Scaling vectors of attoJoule per bit modulators

To cite this article: Volker J Sorger et al 2018 J. Opt. 20 014012

View the article online for updates and enhancements.
Scaling vectors of attoJoule per bit modulators

Volker J Sorger, Rubab Amin, Jacob B Khurgin, Zhizhen Ma, Hamed Dalir and Sikandar Khan

1 Department of Electrical and Computer Engineering, George Washington University, 800 22nd St., Science & Engineering Hall, Washington, DC 20052, United States of America
2 Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, United States of America
3 Omega Optics, Inc., 8500 Shoal Creek Blvd., Bldg. 4, Suite 200, Austin, TX 78757, United States of America

E-mail: sorger@gwu.edu

Received 14 September 2017, revised 23 November 2017
Accepted for publication 29 November 2017
Published 27 December 2017

Abstract
Electro-optic modulation performs the conversion between the electrical and optical domain with applications in data communication for optical interconnects, but also for novel optical computing algorithms such as providing nonlinearity at the output stage of optical perceptrons in neuromorphic analog optical computing. While resembling an optical transistor, the weak light–matter-interaction makes modulators 10^5 times larger compared to their electronic counterparts. Since the clock frequency for photonics on-chip has a power-overhead sweet-spot around tens of GHz, ultrafast modulation may only be required in long-distance communication, not for short on-chip links. Hence, the search is open for power-efficient on-chip modulators beyond the solutions offered by foundries to date. Here, we show scaling vectors towards atto-Joule per bit efficient modulators on-chip as well as some experimental demonstrations of novel plasmonic modulators with sub-fJ/bit efficiencies. Our parametric study of placing different actively modulated materials into plasmonic versus photonic optical modes shows that 2D materials overcompensate their miniscule modal overlap by their unity-high index change. Furthermore, we reveal that the metal used in plasmonic-based modulators not only serves as an electrical contact, but also enables low electrical series resistances leading to near-ideal capacitors. We then discuss the first experimental demonstration of a photon-plasmon-hybrid graphene-based electro-absorption modulator on silicon. The device shows a sub-1 V steep switching enabled by near-ideal electrostatics delivering a high 0.05 dB V^−1 μm^−1 performance requiring only 110 aJ/bit. Improving on this demonstration, we discuss a plasmonic slot-based graphene modulator design, where the polarization of the plasmonic mode aligns with graphene’s in-plane dimension; where a push–pull dual-gating scheme enables 2 dB V^−1 μm^−1 efficient modulation allowing the device to be just 770 nm short for 3 dB small signal modulation. Lastly, comparing the switching energy of transistors to modulators shows that modulators based on emerging materials and plasmonic-silicon hybrid integration perform on-par relative to their electronic counterpart parts. This in turn allows for a device-enabled two orders-of-magnitude improvement of electrical-optical co-integrated network-on-chips over electronic-only architectures. The latter opens technological opportunities in cognitive computing, dynamic data-driven applications systems, and optical analog computer engines including neuromorphic photonic computing.

4 This article belongs to the special issue: Emerging Leaders, which features invited work from the best early-career researchers working within the scope of the Journal of Optics. Professor Volker Sorger was selected by the Editorial Board of the Journal of Optics as an Emerging Leader.
Electro-optic modulator challenges

Electro-optic modulation is a key function in modern data communication as it performs the conversion between the electronic data originating from computing cores, to the optical domain of low-loss data routing. While this function is universally used around the globe in long-haul, metro, and short-haul communications [1], such as data centers [2, 3], the case for on-chip optical interconnects was made [4] mainly to address the widening discrepancy between the data handling capability of electronic cores versus delays and power overheads in the communication-handling network-on-chip [5–8].

The parallels between an electro-optic modulator (EOM) and a field-effect-transistor (FET) however are noticeable; both control a ‘channel’ via an electrostatic gate. The discrepancies of the physical device lengths are, nevertheless, significant, as state-of-the-art silicon photonics modulators are of millimeter dimensions [9], while FETs are just tens of nanometers short. The reason for this is known, and lies in the weak light–matter–interaction (LMI), and inefficient material ability to change its optical refractive index upon applying a gate bias [10–14].

Given the resemblance of an EOM to an FETs, we interchangeably use the words ‘switching’ and ‘modulation’, while accepting the slight discrepancy—switching technically refers to a strict 2-level system, while modulation is an analog function. In reality, both EOMs and FETs have analog transfer functions, hence justifying the terminology definition used here.

This work focuses on charge-driven electro-absorption modulators (EAM) only, as opposed to electric field-driven designs, such as those based on Franz–Keldysh or Pockels’s effect [15–18]. We mainly discuss three active materials only, namely silicon, indium–tin-oxide (ITO) and graphene, but briefly mention results regarding expected absorption of other two dimensional (2D) materials [19–21], quantum dots [22], and quantum wells [23, 24]. The discussion thence includes a parametric study of achievable optical effective index changes as a function of the optical mode overlap factor with the active material, the materials own index change potential, and the effective mode’s group index with bias. We then show experimental results of a hybrid-photon-plasmon modulator based on graphene on silicon photonics, and a mode-overlap improved dual-gated graphene plasmon-slot waveguide modulator design allowing for sub-1 μm short device lengths, realizing ~4 dB μm⁻¹ strong signal modulation. Lastly, we compare our plasmonic silicon EOMs with FETs showing an energy-convergence of these two classes of devices.

Optoelectronic devices are significantly more bulky compared to their electronic counterparts. For instance, electrooptic modulators based on silicon’s plasma dispersion in photonic integrated waveguide modes are several millimeter long in order to obtain the desired phase shift leading to amplitude modulation in Mach–Zehnder-interferometers (MZI) [9].

Three-terminal switching devices usually consist of a ‘source’, a ‘drain’ and some form of a control ‘gate’. In FET both the channel and the gate are electronically controlled. The reason why this is possible within only tens of nanometers is based on the fact that both the channel (electron current) and the gate (electron-loading via a capacitive gate) have the same length-scale, namely their spatial extents are bound by their fermionic wave function being on the order of nanometers. Photons being bosons, on the other hand, do not interact with one-another. The only option for them to interact is via matter in a nonlinear way. This means that the opto-electronic response of a device is fundamentally governed by the ability and efficiency of the photons (or plasmons) to interact with the electronic wavefunctions of matter.

Cavity resonance for modulation

The fundamental inefficiency of optoelectronics, however, is that the spatial length scales of the electronic versus photonic wavefunctions are three orders of magnitude apart, considering wavelengths in the visible or near IR range. Incidentally, here we only consider telecom wavelengths, namely 1550 nm (0.8 eV). This large mismatch is the physical cause of the weak interaction between light and matter, and has technologically led to bulky opto-electronics with low chip integration densities [25]. Thus, in order to shrink down (i.e. ‘scale’) the device lengths and footprints, one has two general options to increase the weak light–matter-interactions (LMI, figure 1); (a) one can either increase the photon lifetime using resonators, aiming to enhance the cavity quality-Q factor, or, (b) one can enhance the electric field density non-resonantly by shrinking the optical mode possibly beyond the diffraction limit of light by deploying polaritonic modes. This is possible for instance by exploiting discontinuities in the permittivity leading to plasmonics, or metal optics [26–34]. High-Q cavities, however, introduce several technological disadvantages, but their spectral sensitivity allows reducing the applied voltage needed to shift the effective mode’s index in and out of resonance of the cavity, thus lowering the required drive voltage for modulation [35, 36]. However, the spectral sensitivity gains are overshadowed by spectral tuning overheads in form of thermal tuning heaters, thus increasing the devices’ dynamic power consumption. Moreover, the long photon lifetimes (τ photon = Q nm λ0/2πc, where λ0 is the operation wavelength of the device, nm the cavity’s modal refractive index, and c the speed of light in vacuum) in cavities lead to slow electro-optic modulation response times. For example, a ring resonator with a Q-factor of 40 000 results in modulation frequencies of only ~12 GHz at telecom wavelengths, which is below the current industry standard.

However, it is interesting to ask what net-benefits resonant photonic structures do enable. For instance, the footprint of a modulator can be lowered by folding the linear length of
a MZI-based modulator into a cavity. Here, the footprint of a cavity-based EOM can be lowered by a factor proportional to the resonator’s finesse, \( F \) [19, 20]. In brief, the lengths of the MZI and ring are given by \( L_{\text{equiv-MZ}} = L \cdot \lambda / 2n_g \) 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 proportional to the finesse of the resonator. Secondly, the sensitivity of the MZI is actually equal to that of the ring EOM, which is somewhat surprising, however can be seen by the following arguments; \( \Delta \nu \approx \Delta \nu_{\text{FWHM}} = \frac{1}{2\pi \tau_{\text{ph}}} \) with \( \Delta n / n_{g} \approx \Delta n / n_{g} \) requires \( \Delta n = \Delta \nu_{\text{FWHM}} n_{g} \), where \( \Delta \nu_{\text{FWHM}} \) is the resonator linewidth, \( n_{g} \) the group index, \( \Delta n \) the modulated index (i.e. a function of applied voltage), and \( \tau_{\text{ph}} \) the photon lifetime in the resonator. A single MZI arm requires an index change of \( \Delta n = \lambda / 2L \), extending the MZI length to be \( L = \frac{Q \cdot \lambda}{2n_{g}} \), which equals the same sensitivity as the ring design. Thirdly, a question to ask is, whether a long linear device (i.e. larger capacitance) or a 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 MZI is not parasitic capacitance limited as shown next. The MZI transit time, \( T_{\text{tr}} \) (treating the MZI as a lumped-element with length \( L \)) is given by \( T_{\text{tr}} = \frac{\lambda}{2n_{g}} L \) and hence the transit-time limited bandwidth is \( (f_{\text{dB}})^{\text{MZ}} = 0.44 \cdot \frac{1}{T_{\text{tr}}} \approx \frac{1}{2\pi \tau_{\text{PH}}} = \frac{1}{2n_{g} L} = \frac{Q}{Q} \). The cavity’s photon-limited bandwidth due to the long photon lifetime can be estimated via \( (f_{\text{dB}})^{\text{ring}} \approx \frac{1}{2\pi \tau_{\text{PH}}} \), closing the above argument that the MZI and ring (or Fabry–Pérot (FP)) cavity EOMs have about the same modulation speed. For example, a modulator with a \( Q \) of 10,000 has a cut-off speed of about 34 GHz (for an electro-optic coefficient of 300 pm V\(^{-1}\), and extinction ratio = 3 dB).

We recently investigated the effect of a cavity on both the shift in resonance wavelength, \( \Delta \lambda \) corresponding to the change (tuning) in the effective refractive index (real part), \( \Delta n_{\text{eff}} \) from the modal tuning of the underlying waveguide mode [19, 20]. However, tuning also increases loss, \( \Delta \alpha \), as a direct result from the Kramers–Kronig relations. Thus, the design challenge is to optimize the ratio of obtainable cavity tuning which improves the modulators extinction ratio, ER, (i.e. modulation depth) relative to these incurred losses. Hence an appropriate cavity-based EOM figure of merit is FOM\(_{\text{EOM-cavity}} = \Delta \lambda / \Delta \alpha \), i.e. the change in the loss, \( \Delta \alpha \) is a function of the modal effective extinction coefficient change, \( \Delta \kappa_{\text{eff}} \). A timely example of a deployment of this
FOM is the target EOM values of the AIM Photonics consortium’s device roadmap [37].

With respect to cavity-enhanced FOM performance, one can show that longer devices perform better when optical losses are low [19, 20], as known from millimeter-long foundry designs [9]. However, there are multiple challenges with this approach: (a) their sizable footprints on the order of millimeters lead to not insignificant insertion losses, IL, and (b) obtaining high-speed circuit designs are challenging given that traveling-wave designs must be adhered to. While sub-volt driving voltages switching at tens of GHz have been demonstrated [24], the device capacitance of these sizable modulators, limits their potential for sub 1/bit efficiency required for next generation modulators [25]. That is why in this work, we focus on LMI enhancements using field enhancements in sub-diffraction limited waveguides without resonance enhancement. The latter also enables spectral broadband devices to allow these modulators to be used in wavelength-division-multiplexing (WDM) photonic circuitry (figure 1). Interestingly, the ratio of both, namely a high Q-factor and a small mode-volume, $V_m$, are proportional to the Purcell factor—a merit that can be regarded as an optical concentration factor [38]. Next, we discuss how this factor allows predicting nanophotonic scaling laws.

**Modulator scaling laws**

The hypothesis that ‘smaller-is-better’ has motivated optical engineers to build various nanophotonic devices in an ad-hoc manner thus far. Here, we analyze scaling laws of EOMs with a focused interest in the micrometer to sub-micrometer scale. With technology options such as plasmonics [11–14], nanoscale dielectric resonators [39, 40] and slot-waveguides [41], we are able to surpass the diffraction limit of light by engineering the effective refractive index. Still, decreasing the optical mode volume, $V_m$, introduces adverse effects, such as bending- and ohmic losses for polaritonic modes. It is therefore not straightforward to predict modulator performance scaled into sub-micron size regime [42], leading to a rigorous analysis of fundamental scaling laws for nanophotonics as a function of critical device length [38]. In this scaling analysis, we assume three types of optical cavities, (a) a traveling-wave ring resonator (RR) [43], (b) a metal-mirror based FP cavity [39], and (c) a plasmonic metal nano-particle (MNP) [44] (figure 2(b)), that enhance the fundamentally weak interaction between light and matter via the ratio of $Q/V_m$, where $Q$ is the cavity quality factor, $V_m$ is the effective volume of electromagnetic energy of a resonant mode. An interesting, although expected result is that all cavity types do not perform equally well for vanishing critical dimension due to their respective non-monotonic Purcell factor scaling. The critical length for these cavities are the radius for the RR and the MNP, and the physical distance between two mirrors for the FP. Analytical expressions for both the cavity quality factor $Q$ and the optical mode volume $V_m$ for the RR and FP cavities are given in [38]. The resulting Purcell factor, defined as $F_p = \frac{1}{4\pi} \left( \frac{\lambda_R}{2\lambda} \right)^2 \left( \frac{\lambda}{\lambda_R} \right)$, where $\lambda_R$ is the resonant wavelength of the cavity, and $n$ is the cavity material refractive index shows a significant influence on the EOMs power consumption, $E/\text{bit} \sim (F_pQ)^{-1}$ [38]. This can be understood from the dimensional schematic of an EOM (figure 2(a)); the required electric field, $E/\text{field} = V_{bias}/h$, to obtain a desired bit-errortrate (BER) at the detector downstream and the device capacitance, $C = \varepsilon_0 A / \ell = (C = \text{capacitance}, A = \text{lateral device area}, \ell = \text{gate oxide}, \varepsilon_0$ and $\varepsilon_\alpha = \text{relative permittivity of the gate dielectric and free space permittivity, respectively})$ determine the energy efficiency, $E/\text{bit} = 1/2 CV^2$ of an EOM (figure 2(a)). Once substituted, we find that $E/\text{bit}$ is proportional to the physical volume of the modulator divided by the cavity $Q^2$, or $E/\text{bit} \sim (F_pQ)^{-1}$. $Q$ for the three selected cavities overall decreases with scaling as expected since light is either less confined (i.e. bending losses), or polaritonic (i.e. matter-like) bound to the metal’s electron sea, thus increasing loss. The mode volume, on the other hand, scales linearly for the one-dimensional scaling for microrings, but cubically for the plasmon cavity. The FP cavity exhibits an inverse scaling, due to the mode character changing from traveling wave to a metal–insulator–metal (MIM) plasmon mode once both mirrors start to couple. The scaling for $Q$ and $V_m$ thus shows optimia for the Purcell factor, which quantify the point of maximum feedback (highest $Q$, low losses), while offering high optical confinement. Beyond $F_p$’s maxima, the parasitic losses of the cavity become dominant (figures 2(c)–(e)). Then, the lowest $E/\text{bit}$ matches the $F_p$ maxima well, where discrepancies originate from the strong $\sim Q^{-2}$ dependency (figure 2(f)). The results show that devices (optimized for energy consumption only) are high-$Q$ EOMs, which are, however spectrally sensitive, tens of micrometers in footprint, have slow response times, and require thermal tuning. The latter was not taking into account in this analysis here. Based on these limitations, plasmonic modulators offer an interesting alternative as they can approach 100’s of $\text{aJ}/\text{bit}$ efficiencies while allowing for sub-micron short device lengths, which supports small electrical capacitances enabling fast switching, provided the active material allows for such.

A modulator’s key performance is the ability to change the waveguide mode’s effective index most efficiently, i.e. with the lowest voltage bias. Thus, the aim is interestingly similar to that of FETs, where the steepness of the current–voltage transfer function is quantified by the sub-threshold swing, i.e. the amount of voltage change required to induce a 10-fold current change. For EOMs, this translates into an effective modal index change for applied voltage bias.

Similar to optical gain building devices like lasers [26, 28, 30, 31, 46–48], EOMs require optimization with respect to the modal overlap, $\Gamma$, with the active material [45, 49, 50]. A high extinction ratio (ER) does critically depend on the obtainable index change upon biasing the device. For instance, for electro-optic (phase) modulation the effective change of the $k$–vector is given by $\delta k = \delta \varepsilon \varepsilon_0 k_0 n_e \frac{n_e}{n} = \frac{\delta n_e}{n} \frac{\delta k_0}{k_0} n_e = \frac{\delta n_e}{n} \kappa_0 \frac{\Delta n}{n} L$, where $\kappa_0$ is the group index in the waveguide mode, and $\delta n$ the modulated index change. The phase change then becomes $\Delta \phi = \frac{2\pi}{\lambda} \Delta n \ell L = \varepsilon_\Gamma \ell \Delta n \kappa_0 \frac{\Delta n}{n} L$, where $\Gamma$ is the optical confinement factor and
Scaling laws of electro-optic modulators, answering the question whether scaling improves modulator performance similarly to transistor scaling. (a) Electro-optic modulator schematic showing spatial dimensions and light propagation. (b) We explore three device-underlying cavity types: a ring resonator cavity with the waveguide width, \( w \), and radius, \( r \); a Fabry–Pérot (FP) cavity comprised of a dielectric material sandwiched by a pair of highly reflecting metal mirrors (reflectivities \( R_1 \) and \( R_2 \)); and a plasmon cavity formed by metal nanoparticle (MNP) embedded in a dielectric, where \( a \) is the particle radius. The modulator scaling parameters are \( r \) (RR), \( l \) (FP), and \( a \) (MNP) cavity. (c)–(e) Cavity performance as a function of scaling. (c) Quality \((Q)\) factor, (d) optical mode volume, \( V_m \), and (e) Purcell factor, \( F_p \). While the general trend shows a reduced \( Q \) upon scaling, significant differences between the three cavity types exist. A quality factor below 1 (overdamped regime) is strictly unphysical and given by the decay rate. Parameters: propagation loss of a diffraction limited beam, \( \alpha_p = 1.0 \text{ dB cm}^{-1} \) used in the RR; Silver metal mirrors, \( \tilde{n}_{Ag} = 0.41 + 10.05i \), the dielectric refractive index \( \tilde{n}_D = 3.0 - 0.001i \), Silver conductivity \( \sigma_{Ag} = 6.3 \times 10^7 \text{ mho m}^{-1} \), and the damping rate for the MNP to be \( \gamma_d = 2.0 \times 10^{15} \text{ rad s}^{-1} \). Each cavity type shows a maximum Purcell factor given by the ideal optical confinement-to-loss point. (f) The energy efficiency of the modulator \((E/\text{bit})\) scales with \( \sim (QF_p)^{-1} \). Thus, improving the optical confinement helps to reduce the energy efficiency, while a higher \( Q \) offers further support. 100’s of atto-Joule efficient modulators are possible with pure plasmonic modes allowing for high-density optical circuits due to sub-micron compact critical device lengths. Reproduced from [38]. CC BY 4.0.
Figure 3. Optical mode exploration for studying the material-field interactions with the modal overlap factor and the effective mode’s group index upon modulation. Materials considered are traditional silicon and emerging materials (ITO, graphene). (a)–(i) Schematic of the mode structures. The relevant parameters are $\lambda_{Si}/\lambda_{b}$ = 451 nm, $\lambda_{hSi}/\lambda_{b}$ = 200 nm, $d_{Si} = h_{Si} = 300$ nm, $d_{hSi} = 100$ nm, $w_{Si} = 550$ nm, $h_{Si}$ = 20 nm, $w_{iTO}$ = 300 nm, $h_{iTO}$ = 10 nm and $w_{hTO}$ = 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 a thickness of $h_{ox} = 5$ nm to ensure similar electrostatics [19, 20]. (a’)–(i’) Respective optical field profiles from FEM simulations for all the structures at their respective starting point in the material dispersion at $\lambda = 1550$ nm.

$\epsilon_{ri} = \int_{\epsilon_{r}}^{\epsilon_{i}} \delta n_{eff} \, dx$. The index change inside the modulator ($\Delta n_{eff}$) is then given by $ER \propto \Delta n_{eff} = \epsilon_{ri} \delta n_{eff} \, \frac{d_{eff}}{d_{true}}$ [45, 49, 50]. For discrete modulation states, this relationship can be expressed as the ratio of the active material index relative to its initial condition ($\delta n_{eff} = \Delta n_{mat}/n_{true}$) multiplied by its modal confinement factor ($\Gamma$), relative modal permittivity enhancement ($\epsilon_{ri}$) and effective group index ($n_{eff}$) [45]. Here, the group index $n_{g}$ corresponds to dispersive propagation in the longitudinal direction given by $n_{g} = n_{eff} - \lambda \frac{dn_{eff}}{dx}$, which applies to isotropic index materials such as silicon and ITO based structures. Due to the unique electro-optic nature of graphene and anisotropy of its indices (tensor), this simple equation insufficiently describes modulation performances in the graphene-based modulators. Note that graphene’s 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. 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. Next, we discuss the various optical modes considered here, while focusing on the confinement factor, and finally discuss obtainable effective index changes governing modulation performance [49, 50].

We study three different mode structures for each of the three active materials introduced above (figure 3), while our aim is to explore modulator-suitable material-mode combinations for electro-absorption modulation mechanisms. The target is to increase the LMI towards ultra-compact modulators while preserving ER. 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 MIM configuration, and a hybrid plasmonic polariton (HPP) design in a metal-insulator-semiconductor configuration [26–34]. In order to understand the LMI enhancement effect from modal compression, we compare each active material with a bulk case where the waveguide consists of the active material only. The resulting design space is a 3 × 3 matrix, where we capture both the modal overlap factor and the group index as a function of carrier concentration for silicon and ITO, whereas for graphene as a function of chemical potential (figures 4 and 5).

As expected, the overlap factor for a bulk silicon-based modulator is rather high (~80%) but does not change significantly with carrier concentration. Hybridizing silicon as active material with plasmonics worsens the overlap since the optical field is partly confined inside the plasmonic slot and not in the actively tuned silicon layer underneath (figure 4(a)). Photonic hybridization provides performance between these two extreme cases. Changing from silicon to ITO shows a bias (i.e. carrier-sensitive) overlap factor approaching 70% near ITOs epsilon-near-zero region close to $7 \times 10^{-20}$ cm$^{-3}$ [51, 52]. Such change helps in EAMs where absorption in the lossy OFF state (high carrier concentration) should have a high overlap factor, while the light ON state losses should be reduced. Since the ITO’s capacitive-gated index change occurs only at a thin 5–10 nm layer corresponding to a graded-index accumulation layer, the bulk modes do not provide any advantage for ITO modulators (figure 4(a)). Graphene’s atomistic cross-section naturally leads to low overlap factors about 2 orders lower compared to ITO (figure 4(b)), but can reach almost 1% for a single graphene layer, in slot-waveguide structures (figure 5(c)) [49]. Generally for all modulators, the effective group index change with voltage swing, $\Delta n_{g}$, should be maximized. For bulk modes this dispersion, or slow light-effect, while present, is relatively weak.
\[ S(x) = \frac{\omega^2(x)cE_0^2}{2\beta} = \frac{n^2(x)E_0^2}{2n_a\eta_0}, \]

where the effective index is \( n_{\text{eff}} = \beta c/\omega \), and the free space impedance, \( \eta_0 \). The total power flow is then

\[ P = W \int_{-\infty}^{\infty} S(x)\,dx = \frac{W}{2n_a\eta_0} \int_{-\infty}^{\infty} n^2(x)E_0^2\,dx = \frac{n_aE_0^2}{2\eta_0} W_{\text{eff}}, \]

where \( n_a \) is the refractive index and \( E_a \) is the transverse electric field in the active layer. The effective thickness, \( t_{\text{eff}} \), a value proportional to the modal overlap factor of the active material with the waveguide mode, is given by:

\[ t_{\text{eff}} = \frac{1}{n_a n_\text{eff}} \int_{-\infty}^{\infty} n^2(x)E_0^2\,dx, \]

where \( E_a \) and \( n_a \) are the maximum field in the waveguide cross-section and the index at that location, respectively. Note that the former does not necessarily need to be at the center of the waveguide, thus allowing for a wide variety of designs. The \( x \)- and \( y \)-directions are the cross-section plane of a waveguide mode propagating in the \( z \)-direction. From here, one can ask what is the absorption of a given active material, and we find a universal dependency of the latter on the effective thickness, \( t_{\text{eff}} \), and the absorption cross-section, \( \sigma_{\text{ab}} \). The explicit expression of \( \sigma_{\text{ab}} \) however depends on the material dependent absorption mechanism and will be reported elsewhere.

Next follows a discussion of graphene as per the example; the intensity \( I_L = E_a^2/2\eta_0 \) depends on the maximum field \( E_a \) and fundamental constants such as the free space impedance, \( \eta_0 \), and the modulated intensity change as:

\[ \delta I_L = \frac{\hbar \omega}{\eta_0} \frac{dN_{\text{gr}}}{dt} \frac{1}{N_{\text{gr}}} = \frac{e^2E_0^2}{8\hbar}N_{\text{gr}}[1 - f_c] \]

where \( N_{\text{gr}} \) is the number of graphene sheets, and \( f_c \) the Fermi–Dirac function. Then, the absorption coefficient inside a waveguide is given as:

\[ \alpha_{\text{gr}} = \frac{\pi \alpha_0 N_{\text{gr}}}{n_a} \frac{1}{t_{\text{eff}}} \frac{f_c}{1 - f_c}. \]
where $\alpha_0$ is the fine structure constant. The absorption is a function of the injected two-dimensional carrier concentration in graphene, $n_{2D}$ ($t_{ox} = 100 \text{ nm}$ was used, figures 5(a), (b)). We find that various materials not only have (trivially) different absorption values for a given voltage (or carrier concentration), but more importantly that they exhibit inverse scaling trends; for instance free-carriers in silicon or ITO increase absorption, since more carriers simply increase the loss in the Drude formula. For quantum dots (QD), 2D materials and quantum wells (QW), the trend is inverse since more carriers occupy states elsewise available for electron-hole pairs upon absorption. For such state-filling materials, the trend for absorption versus voltage is similar, however appears with different magnitude and voltage scaling. The latter ‘modulation steepness’ is, however, relevant from an efficiency point of view similar to Landau limit of 60 mV dec$^{-1}$ in transistors; that is the less voltage for switching, the lower the dynamic power consumption of the modulator. Here, quantum dots and wells, and graphene perform particularly steep, while the highest absorption is found for WSe$_2$, a 2D material representing the class of transition metal dichalcogenides (TMD). This state blocking in graphene is referred to as Pauli-blocking [53–55]; here a photon can only be absorbed when an electron transition occurs which requires an empty state in the conduction band. If the Fermi level is above the energy level given by the sum of the electron’s initial state and the photon energy, all states are filled, and the photon is not absorbed, ‘blocked’. The absorption steepness, thus depends on the energy-sharpness of the density of states, which at first order is $\sim k_B T = 10$–100s meV, where $T$ is the temperature, and $k_B$ the Boltzmann constant.

The weak index modulation of silicon is exemplified in figure 5(c), by requiring close to unity-high overlap factors in order to obtain a reasonable modulation strength, defined as the absorption change upon modulation [49]. However, for

Figure 5. Material impact on absorption modulation. (a)–(b) Material-based optical absorption for electro-absorption modulators versus charge (i.e. carrier concentration, $n_{2D}$). Free-carrier-based absorption (silicon, ITO) scales with carrier concentration (and bias, (b)) due to the Drude model. All other modulation mechanisms rely on some form of ‘band-filling’ leading to absorption blocking, such as the Pauli blocking in graphene. 2D material TMDs show in general the highest absorption (see main text), but quantum wells and quantum dots offer the steepest switching (low voltage). (c) The modulation performance of an electro-absorption modulator improves with overlap factor. However, the weak effective mode index change of silicon is outperformed by emerging materials such as ITO or graphene despite the low optical overlap factor [49]. $\lambda = 1550 \text{ nm}$. 

\[ J. \text{ Opt.} \ 20 \ (2018) \ 014012 \ V \ J \text{ Sorger et al} \]
realistic on-chip waveguide designs, the modal overlap can never reach 100% even in bulk modes, since some amount of field always leaks into the cladding (see forbidden region, figure 5(c)). Turning to emerging active materials such as ITO and graphene for modulation shows that the strong index modulation of unity order is able to compensate for their small overlap factor. Graphene in particular is remarkable that modulation of unity order is able to compensate for their realistic on-chip waveguide designs, the modal overlap can never reach 100% even in bulk modes, since some amount of field always leaks into the cladding (see forbidden region, figure 5(c)). Turning to emerging active materials such as ITO and graphene for modulation shows that the strong index modulation of unity order is able to compensate for their small overlap factor. Graphene in particular is remarkable that a 0.35 nanometer thin sheet of material with an overlap factor if 0.5 × 10−4 provides a stronger modulation than the best silicon mode ever can [53]. A single 2D material like graphene is able to show an overlap factor approaching 1% when optimized for in-plane polaritons such as in slot-waveguides (blue squares, figure 5(c)). In fact, stacking multiple layers of 2D materials may be an interesting approach to increase the modulation efficiency further, enabled by overlap factors of up to 40%.

Since there has been much interest in 2D materials beyond graphene for opto-electronics due to their unique excitonic properties originating from their high anisotropy given by the contrasting dimensionality [56–58]. The high exciton binding-energy in 2D materials originates from the reduced amount of coulomb screening leading to high absorption. Therefore, it was suggested that such ‘robust’ excitons can be used for efficient light modulation. Here, the 2D material exciton is characterized by its 2D Bohr radius

\[ a_{\text{ex}} = \frac{2\pi \rho_{\text{eff}} r_{\text{eff}}^2}{e^2 m} \]  

(6)

which is 50% as large as the radius of 3D-material excitons. The effective dielectric constant \( \epsilon_{\text{eff}} \) in 2D materials approaches unity while the effective mass \( m_{\text{eff}} \) is somewhat larger than in III–V semiconductors, thus given a TMD exciton Bohr radius of just a few nanometers leading to a high corresponding exciton binding energy of 0.5–0.7 eV [59].

\[ E_{\text{ex}} = \frac{\hbar^2}{2m_{\text{ex}}} \sim 0.5 \text{ eV}. \]  

(7)

The absorption, \( \alpha \), of the exciton can be obtained by using the value of the exciton envelope wavefunction at the origin resulting in

[\[ \alpha(\omega) = 4\sigma'(\omega)/\pi\epsilon_{\text{eff}}^2 \]]

operating on the excitonic resonance

[\[ \sigma'(\omega) = 4\epsilon_0\epsilon_{\text{eff}} \frac{\omega}{\omega - \omega_{\text{ex}}} \approx \frac{2\pi\alpha_0}{\hbar} \frac{\hbar}{m}\omega \]]

where the dipole transition is given by

[\[ r_{\text{ex}}^2 = \frac{p_{\text{ex}}}{m_0^2\omega^2} = \frac{\hbar}{2m_{\text{ex}}\omega} \]]

Modulating 2D materials results in (i) state filling, (ii) bandgap renormalization, and (iii) screening. As a result the exciton bleaches due to Mott transition with a screening radius comparable to the exciton radius. Therefore, exciton bleaching most likely takes place because of state filling. The exciton wavefunction can be considered a coherent superposition of the electron hole pair states with wavevectors ranging from 0 to \( \sim (\alpha_{\text{ex}})^{-1} \) density of these states. Here, the exciton radius and binding energy do not impact the switching charge significantly. Since the effective mass in TMDs is typically larger than in III–V semiconductors, using excitons does not change the fundamental fact that each time a single electron is injected inside the active layer a single transition is being blocked, and, given the fact that the oscillator strength for each allowed transition is about the same, the expected obtainable absorption change is roughly similar.

Temperature effects in graphene

Given the high modulation potential of graphene, we next consider more modulation physics relating to graphene’s Pauli blocking electro-optic modulation mechanism. Graphene is an anisotropic material given its dimensions: in its honeycomb like lattice plane, the in-plane permittivity (\( \epsilon_{\parallel} \))
can be tuned by varying its chemical potential, $\mu$, whereas the out-of-plane permittivity is reported to remain constant around 2.5 \cite{19, 20}. We model graphene with two different temperatures by the Kubo model at $T = 0$ K and $T = 300$ K (figure 6). At higher temperatures, the imaginary refractive index versus chemical potential is smeared due to the natural temperature dependency of the Fermi–Dirac distribution function. Upon doping (i.e. biasing) graphene to a chemical potential near half of the photon’s energy, a small switching energy is needed for of only $>30$ meV for $T = 0$ K versus $~100$ meV at $T = 300$ K (figure 6). This difference in the minimum voltage of about $3\times$ is equivalents to an energy saving of about 10-fold improvement of devices when operated at cryo-temperatures. However, the voltages required for actual devices needs to be considered for modulating from zero-chemical potential to the Pauli-blocking regime of about $\mu_c = 0.4$–0.5 eV. In a real device, one would only modulate around the point of half the photon energy, here $\frac{\hbar \nu}{2} = 0.4$ eV, given the telecom wavelength. The latter introduces a voltage drop across the contacts, lowering the actual applied voltage range across the device capacitor. The metal contact from a plasmonic modulator offers a unique advantage over photonic devices, since the drive voltage suffers no degradation in the contacts.

Graphene’s optical absorption ($\alpha$) for telecom wavelength ($\lambda = 1550$ nm) relates to an optical transmission change with bias (figures 7(a) and (b)). Graphene placed on a photonic mode separated by a thin oxide creates an electrical capacitor (figure 7(d)). The voltage steepness required to exploit the Pauli blocking modulation is however relatively high, $<0.1$ dB V$^{-1}$ when weak LMI photonic modes are used (figure 7(b)) \cite{60}. With the aim of creating a steepest electro-optic switching behavior, obtainable performance dependencies based on the quality of both the electrostatics and the ability to deliver a voltage change to the optical mode without resistive losses, i.e. a low contact resistance is desired. This is in fact identical, from an electrostatics point of view, to FET device physics, where the switching steepness is quantified by the sub-threshold swing. This electrostatics can be optimized in two ways; (i) improving the gate’s control of the electro-optic effect, and (ii) reducing the series resistance. Thus, a higher capacitor while hurting the RC-delay of the modulator, is actually desired, since it enables a steeper switching transfer function. Practically, this can be realized by reducing the capacitor’s oxide thickness, and/or introducing high-$k$ dielectrics. Secondly, the applied bias voltage drop should only appear at the capacitor (the device itself), and not at the contacts, or channel leading to the device. This means that photonic modulators are fundamentally challenged by the required doping to reduce series resistance near the contact regions (figure 7(d)). Plasmonic devices, in contrast, use their optically field-confining metal synergistically as a low-resistive contact (figure 7(e)). Simply put, plasmonics enables to obtain a capacitor with seamless spatial overlap relative to the device region, thus minimizing series- and contact resistances. While photonic modulator designs do have the flexibility of spatial selective doping, such as increasing the doping level near the contacts, the accompanying effect of the higher carrier concentration introduces parasitic optical losses due to increased absorption. Indeed, the main difference between photonic and plasmonic modulators is that in plasmonics losses are high but tolerable due to the increased LMI allowing $\lambda$-size short devices, but in photonic losses are to be avoided at all costs, since the required device lengths are 100–1000 times $\lambda$.

Focusing on graphene-based modulators next, experimentally obtained power transmission changes per voltage-length, $\Delta P_{\text{opt}} = 0.013$ dB V$^{-1}$ $\mu$m$^{-1}$ \cite{53}, which is then improved by about $2\times$ in a double layer graphene push–pull configuration resulting in 0.025 dB V$^{-1}$ $\mu$m$^{-1}$ \cite{54}. Here, we present the first results of a plasmonic graphene-based modulator in a hybrid plasmon-silicon integration configuration improving this performance by another $2\times$ to 0.050 dB V$^{-1}$ $\mu$m$^{-1}$ (see next section for device details). This novel device utilizes the plasmonic LMI benefits, resulting in device length shrinkage of about $5\times$ (from 40 $\mu$m down to 8 $\mu$m) when comparing with photonic designs \cite{53, 54}. We note that the focus of this work is to explore devices that are optimized for power consumption and not for speed. However, when optimized for speed, this power metric drops to 0.003 dB V$^{-1}$ $\mu$m$^{-1}$, but enables 35 GHz fast modulation \cite{60}. Similarly, the contact resistances for experimental photonic devices is about 1000 $\Omega$, which is one order of magnitude lower for plasmonic modulators; 50–200 $\Omega$ for the graphene contact and channel (depending how close this contact is to the device), and almost no resistance at the plasmonic contact $\sim$10s $\Omega$ (figure 7(e)). The latter indeed provides a unique opportunity to design highly-energy efficient devices. A goal should be to design and demonstrate $>3$ dB V$^{-1}$ $\mu$m$^{-1}$ devices, which would enable sub-$\lambda$ long and sub-1V efficient devices if a minimum of ER $> |-3|$ dB are required. This, however, necessitates an improvement of over $10\times$ compared to our latest plasmon-photonic hybrid modulator. A possible roadmap towards this is to consider the optical polarization to match the in-plane component of 2D materials, thus increasing the optical overlap factor as discussed in the next section.

**Experimental plasmon graphene modulators on silicon**

The schematic of this first graphene-based hybrid-photonic-plasmon EAM on a silicon photonics platform uses the discussed gating scheme in figure 7. Based on the above considerations, we here show experimental results of a hybrid-plasmon graphene-based EAM operating in the telecom C-band. The device comprises of a single graphene layer sandwiched inside a silicon-based HPP mode (figures 8(a) and (b)). This design allows for strong field confinement and decent modal overlaps ($\Gamma \sim 5 \times 10^{-4}$), when surface roughness of the atomic layer deposition (ALD) oxide and metal gate deposition are considered (figure 8(c)). The latter is important since plasmonics dictates that the electrical field lines are always perpendicular to the metal and hence to the graphene, which would result in a vanishing in-plane ($E_z$)
field component in graphene, thus zeroing out the modal overlap. However, the natural process roughness (~10 nm, mainly from the poly-crystalline grain boundaries of the metal deposition with respect to a ‘flat’ ALD process) actually helps in this work providing a small amount of in-plane graphene fields (figure 8(c)). Tuning the Fermi level of graphene sandwiched inside this electrical MOS capacitor we achieve an ER of 0.4 dB μm⁻¹ resulting in an ER efficiency per unit device-length of 0.05 dB V⁻¹ μm⁻¹ (figure 8(d)). This is enabled by a multitude of device improvements; (i) the index change of the active material is high (unity), (ii) the group index is relatively large (~10), (iii) the overlap factor, which is not inherently high, is improved by the intrinsic roughness at the graphene-metal interface. Considering other device performance-related factors there are other fundamental benefits of this design including the contact resistance (R_c = 210 Ω) can be fundamentally lower compared to any photonic (non-plasmonic) mode and cavity structures where any placement of the metallic contact close to the optical mode will introduce intolerable losses. This is different for our plasmonic mode, which is inherently lossy, but the polaritonic (matter-like) mode allows to scale-down the device into a few micrometer small device (a reduction of a factor of 100) compared to traveling-wave silicon-based modulators. We refer to this design as an ‘in-the-device-biasing’, as opposed to biasing the device a few to tens of micrometers away from the active region. As such, the overall design allows for a more compact overall footprint reducing electronic capacitive overheads. Lastly, reducing the dielectric thickness (t_ox = 5 ± 1 nm), improves the electrostatics enabling a sub-1 V modulation performance. In determining the power consumption of the device, here a designed circuit has options with respect to bias conditions; for instance, (i) a bias voltage of 0.75 V enables 0.5 dB μm⁻¹
of $ER$, while (ii) a bias of 0.1 V just 0.2 dB $\mu$m$^{-1}$. Assuming small signal modulation requiring a minimum ER of $-3$ dB a device length required are 6 $\mu$m and 15 $\mu$m, respectively. This results thence in an $E/\text{bit} = \frac{1}{2} CV^2$ of 2.6 $\Omega$/bit and 110 aJ/bit, respectively. Here the device areas are just 3.6 and 8.8 $\mu$m$^2$, taking the waveguide width of 600 nm as a lateral capacitor dimension. The latter, however, could be reduced by another factor of $\sim 3$ approaching the silicon waveguide cutoff, entering the tens of aJ/bit regimes. It is interesting to ask what the fundamental lower limit for modulator energies are given a desired BER and operating temperature, which is however not part of this work. Suffice to say, for any charge-driven devices such as the graphene-based devices considered here, the ultimate limit is set by broadening, $\gamma$, of the Fermi-level or any other relevant absorption states (depending on the modulation mechanism). Thus, the minimum required drive voltage is therefore expected to be Boltzmann approximation smeared' on the order of a few $k_B T$.

An improvement from the HPP-based graphene EAM discussed in figure 8 is a device that can ensure that the optical field density is in-plane with the lateral dimension of the 2D material. This can be achieved, for instance, when the 2D material is combined with slot-waveguides; here graphene could be either above the slot [61] or below as discussed in this work. Results show a high modulation performance of 1.2 dB $\mu$m$^{-1}$ for thin and narrow metal slots, given an ER metric of 2 dB $V^{-1} \mu$m$^{-1}$ for 0.5 V of bias change (figure 5(c)) [49]. This is indeed close to the set grand
challenge of $3 \text{ dB V}^{-1} \text{um}^{-1}$. To provide an outlook, further device improvements should consider dual gating in a push–pull configuration similar to [54], but with two pairs of graphene layers above and below the slot (figure 9). The latter could result in about $4 \text{ dB um}^{-1}$ of switching, enabling a device just 770 nm long, or about $\frac{1}{2}\lambda$. While not as compact as the atomistic switch from [62], where just a few atoms (possibly even one single atom) control the modulation by cutting off a surface plasmon gap mode, a graphene modulator is expected to switch faster than metal migration-based switching mechanisms. An estimated energy consumption for the 4-layer graphene modulator assuming the same lateral and gate-oxide dimension as for our actual device from figure 8, gives a miniscule power consumption of 170 aJ/bit for 0.75 V of bias, respectively (taking $\text{ER} = 3 \text{ dB}$, figure 9). We note that the latter is about as efficient as a single FET without contacts (figure 10).

### Energy-per-bit function modulator roadmap

Coming back to drawing a parallel between the energy-per-bit function of FETs and modulators, one can show a declining trend with time (figure 10) [63]. Based on the IBM device scaling, physical dimensions of the FET and other parameters such as doping level and footprint scale as a function of a universal scaling factor [64]. As the critical dimension of the FET scales down, its energy consumption also declines simply from reduced capacitance mainly driven by improved electrostatics (blue dots, figure 10). Now, nearing the end of Moore’s law, sub 10 nm feature sizes allow energy functions $<10^4 k_B T$ corresponding to tens of aJ/bits for the FET gates alone. However, adding the wires to control this MOS capacitor adds about 2-orders of magnitude to the power consumption. The key question for electronics is therefore how long does the connection to the device need to be, and the answer lies in both circuit and interconnect designs. Still, the charging and discharging of wires is a fundamental challenge in electronics [4], impacting emerging technologies such as neuromorphic memristor devices and crossbar architectures [65]. It is therefore interesting to ask how the energy consumption at the device level compares between electronics and photonics; where the promise in the latter is based on the ‘wires’ in optics being only limited by light propagation delay and power consumption of the laser determined by the detector’s sensitivity and desired BER or SNR. Thus, if an equivalent energy consumption per switch is possible, photonics would see an performance improvement over...
of 1000 as from thermal and other effects. Thus, a technological gap versus the device and wire in electronics allows for shorter RC delays, but also for lower dissipative energy consumption. In addition, WDM offers data processing parallelism over electronics, thus resulting in an improvement factor of about 250×. This is just a device-level analysis, and a proper benchmark should include the link, and circuit level. The minimal fundamental switching energy is given by electronic links of up to about 250× using WDM and medium fast EOM drivers (tens of GHz) [66]. Mapping the recent advances of modulators into this ‘other Moore’s law’ (figure 10), shows that consistent improvements in active material selection, photonic-plasmonic hybridization (mode and plasmon/photon devices for active/passive light control), allow approaching energy-per-bit functions comparable to FETs (i.e. tens of aJ/bits). Yet, 1 k_BT is the ultimate limit and be about 5 zJ/bit given T = 300 K. Energy levels below 1 k_BT are fundamentally unattainable given broadening such as from thermal and other effects. Thus, a technological gap of 1000× (between 1 and 1000 k_BT) exists, where bridging this gap is indeed an interesting scope for future research.

Conclusions

In conclusion, we have discussed scaling vectors for attojoule efficient EOMs. We discussed LMIs required to approach this goal leading to a required optical concentration factor proportional to the Purcell factor. Selecting three material-modulation mechanisms utilizing charge, namely silicon, ITO, and graphene, we showed how optical mode designs impact obtainable effective index changes as a function of the optical overlap factor, the obtainable material index change, and the mode’s effective group index. Results show that the weak plasma dispersion of silicon limits achievable extinction ratio per nominal device length, whereas the strong index modulation of (and above) unity of TCO and 2D materials such as ITO or graphene overcompensates the low optical overlap factor (10^-5–10^-4). Cooling a graphene-based modulator enables lowering the energy-per-bit function by about 10× via improving the Fermi–Dirac function-based switching steepness of graphene based on Pauli blocking. Furthermore, we showed that improving the capacitive electrostatics of any modulator improves the achievable extinction ratio per applied voltage, a value similar to the sub-threshold swing in transistors. We experimentally demonstrate a reduction of 10× in the switching steepness in a hybrid-photon-plasmon mode graphene-based EAM on silicon with a modulation efficiency of 0.05 dB V^{-1} μm^{-1} requiring sub-1V of drive voltages. The plasmonic metal serves here synergistically as a gate to drive the capacitor, by lowering the series resistance (∼200 Ω) leading up to the device—a fundamental advantage.

Figure 10. The other Moore’s law: energy/bit versus time for FETs and modulators. FETs are able to switch at the aJ/bit at the device level, but require high parasitic electrical connectors, which increase the energy-per-bit function by about 100× [62]. Optoelectronics is limited to 10–100× J/bit efficiencies, dramatically lagging in performance from electronic transistors, which illustratively highlights the weak light-matter-interaction. Applying the design physics criteria discussed in this work shows that tens of aJ/bit efficient devices are possible, when combining highly index-changing emerging materials with polaritonic modes while optimizing polarization, mode overlap, effective index, contact resistance, and capacitance. Our ITO and graphene-based plasmon silicon hybrid integrated devices already perform at the hundreds of aJ/bit level. Going beyond this level, one can introduce low Q-cavities and multi-gating schemes as discussed in figure 9, or explore other stronger (steeper) switching materials such as quantum dots in conjunction with plasmonics or polaritonic modes. Having almost reached the long-standing goal of merging lengths scales of photonics and electronics, one can estimate the fundamental performance benefits of optics over electronics in data communication; the faster device speed enabled by only driving the sub-micrometer small photonic device capacitor (versus the device and wire in electronics) allows for shorter RC delays, but also for lower dissipative energy consumption. In addition, WDM offers data processing parallelism over electronics, thus resulting in an improvement factor of about 250×. This is just a device-level analysis, and a proper benchmark should include the link, and circuit level. The minimal fundamental switching energy is given by broadening effects such as temperature ‘smearing’ of the Fermi level of about a few k_BT. Brown arrow links FET data to the corresponding y-axis (critical device dimension, i.e. gate length). Temperature T = 300 K.
of plasmonics over photonic-based devices. The latter performance can be further improved by ensuring that the polarization of the optical or plasmonic mode is parallel in-plane with the 2D material, such as by a proposed metal slot-based graphene modulator. Deploying two pairs of graphene, one above and one below the slot, we showed a design for a 4 dB μm⁻¹ efficient modulator, resulting in a 770 nm short modulator device length for ∼3 dB of signal modulation. We explored two drive voltage options of 0.75 V or 0.1 V, where the latter increases the device length to achieve the required 3 dB modulation depth. The experimental hybrid photon plasmon modulator requires just 2.6 fJ/10 GHz given circuit the RC-delay still supports the desired driver speeds. The expected benefits for treading-in voltage for length, provided depending on which bias voltage was used, which shows the expected benefits for treading-in voltage for length, provided.

Acknowledgments

VS is supported by ARO (W911NF-16-2-0194), by AFOSR (FA9550-17-1-0377). HD is supported by AFOSR (FA9550-17-P-0014) of the small business innovation research (SBIR) program.

ORCID iDs

Volker J Sorger https://orcid.org/0000-0002-5152-4766

References

[19] Amin R et al 2017 Active material, optical mode and cavity impact on nanoscale electro-optic modulation performance Nanophotonics 0 1–18
[27] Ma R-M et al 2012 Plasmon lasers: coherent light source at molecular scales Laser Photonics Rev. 7 1–21


[38] Liu K et al 2016 Fundamental scaling laws in nanophotonics Sci. Rep. 6 37419


[40] Liu K and Sorger V J 2015 Enhanced interaction strength for a square plasmon resonator embedded in a photonic crystal nanobeam cavity J. Nanophotonics 9 093790


[45] Haffner C et al 2015 All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale Nat. Photon. 9 525–8


[61] Ding Y et al 2017 Efficient graphene based electro-optical modulator enabled by interfacing plasmonic slot and silicon waveguides Nanoscale 9 15576–81


[63] Class notes from Eli Yablonovitch (2009)

