material model. This spectral shift-to-loss relation is cavity dependent as we can relate the shift in resonance ($\Delta\lambda$) to the dominant longitudinal mode while the incurred loss ($\Delta\alpha$) depends on the Q factor of the cavity. Hence, we are interested in an optimum FOM for cavity-based EOMs that is proportional to $\Delta\lambda/\Delta\alpha$.

2.4.1 The Fabry-Pérot cavity

For the FP cavity an established formula for analytical manipulations exists to aid our approach [48].

$$Q = -\frac{2\pi}{\lambda} \frac{2n_{\text{eff}}L}{\log[R_1R_2(1-T_{\text{loss}})^2]} \quad (4)$$

where Q denotes the Q factor of the cavity, R_1 and R_2 are the optical reflectivity of the cavity facets, T_{loss} is the fractional internal loss per pass and L is the length of the cavity. Silver (Ag) mirrors are used to form the FP cavity because the air reflection in ITO is rather low allowing only for weak feedback, lacking the necessary finesse. Hence, assuming near-normal reflection at the mirrors, the corresponding mirror reflectivity becomes

$$R_1 = R_2 = \left(\frac{\tilde{n}_{\text{Ag}} - \tilde{n}_{\text{eff}}}{\tilde{n}_{\text{Ag}} + \tilde{n}_{\text{eff}}} \right)^2 \quad (5)$$

The total loss in the cavity has two major components – material loss and metallic loss. Material loss is due to the active material itself being lossy with considerable effective extinction coefficients. The use of metals for mirrors in the cavity does allow improved reflectivity, but the high extinction coefficient of metals incurs some inevitable loss in the system, especially in the small length limit. This metallic loss is due to the fact that trapped light inside the cavity can penetrate the metal mirrors based on skin depths at respective frequencies resulting in field penetration loss in the mirrors.

$$T_{\text{loss}} = \alpha_{\text{abs_total}} + \alpha_{\text{pen}} \quad (6)$$

$$\alpha_{\text{abs_total}} = 1 - e^{-\alpha_{\text{abs}}L} \quad (7)$$

$$\alpha_{\text{abs}} = \frac{4\pi\kappa_D}{\lambda} \quad (8)$$

$$\alpha_{\text{pen}} = \left(\frac{4\pi\kappa_m}{\lambda} \right) \delta_s \cdot \frac{2\delta_s}{2\delta_s + L} \quad (9)$$

where α_{abs} is the absorption coefficient of the active material, κ_D is the effective extinction coefficient of the active material and λ is the operating wavelength. $\alpha_{\text{abs_total}}$ denotes the material loss, i.e. light absorption in the cavity, and α_{pen} denotes the metallic loss due to field penetration into the mirrors. κ_D and κ_m are the extinction coefficients of dielectric and metal materials, respectively, δ_s is the penetration or skin depth of the metal at a corresponding frequency of the trapped light at which the field magnitude drops to $1/e$ of the surface value and is given by $\delta_s = \sqrt{\lambda\epsilon_0 c / \pi\sigma}$, σ is the electrical conductivity in $\Omega^{-1}\text{m}^{-1}$, and ϵ_0 is the vacuum permittivity [48]. κ_m is high for metals, i.e. Ag in this case. The dominant mode in the cavity is found from the finite element method (FEM) simulations. The corresponding

longitudinal mode order in the cavity is then found by $m = 2n_{\text{eff}}L/\lambda$ [54]. In theory, the change in the resonance can be found by setting the longitudinal mode order from the off to on states; i.e. the shift in the wavelength for different states reflects the tunability of the cavity.

$$\Delta\lambda = \frac{2L}{m} |n_{\text{eff,ON}} - n_{\text{eff,OFF}}| \quad (10)$$

3 Application space – electro-optic modulation

3.1 Cavity impact on electro-optic modulation

Utilizing resonances allows for a variety of advantages for modulators. Using a physical picture, the performance of an EOM is intimately tied to the ability to deliver strong LMI to modify an optical signal. The deployment of an optical cavity can thus enhance the LMI both spatially and temporally; spatially, a cavity concentrates the optical field by “folding-back” photons interacting with the modulated optical index and stores photons in time inside the cavity. The effect of a cavity for a modulator can be compared to a linear device (e.g. MZ, EAM), and one can reach the following qualitative results: the cavity (i) allows for more compact device designs, (ii) increases modulation sensitivity, (iii) however, limits modulation speeds and spectral bandwidth, and (iv) allows for a reduction in power consumption [48]. Next, we discuss some of these qualitatively. The footprint of a cavity-based EOM can be lowered by a factor about equal to the resonators finesse compared to the MZ [55]. In brief, the lengths of the MZ and ring given by $L_{\text{equiv-MZ}} = Q \cdot \lambda / 2n_g$, and $L_{\text{ring}} = 2\pi R$, where R and Q are the ring’s radius and quality factor, respectively. Thus, the ratio of the lengths of the two EOMs is given by the equation in Figure 5, where \mathcal{F} is the finesse of the resonator. Secondly, the sensitivity of the MZ is actually equal to that of the ring EOM, which is somewhat surprising, however can be seen by the following arguments: $\Delta\nu \approx \Delta\nu_{\text{FWHM}} = \frac{1}{2\pi\tau_{ph}}$ with $\frac{\Delta\nu}{\nu} \approx \frac{\Delta n}{n_g}$ requires $\Delta\nu = \frac{n_g}{Q}$, where $\Delta\nu_{\text{FWHM}}$ is the resonator linewidth, n_g the group index, Δn the modulated index (i.e. a function of applied voltage), and τ_{ph} the photon lifetime in the resonator. Since a single arm MZ requires an index change of $\Delta n = \lambda/2L$, extending the MZ length to be $L = Q \cdot \lambda/2n_g$ the same sensitivity to the ring design. Thirdly, the question is whether a long linear device (i.e. larger capacitance)

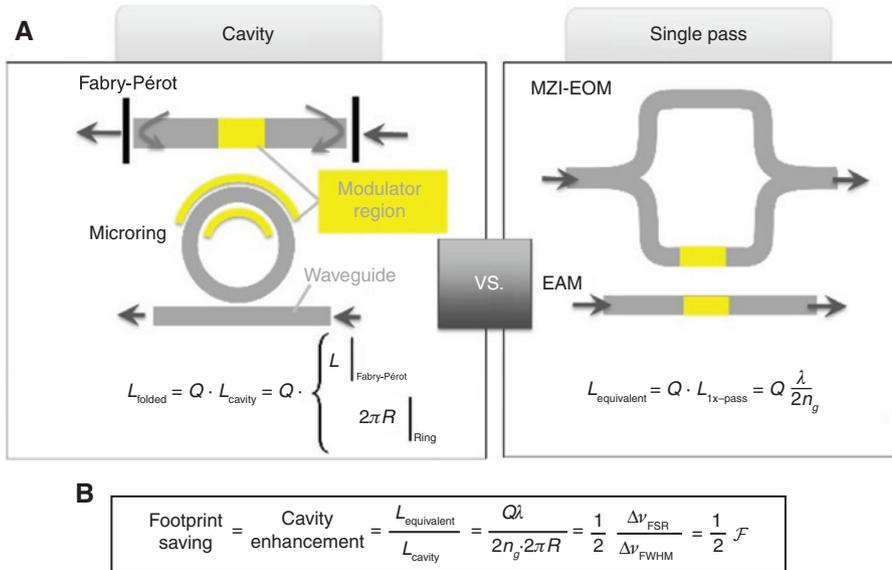


Figure 5: Impact of a cavity on electro-optic modulation.

(A) Schematics and effective photon interaction lengths for cavity-based EOMs vs. single-pass devices. (B) Footprint benefit from cavity is proportional to the finesse, \mathcal{F} . Definitions: L = physical lengths, Q = quality factor, R = ring radius, n_g = group index, λ = wavelength.

or more compact due to a cavity (but potentially photon lifetime limited) design has a higher modulation speed. The answer is that they are in fact equal, provided that the MZ is not parasitic capacitance limited as shown next. The MZ transit time (treating the MZ as a lumped element) is given by $T_d = \frac{n_g}{c}L$, and hence the transit-time limited bandwidth is $(f_{3\text{dB}})_{\text{MZ}} = 0.44 \cdot \frac{1}{T_{\text{trans}}} \approx \frac{1}{2T_{\text{trans}}} = \frac{1}{2n_g L} = \frac{v}{Q}$. The cavity's photon-limited bandwidth due to the long photon lifetime can be estimated via $(f_{3\text{dB}})_{\text{ring}} \approx \frac{1}{2\pi\tau_{\text{ph}}} = \frac{v}{Q}$, closing the above argument that the MZ and ring/FP cavity EOM have about the same modulation speed. For example, a modulator with a Q of 10,000 has a cutoff speed of about 34 GHz (for an EO coefficient of 300 pmV⁻¹, and ER=3 dB) [56].

We note that any slow-light effect used to increase the light-matter interaction for EO modulation can in principle be treated similarly to a cavity due to the increased photon interaction time with the active EO altered material, and we refer the reader to the tutorial on this subject by Khurgin [34]. However, the main difference between slow-light effects and a cavity is the spectral limitation of the latter. This introduces limitations with respect to wavelength division multiplexing (WDM)-based applications. However, here we argue that the footprint overhead of the necessary multiplexer/demultiplexer on-chip needs to be photonic (i.e. diffraction-limited photonics) and is

therefore significantly larger (millimeters) compared with the micro- and nanoscale modulator devices considered in this work. Hence, any WDM attempt would render the compactness of nanophotonic-based EOMs useless. A PIC utilizing such micro- and nanoscale EOM is therefore better suited with multiple point-to-point single-wavelength busses in parallel, yielding a higher bit-flow density [57].

Given the above rationale that a cavity, or slow-light effect waveguide, provides extended photon interaction time feedback system, such a system should be mainly intended for EO applications rather than for EA. We note that cavities with extraordinary high Q factors have been demonstrated, even on-chip [58–60]. Since our considered active materials are somewhat lossy, and we are interested in scaled-down devices using micro- and nanoscale cavities, we therefore focus here on realistic (lossy) cavities only. If the corresponding losses are sufficiently large such that all the light injected into the cavity is absorbed along the propagation direction, there will be no formation of resonances. In other words, EA operations inside a cavity system do not offer any meaningful insights, since the absorptive nature of the cavity upon biasing pulls any Q factor of the cavity to 0. However, it is perceivable that there is a cavity-material combination over which some amount of Q could provide enhanced LMIs and should be studied by the community. Here we focus on the discussion of the EO aspects. The shift in resonance with the change in carrier concentration, i.e. tuning, $\Delta\lambda$ corresponds to

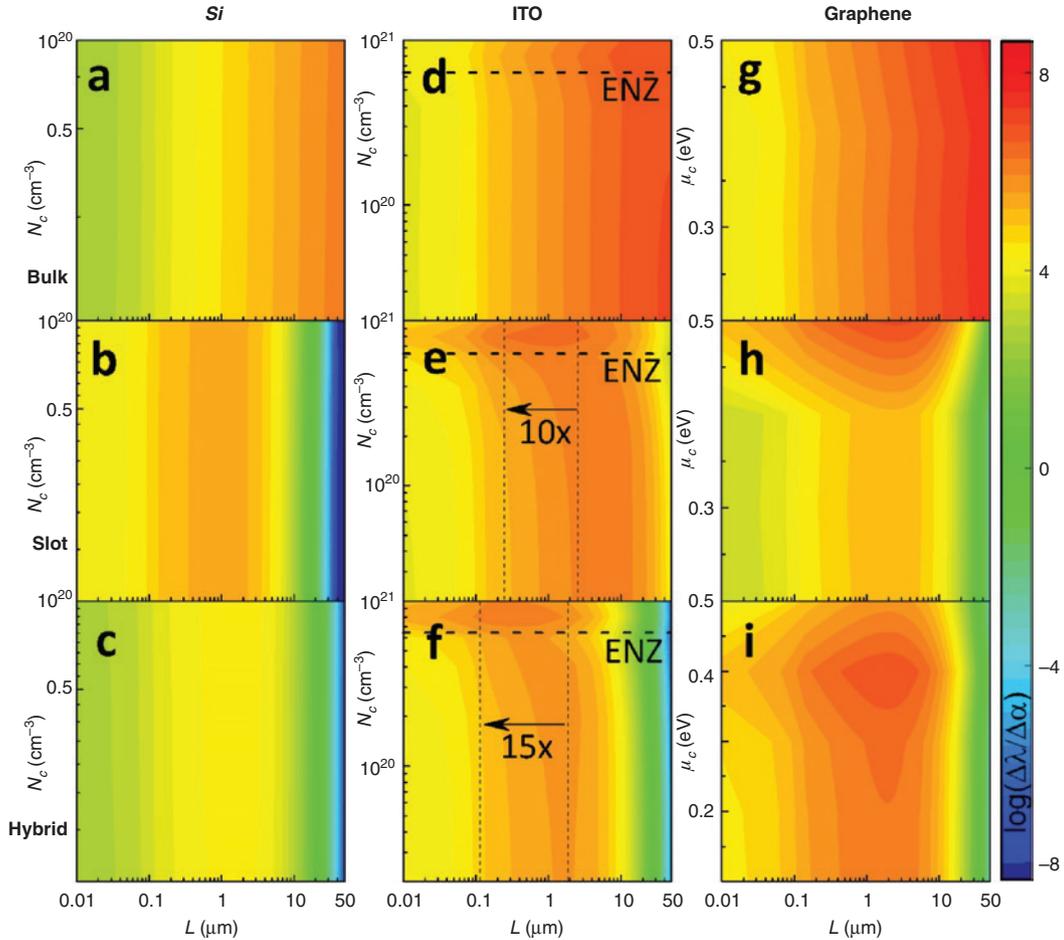


Figure 6: Fabry-Pérot-based cavity resonance shift-loss-ratio, i.e. $\log(\Delta\lambda/\Delta\alpha)$, vs. carrier concentration levels in cm^{-3} (for Si and ITO) and chemical potential in eV (for graphene) vs. scaling (μm) for all nine cavities based on the modes from Figure 2. Every point on the y axes on these graphs represent corresponding tuning from the start point in either carrier concentration (a–f) or chemical potential (g–i) to that point. Each color level in the color bar corresponds to a change of approximately $2.27 \times 10^4 \text{ cm}^2$ in the actual $\Delta\lambda/\Delta\alpha$ ratio.

the change in the effective refractive index (real part), Δn_{eff} from the modal tuning properties section above. The loss also increases with tuning as a direct result from the Kramers-Kronig relations, and our aim is to optimize the ratio of obtainable tuning which improves ER relative to these incurred losses, i.e. $\text{FOM}_{\text{EOM-cavity}} = \Delta\lambda/\Delta\alpha$. This change in the loss, $\Delta\alpha$ is a function of the modal effective extinction coefficient change, $\Delta\kappa_{\text{eff}}$. We note that with the introduction of a cavity other losses are also being added, but the fundamentally strong relationship is mainly corresponding to the material and subsequent effective modal extinction coefficients. For better EO performances, it is typically desired that the tuning aspect should outweigh the losses concerned. This essentially means that we want to maximize $\Delta\lambda$ without making $\Delta\alpha$ too large. It is apparent that we have to encounter the rising losses as a result of tuning, but we want to minimize this effect as much as possible for these cases. The $\Delta\lambda/\Delta\alpha$ ratio for all

the cavities are calculated and plotted with respect to the key tuning parameter changes, i.e. carrier concentration for Si and ITO structures and chemical potential for the graphene structures on the vertical axis and scaling of the cavities on the horizontal axis (Figure 6). To evaluate the results in the same scale, the cavity lengths are swept from 10 nm to 50 μm . The logarithm of the $\Delta\lambda/\Delta\alpha$ ratio is plotted for a better dynamic range. Here, a negative value of the ratio signifies that $\Delta\alpha$ is greater than $\Delta\lambda$ for those values. This essentially means that the losses overturn the tuning aspects for those levels, and the loss takes over in the cavity and EO modulation is limited. The material models suggest monotonically decreasing functions for the real parts of the index, n for Si and ITO. The graphene n is also decreasing in the n -dominant range. This necessarily means that the tuning of resonances can only blueshift the resonances at the start point in tuning. We also investigate the scaling aspects of the cavities utilizing

the cavity length dependence of Q ; typically with low loss, higher cavity length yields higher Q corresponding to lower energy per bit (lower tuning is required) and higher ER. This automatically puts the bulk structures ahead because of their low effective extinction coefficients. However, the loss term, $\Delta\alpha$ is scaling-dependent, whereas the shift in resonance $\Delta\lambda$ does not depend on the length of the cavity. The modal aspects of tuning can be observed in the vertical axes of all of the plots in Figure 6, and the scaling aspects, on the other hand, can be observed in the horizontal axes. In regards to the mode structures, it is noticeable that the bulk structures are the most favorable ones for tuning. If scaling is not a factor in consideration, the bulk structures can offer decent modulators but require large footprint. However, since the cavity length is proportional to the electrical capacitance and energy consumption (joules per bit), a shorter cavity size is desired. The bulk graphene case shows high performance. But if we upscale to longer cavity lengths, the Si bulk structure can outperform the ITO bulk one. This happens due to the inherent higher extinction coefficient of the ITO material; the $\Delta\lambda/\Delta\alpha$ ratio starts to decrease after a certain cavity length (near 200 μm) for the bulk ITO structure, whereas the ratio keeps increasing for the bulk Si one (supplementary Information, Section 4). Hence, if footprint is not a factor, a cavity with bulk Si can fetch higher tuning than the bulk ITO one at higher cavity lengths (Figure S4). The bulk Si mode exhibits a substantial change in the effective extinction coefficient (Figure 4B), but this feature is not noticeable in Figure 6 because of the very small value of the effective extinction coefficient (Figure 4B) along the entire tuning range. Q for bulk Si cavity design is sufficiently larger with tuning compared to the others (Figure S5). So the higher Q factor outweighs the imaginary part of the effective index change effects.

Scaling also plays a vital role in the tuning performances in cavities. All modes considered exhibit little to no voltage tuning for high scaling. This is due to the metal mirror feedback systems of the cavities. For shorter cavity lengths, the FP structure effectively becomes a MIM structure causing the Q factor of the cavity to plummet. Also, at longer scaling, all the mode structures with sufficient modal loss to begin with, i.e. the more confined ones (all except the bulk ones), exhibit a descent in the $\Delta\lambda/\Delta\alpha$ ratio. This, in turn, reflects the fact that the more confined the mode is, the shorter the device can be. For the more confined structures (slot and hybrid), around the micron sizes, prove to be optimal (Figure 6). Even with optimal scaling, it is noticeable that the ITO and graphene structures outperform the Si ones mainly because of the material traits following steeper index change. Near the

micron scale, the slot structure is better than the hybrid structure for EO operations and suggests high-performance micrometer small Mach-Zehnder interferometer-based EOMs. The Si slot exhibits about 11.5 times higher FOM than its hybrid counterpart, the ITO slot is about 1.5 times higher than its hybrid counterpart and the graphene slot is about 1.4 times higher than its hybrid counterpart, all for the highest possible tuning that corresponds to the highest values in the $\Delta\lambda/\Delta\alpha$ ratio near the micron level scaling. In the ITO plasmonic structures, i.e. the slot and hybrid, it is noticeable that similar performance devices can be obtained with a reduced footprint of about an order of magnitude by employing near ENZ modulation off state. The slot structure can fetch about one order of magnitude reduced scaling for similar performance in the optimal scaling region, and the HPP structure can fetch about 1.5 orders of magnitude reduced scaling similarly. If the length is fixed near the micron scale, we can compare the different material-based cavity performance with the same mode structures, for example, slot Si with slot graphene. A comparison for our intended purpose can be done for the bulk structures such that the graphene and ITO ones show nearly 43 and seven times higher FOM, respectively, than the Si one. Similarly for the slot structures, graphene and ITO ones are nearly 49 and 11 times higher in FOM, respectively, than their Si counterpart. Finally, for the hybrid cases, the graphene and ITO designs are 410 and 84 times higher than the Si case, respectively. Thus, overall, graphene appears to be a very formidable opponent to the bulk materials. Within the bulk materials, ITO outperforms Si in more confined and shorter structures. It should be noted that graphene-based EOM performances are overestimated in our work due to the interband transitions approximation we used in the Kubo formalism. Although Δn_{eff} for the graphene-based modes are lower than ITO cases, graphene modes experience almost an order of magnitude lower change in $\Delta\kappa_{\text{eff}}$ even with maximum tuning in the allowed range. Furthermore, the graphene-based modes exhibit lower κ_{eff} for higher tuning ranges, in turn leading to a considerable Q and subsequent high FOM results in associated cavities. Increase in temperature broadens (smears out) the transition. Results from Ooi et al. could be realistic from the various approximations to the Kubo formalism (Figure S1). Although an increased extinction coefficient results in higher insertion losses for a phase-sensitive modulator, the cavity Q -factor can compensate this added loss (Figure S5). Depending on the desired data-signal quality (i.e. extinction ratio of the modulator), the operation voltage could be reduced by a factor of about 3. Thus, if the entire system would be cooled to cryogenic levels

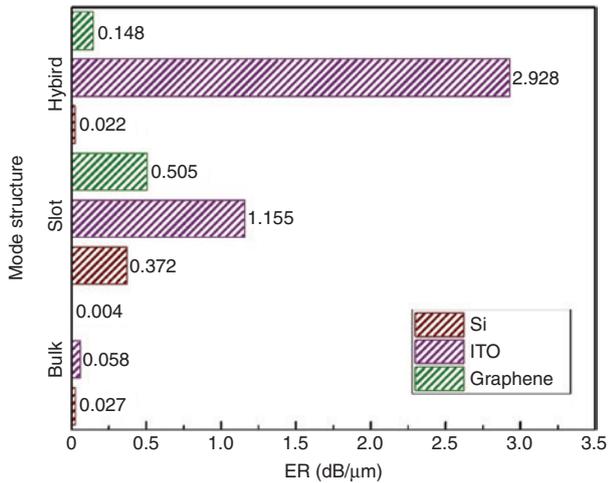


Figure 7: Electro-absorptive modulation (EAM) linear device performances characterized by extinction ratio (ER) for all the structures.

already, the energy per bit could be lowered by about a factor of $\sim 10 \times$ [61]. Indeed cooling devices to cryogenic temperatures could actually become an interesting option in single and few photon systems [62].

3.2 Electro-absorptive linear performance

The cavity-based EO phase modulator is a crucial component in the on-chip photonic network. Although cavity size can be adjusted, it is usually large compared to other on-chip devices, which limits the on-chip density when optimizing the link and network performance. Compared to phase modulation by using cavities, EA modulators, by tuning the loss term of active materials, provide a rather compact device footprint, giving ways to further increase the on-chip bit area density. Using the results from above, we numerically analyzed the EA property of aforementioned materials and optical modes. The performance is characterized by ER in dB (Figure 7). These results are analytically calculated from the effective mode index retrieved from FEM analyses. For Si modes, as expected, the ER for all three modes are low due to the fact that the imaginary part of refractive index changes insignificantly during carrier injection. Similarly, although graphene has a drastic refractive index change when tuning chemical potential, the EA performance for graphene is rather poor, mainly because graphene modes suffer from the low confinement within the single atom layer thickness. ITO has a modest index change compared to Si, and graphene, however, has the strongest EA response. For ITO bulk mode, as the confinement factor is quite low which is depicted in Figure 3, the extinction changes slightly

between on and off states. With the help of plasmonic modes, most portion of light could be squeezed into the active layer of ITO (Figure 3), which leverage the ER to a much higher level; especially in the hybrid mode the ER is as high as 2.9 dB/μm.

4 Conclusion

In this work we performed an EOM and EAM performance analysis for current-driven devices. Considering three active materials (Si, ITO and graphene), and three optical mode types (bulk, plasmonic slot and plasmonic-photon-hybrid modes), we investigated the effective index change upon biasing these modulator configurations while observing the device intricacies with respect to (a) material index change, (b) modal confinement factor and (c) group index. Furthermore, we consider a FP cavity model for these nine-mode-material cases to investigate cavity modulator FOM given by the resonance shift-to-loss-change ratio, motivated by the fundamental Kramers-Kronig relation, aiming to find a design optimization region. Out of the three materials chosen for this study, ITO and graphene clearly outperform conventional Si, in both EOM and EAM designs. The different modes discussed have their distinguishable advantages. The more confined slot and hybrid modes lead to shorter device lengths and are beneficial for achieving improved results for high scaling. We note that the bulk structure is the favorable option for overall higher EO modulation. If the scaling or footprint is not a deciding factor, one can easily employ the bulk structure to get enhanced modulation strength using cavity feedback. However, scaling and footprint reduction is of utmost importance to enabling nanoscale operation and improving the on-chip packaging density. The unique EO property of graphene with the sharp real part of the index change and the almost constant infinitesimal imaginary part of the index in that region can make it the highest performing material in terms of modulation strength. But consideration must be given to the corresponding confinement factor and to placing the graphene sheet in such a manner that interaction with the in-plane components of the involved electric field in the mode is maximized. In our choice of the mode structures, the graphene slot served these conditions better, and that is why it was the best performing EOM cavity in the plasmonic ones. Furthermore, employing near ENZ modulator off state in the ITO plasmonic structures can improve cavity optimal scaling by about an order of magnitude. This essentially means that similar performance cavity modulators can be achieved

in a reduced footprint of about an order of magnitude. For EAM operations, the more lossy plasmonic structures should be considered. ITO slot and HPP structures can offer superior performance, whereas the graphene slot is also noteworthy considering the atomic thin layer of graphene. Depending on the desired effects and scaling concerns, the proper material-modal combinations should be employed.

5 Supplementary information

Additional supporting information may be found in the online version of this article at the publisher's website.

Acknowledgments: V.S. is supported by the Army Research Office under the contract number W911NF-16-2-0194 and by the Air Force Office of Scientific Research-Young Investigator Program under grant FA9550-14-1-0215.

References

- [1] Sorger VJ, Oulton RF, Ma RM, Zhang X. Toward integrated plasmonic circuits. *MRS Bull* 2012;37:728–38.
- [2] Liu K, Ye CR, Khan S, Sorger VJ. Review and perspective on ultrafast wavelength-size electro-optic modulators. *Laser Photon Rev* 2015;9:172–94.
- [3] Liu K, Li N, Sadana DK, Sorger VJ. Integrated nanocavity plasmon light sources for on-chip optical interconnects. *ACS Photonics* 2016;3:233–42.
- [4] Theofanous NG, Aillerie M, Fontana MD, Alexakis GE. A frequency doubling electro-optic modulation system for Pockels effect measurements: application in LiNbO₃. *Rev Sci Instrum* 1997;68:2138–43.
- [5] Wooten EL, Kissa KM, Yi-Yan A, et al. A review of lithium niobate modulators for fiber-optic communications systems. *IEEE J Sel Topics Quantum Electron* 2000;6:69–82.
- [6] Bar-Joseph I, Klingshirn C, Miller DAB, Chemla DS, Koren U, Miller BI. Quantum-confined Stark effect in InGaAs/InP quantum wells grown by organometallic vapor phase epitaxy. *Appl Phys Lett* 1987;50:1010–2.
- [7] Liu A, Jones R, Liao L, et al. A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor. *Nature* 2004;427:615–8.
- [8] Xu Q, Schmidt B, Pradhan S, Lipson M. Micrometre-scale silicon electro-optic modulator. *Nature* 2005;435:325–7.
- [9] Reed GT, Mashanovich G, Gardes FY, Thomson DJ. Silicon optical modulators. *Nat Photon* 2010;4:518–26.
- [10] Ye C, Khan S, Li ZR, Simsek E, Sorger VJ. λ -Size ITO and graphene-based electro-optic modulators on SOI. *IEEE J Sel Topics Quantum Electron* 2014;20:40–9.
- [11] Feigenbaum E, Diest K, Atwater H. Unity-order index change in transparent conducting oxides at visible frequencies. *Nano Lett* 2010;10:2111–6.
- [12] Liu M, Yin X, Ulin-Avila E, et al. A graphene-based broadband optical modulator. *Nature* 2011;474:64–7.
- [13] Wang F, Zhang Y, Tian C, et al. Gate-variable optical transitions in graphene. *Science* 2008;320:206–9.
- [14] Ma Z, Tahersima M, Khan S, Sorger VJ. Two-dimensional material-based mode confinement engineering in electro-optic modulators. *IEEE J Sel Topics Quantum Electron* 2017;23:81–8.
- [15] Ye C, Liu K, Soref RA, Sorger VJ. A compact plasmonic MOS-based 2×2 electro-optic switch. *Nanophotonics* 2015;4:261–8.
- [16] Huang C, Lamond RJ, Pickus SK, Li ZR, Sorger VJ. A sub- λ -size modulator beyond the efficiency-loss limit. *IEEE Photon J* 2013;5:2202411.
- [17] Vasudev AP, Kang J, Park J, Liu X, Brongersma ML. Electro-optical modulation of a silicon waveguide with an “epsilon-near-zero” material. *Opt Express* 2013;21:26387–97.
- [18] Ye C, Li Z, Liu K, Soref R, Sorger VJ. Strong ITO index modulation for switching devices. In: *Advanced Photonics for Communications, OSA Technical Digest (online)* (Optical Society of America, 2014), paper JM3B.1.
- [19] Baek J, You JB, Yu K. Free-carrier electro-refraction modulation based on a silicon slot waveguide with ITO. *Opt Express* 2015;23:15863–76.
- [20] Shi K, Haque RR, Zhao B, Zhao R, Lu Z. Broadband electro-optical modulator based on transparent conducting oxide. *Opt Lett* 2014;39:4978–81.
- [21] Dionne J, Diest K, Sweatlock L, Atwater HA. PlasMOSstor: a metal–oxide–Si field effect plasmonic modulator. *Nano Lett* 2009;9:897–902.
- [22] Lee HW, Papadakis G, Burgos SP, et al. Nanoscale conducting oxide PlasMOSstor. *Nano Lett* 2014;14:6463–8.
- [23] Lafone L, Sidiropoulos TPH, Oulton RF. Silicon-based metal-loaded plasmonic waveguides for low-loss nanofocusing. *Opt Lett* 2014;39:4356–9.
- [24] Manolatu C, Lipson M. All-optical silicon modulators based on carrier injection by two-photon absorption. *J Lightwave Tech* 2006;24:1433–9.
- [25] Xu Q, Manipatruni S, Schmidt B, Shakya J, Lipson M. 125 Gbit/s carrier-injection-based silicon micro-ring silicon modulators. *Opt Exp* 2007;15:430–6.
- [26] Melikyan A, Lindenmann N, Walheim S, et al. Surface plasmon polariton absorption modulator. *Opt Exp* 2011;19:8855–69.
- [27] Ma Z, Li Z, Liu K, Ye C, Sorger V. Indium-tin-oxide for high-performance electro-optic modulation. *Nanophotonics* 2015;4:198–213.
- [28] Sorger V, Lanzillotti-Kimura N, Ma R, Zhang X. Ultra-compact silicon nanophotonic modulator with broadband response. *Nanophotonics* 2012;1:17–22.
- [29] Zhu S, Lo GQ, Kwong DL. Phase modulation in horizontal metal-insulator-silicon-insulator-metal plasmonic waveguides. *Opt Exp* 2013;21:8320–30.
- [30] Salamin Y, Heni W, Haffner C, et al. Direct conversion of free space millimeter waves to optical domain by plasmonic modulator antenna. *Nano Lett* 2015;15:8342–6.
- [31] Haffner C, Heni W, Fedoryshyn Y, et al. All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale. *Nat Photonics* 2015;9:525–8.
- [32] Soref RA, Bennett BR. Electrooptical effects in silicon. *IEEE J Quantum Electron* 1987;23:123–9.

- [33] Ansell D, Radko IP, Han Z, Rodriguez FJ, Bozhevolnyi SI, Grigorenko AN. Hybrid graphene plasmonic waveguide modulators. *Nat Comm* 2015;6:8846.
- [34] Khurgin JB. Slow light in various media: a tutorial. *Adv Opt Photon* 2010;2:287–318.
- [35] Sun S, Badawy A, Narayana V, El-Ghazawi T, Sorger V. The case for hybrid photonic plasmonic interconnects (HyPPIs): low-latency energy-and-area-efficient on-chip interconnects. *IEEE Photon J* 2015;7:1–14.
- [36] Ye Y, Wong Z, Lu X, et al. Monolayer excitonic laser. *Nat Photonics* 2015;9:733–7.
- [37] Tahersima M, Sorger V. Enhanced photon absorption in spiral nanostructured solar cells using layered 2D materials. *Nanotechnology* 2015;26:344005.
- [38] Gan X, Shiue R, Gao Y, et al. Chip-integrated ultrafast graphene photodetector with high responsivity. *Nat Photonics* 2013;7:883–7.
- [39] Li G, Krishnamoorthy A, Shubin I, et al. Ring resonator modulators in silicon for interchip photonic links. *IEEE J Sel Topics Quantum Electron* 2013;19:95–113.
- [40] Hanson G. Dyadic Green's functions and guided surface waves for a surface conductivity model of graphene. *J Appl Phys* 2008;103:064302.
- [41] Jun YC, Briggs RM, Atwater HA, Brongersma ML. Broadband enhancement of light emission in silicon slot waveguides. *Opt Exp* 2009;17:7479–90.
- [42] Saleh A, Dionne J. Waveguides with a silver lining: low threshold gain and giant modal gain in active cylindrical and coaxial plasmonic devices. *Phys Rev B* 2012;85:045407.
- [43] Veronis G, Fan S. Modes of subwavelength plasmonic slot waveguides. *J Lightwave Tech* 2007;25:2511–21.
- [44] Soref RA. Silicon-based optoelectronics. *Proc IEEE* 1993;81:1687–706.
- [45] Soref RA, Bennett BR. Kramers-Kronig analysis of electro-optical switching in silicon. *Proc SPIE* 1987;704:32–7.
- [46] Ray S, Banerjee R, Basu N, Batabyal AK, Barua AK. Properties of tin doped indium oxide thin films prepared by magnetron sputtering. *J Appl Phys* 1983;54:3497.
- [47] Tuna O, Selamet Y, Aygun G, Ozyuzer L. High quality ITO thin films grown by dc and RF sputtering without oxygen. *J Phys D Appl Phys* 2010;43:055402.
- [48] Liu K, Sun S, Majumdar A, Sorger VJ. Fundamental scaling laws in nanophotonics. *Sci Rep* 2016;6:37419.
- [49] Nielsen MP, Lafone L, Rakovich A, et al. Adiabatic nanofocusing in hybrid gap plasmon waveguides on the silicon-on-insulator platform. *Nano Lett* 2016;16:1410–4.
- [50] Oulton RF, Sorger VJ, Genov DA, Pile DFP, Zhang X. A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation. *Nat Photonics* 2008;2:496–500.
- [51] Almeida VR, Xu Q, Barios CA, Lipson M. Guiding and confining light in void nanostructure. *Opt Lett* 2004;29:1209–11.
- [52] Oulton RF, Bartal G, Pile DFP, Zhang X. Confinement and propagation characteristics of subwavelength plasmonic modes. *New J Phys* 2008;10:105018.
- [53] Jackson JD. *Classical electrodynamics*. 3rd ed. New York, John Wiley & Sons, 1999.
- [54] Kasap SO. *Optoelectronics and photonics: principles and practices*, 2nd ed. Englewood Cliffs, New Jersey, Prentice-Hall, 2013.
- [55] Koch TL. In: 8th IEEE International Conference on Group IV Photonics, London, 2011, 112–4.
- [56] Lin H, Ogbuu O, Liu J, Zhang L, Michel J, Hu J. Breaking the energy-bandwidth limit of electrooptic modulators: theory and a device proposal. *J Lightwave Tech* 2013;31:4029–36.
- [57] Sun S, Badawy A, Narayana V, El-Ghazawi T, Sorger VJ. In: *Advanced Photonics 2016 (IPR, NOMA, Sensors, Networks, SPPCom, SOF)*, OSA technical Digest (online), paper ITu2B.6. Optical Society of America, 2016.
- [58] Vahala KJ. Optical microcavities. *Nature* 2003;424:839–46.
- [59] Armani DK, Kippenberg TJ, Spillane SM, Vahala KJ. Ultra-high-Q toroid microcavity on a chip. *Nature* 2003;421:925–8.
- [60] Lee H, Chen T, Li J, et al. Chemically etched ultrahigh-Q wedge-resonator on a silicon chip. *Nat Photonics* 2012;6:369–73.
- [61] Ma Z, Amin R, Khan S, Tahersima M, Sorger VJ. Temperature dependence of a sub-wavelength compact graphene plasmon-slot modulator. *arXiv:1709.01465*.
- [62] Shainline JM, Buckley SM, Mirin RP, Nam SW. Superconducting optoelectronic circuits for neuromorphic computing. *Phys Rev Applied* 2017;7:034013.

Supplemental Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/nanoph-2017-0072>).