

# Transparent conducting oxide electro-optic modulators on silicon platforms: A comprehensive study based on the drift-diffusion semiconductor model

Georgios Sinatka,<sup>1,a)</sup> Alexandros Pitolakis,<sup>1</sup> Dimitrios C. Zografopoulos,<sup>2</sup> Romeo Beccherelli,<sup>2</sup> and Emmanouil E. Kriezis<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki GR-54124, Greece

<sup>2</sup>Consiglio Nazionale delle Ricerche, Istituto per la Microelettronica e Microsistemi, Via del fosso del cavaliere 100, 00133 Rome, Italy

(Received 14 September 2016; accepted 28 December 2016; published online 12 January 2017)

Electro-optic waveguide modulators exploiting the carrier-induced epsilon-near-zero effect in transparent conducting oxides are comprehensively studied and evaluated using a rigorous multi-physics modeling framework. The examined amplitude modulators integrate indium tin oxide with two representative examples of the silicon-on-insulator technology, the silicon-rib and silicon-slot platform, with the latter design exhibiting superior performance, featuring  $\mu\text{m}$  modulation lengths, switching speeds exceeding 100 GHz, and a sub-pJ per bit of energy consumption. The effect of free carriers is rigorously introduced by combining the drift-diffusion model for the description of the carrier dynamics with near-infrared carrier-dependent permittivity models, leading to a seamless and physically consistent integration of solid-state physics and Maxwell wave theory on a unified finite-element platform. *Published by AIP Publishing.*

[<http://dx.doi.org/10.1063/1.4973896>]

## I. INTRODUCTION

The process of optical modulation consists in impressing an information signal on the amplitude, phase, or polarization of an optical carrier. With the direct current modulation of a light source posing bandwidth limitations, external modulation schemes that exploit thermo-optic, electro-absorption as well as electro-optic (EO) effects have been proposed.<sup>1</sup>

In this work, the *carrier concentration change effect* is utilized for the on-off keying modulation of the optical carrier. Alternatively referred to as the *plasma dispersion effect*, this mechanism induces changes in the free-carrier concentration of the constituent semiconductor materials, modulating both the real and imaginary part of their refractive index and, in turn, the phase and the absorption of the optical wave, respectively, paving the way for carrier-controlled photonic components.<sup>2</sup> With switching times in the order of ps, the examined effect well-exceeds the  $\mu\text{s}$  thermo-optic performance.<sup>3</sup> Additionally, it does not demand for complex layered heterostructures to manifest itself as the quantum-confined Stark electro-absorption effect,<sup>4</sup> suggested as an alternative mechanism for realizing compact and energy efficient modulators.<sup>5</sup>

Classified in the electro-optic (EO) mechanisms, carrier concentration changes are electrically induced either by injecting carriers in a forward-biased *pin* diode or through the field-effect formation of carrier-accumulation and -depletion regions in metal-oxide-semiconductor (MOS) structures and reverse-biased *pn* junctions. With the former option imposing speed limitations due to the slow generation/recombination processes originating from the generally long minority carrier lifetimes, the field-effect option is

preferable, depending solely on the inherently faster majority carrier dynamics, with the modulation speed limited instead by the device resistance-capacitance (RC) time constant.

Optical modulators based on carrier-induced changes in silicon have already been reported and reviewed in Ref. 6, realized either through Mach-Zehnder interferometer or resonance configurations. However, due to the weak impact of carrier concentration changes on the silicon refractive index, the proposed modulators suffer from large footprints and increased energy requirements, thus raising the necessity for alternative material platforms, ideally compatible with the well-established silicon technology as well. The investigation of novel materials with enhanced linear as well as non-linear optical properties indicates the ever-rising demand for highly functional nanophotonic components, exploitable in modulation<sup>7,8</sup> and switching applications.<sup>9–12</sup>

Complying with the standard CMOS fabrication techniques, transparent conducting oxides<sup>13</sup> (TCOs) rise as promising materials for dynamic photonic components, exhibiting optical properties that strongly depend on the concentration of free carriers. TCOs are wide-bandgap, degenerately doped (usually n-type) oxide semiconductors. The most well-known TCO is indium tin oxide (ITO), with other representative examples being indium zinc oxide (IZO) and aluminium zinc oxide (AZO). Exhibiting a direct bandgap greater than 3 eV, TCOs are transparent in the visible and near-infrared (NIR) regime, while their ability to accommodate a high concentration of free carriers, typically between  $10^{19}$ – $10^{21}$   $\text{cm}^{-3}$ , with mobility values around  $50 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  renders them highly conductive. These properties made TCOs ideal for realizing transparent electrodes<sup>14</sup> in optoelectronic applications, earning a leading role in the market of consumer electronics (e.g., flat panel displays) and photovoltaics.

<sup>a)</sup>gsinatka@auth.gr

Furthermore, beyond the conventional applications of TCOs, the dependency of their electrical and optical properties on variations of the free-carrier concentration renders them tunable and, thus, exploitable for dynamically controlled guided wave applications as well. In Fig. 1(a), the Drude-like infrared permittivity of ITO is illustrated as a function of its free-carrier concentration at the wavelength of  $1.55\ \mu\text{m}$ . Plotting the real and imaginary part of its permittivity, it is evidenced that ITO behaves as a lossy dielectric material for low concentration values, gradually obtaining metal-like properties for increasing concentrations, after crossing a region of near-zero permittivity. The imaginary part of ITO permittivity is monotonically increasing for increasing concentration values, indicating a rise in the intrinsic material losses.

The potential for largely negative real parts of the TCO permittivity attracted the attention on these semiconductors as alternative plasmonic materials.<sup>15</sup> In this work, though, the interest is focused on their epsilon-near-zero (ENZ) permittivity region, which is a distinctive feature of the TCO family in the NIR regime, in contrast to conventional doped semiconductors and metals, whose ENZ regions lie in the far-infrared and ultraviolet regime, respectively. This unique property allowed the design of ENZ nanophotonic components<sup>16</sup> and, especially, the realization of optical modulators<sup>17</sup> based on the *ENZ effect*.

The ENZ effect results in increased propagation loss by introducing lossy regions of near-zero permittivity in heterostructure waveguides. Specifically, by integrating a TCO layer in a photonic waveguide, the mode losses can be modulated by toggling the TCO carrier concentration between a pair of values that result in low losses (ON-state), when the TCO behaves as a low-loss dielectric, and a high-loss state (OFF state), observed when the TCO is driven into its ENZ region. The ENZ effect manifests itself through increased Joule dissipation, resulting from the enhancement of the electric field in the interior of the lossy ENZ layer, directly implied by the continuity of the normal component of the dielectric displacement. This indicates that the ENZ effect is a polarization-dependent effect that requires an electric field normally polarized to the ENZ layer in order to manifest itself. The induced propagation loss scales with the  $\text{Im}\{\tilde{\epsilon}_{\text{TCO}}\}/|\tilde{\epsilon}_{\text{TCO}}|^2$  ratio,<sup>18,19</sup> with  $\tilde{\epsilon}_{\text{TCO}}$  being the infrared TCO complex permittivity.

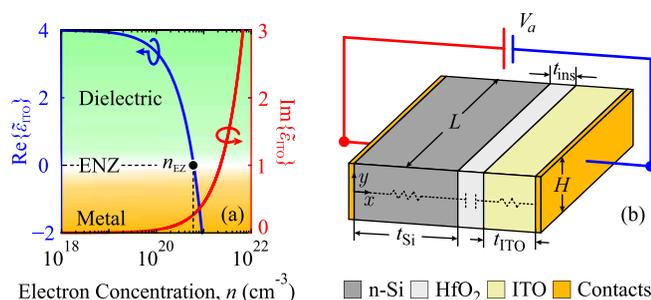


FIG. 1. (a) ITO real and imaginary permittivity parts as functions of the electron concentration at the wavelength of  $1.55\ \mu\text{m}$ . The epsilon-zero carrier concentration  $n_{\text{EZ}}$  is equal to  $6.17 \times 10^{20}\ \text{cm}^{-3}$ . (b) One-dimensional model of the n-Si/HfO<sub>2</sub>/ITO junction, employed to control the ITO free-carrier concentration, along with the equivalent RC electrical circuit.

Referring to the ITO permittivity illustrated in Fig. 1(a), the electron concentration that maximizes the propagation loss is calculated approximately equal to  $n_{\text{EZ}} \approx 6.17 \times 10^{20}\ \text{cm}^{-3}$ .

An abundance of TCO-based modulator designs has been proposed in the literature, collectively reviewed in Ref. 20, employing physical platforms from both silicon photonics<sup>19,21,22</sup> and (hybrid-)plasmonic technologies.<sup>23–33</sup> However, in the majority of the presented works, the carrier concentration change effect is inadequately introduced using averaging techniques, which fail to fully interpret the underlying physics, also occasionally leading to very optimistic results due to an overestimate of the ENZ effect magnitude. Furthermore, current efforts to rigorously describe the carrier dynamics are limited to the use of simplified models, such as the Thomas-Fermi (TF) approximation,<sup>21,26–30</sup> or commercial software simulations<sup>31,32</sup> that provide little insight into the mathematical modeling of the carrier behavior.

In this work, a seamless and physically consistent integration of solid-state physics and Maxwell wave theory is achieved on a unified finite-element platform (COMSOL Multiphysics®). The proposed model is utilized for the design and evaluation of ITO-comprising EO modulators, developed on the silicon-on-insulator (SOI) platform and employing the semiconductor-insulator-semiconductor (SIS) junction illustrated in Fig. 1(b) to control the free-carrier concentration. Opting for a high-k dielectric such as hafnium dioxide (HfO<sub>2</sub>) is proven essential for the field-effect control of free carriers. Beginning with a thorough review of the fundamental silicon-rib (Si-rib) structure, a novel modulator design is subsequently presented, relying on the silicon-slot (Si-slot) waveguide as the underlying platform. The proposed modulator outperforms the Si-rib design in all metrics and is proven comparable to the performance of plasmonic implementations.

Having reviewed the state of the art in TCO-based modulators and highlighted the contribution of this work, the paper continues in Sec. II with presenting the examined SOI platforms and the resulting modulation configurations. Section III contains the mathematical modeling of the material solid-state physics, reviews the salient points of the field-effect carrier-control mechanism, and discusses issues on the speed and the energy requirements of the switching mechanism. Section IV contains the carrier-dependent material models employed to map carrier concentration changes to variations in the infrared material permittivity, together with a consistent electromagnetic analysis of the reviewed Si-rib and the proposed Si-slot configuration. A summary and the main conclusions of this paper are given in Sec. V.

## II. PHYSICAL PLATFORMS

The proposed EO modulators are illustrated in Fig. 2. In Fig. 2(a), the Si-rib modulator consists of a donor doped Si-rib waveguide, conformally coated by a thin HfO<sub>2</sub> dielectric layer, followed by an ITO cladding. These successively deposited material layers form the necessary capacitor-like configuration presented in Fig. 1(b) for controlling the ITO free-carrier concentration. In a similar manner, the Si-slot modulator is depicted in Fig. 2(b) with ITO filling the slot

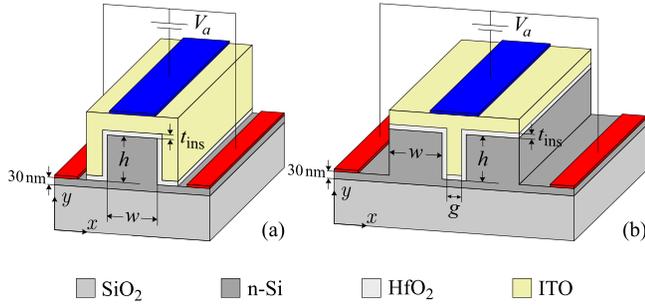


FIG. 2. SOI-based EO modulators based on the (a) Si-rib and (b) Si-slot waveguide. The biasing circuitry is schematically presented as well, along with the pertinent geometric parameters. The silicon slab thickness is fixed to 30 nm.

region. In both designs, the bias is applied between the n-Si and ITO layers (red and blue contacts in Fig. 2) with the rib structure allowing easy access to the doped Si.

The modes supported by the waveguides introduced are presented in Fig. 3, considering that the ITO is uniformly doped at  $10^{19} \text{ cm}^{-3}$ . Both TE and TM modes are illustrated for the Si-rib waveguide, while the TE slot mode of interest supported by the Si-slot waveguide is depicted in Fig. 3(c). According to Fig. 1(a), the selected ITO doping results in a low-loss dielectric permittivity value equal to  $\bar{\epsilon}_{\text{ITO}} = 3.9204 - j0.0044$ , with the refractive indices of the remaining materials listed in Table I. Such a lightly doped ITO introduces low losses to the SOI platform, preserving its comparative advantage over the intrinsically lossy plasmonic implementations.<sup>23–33</sup> Indeed, the propagation loss for the guided modes in Fig. 3 is estimated around  $0.02 \text{ dB}/\mu\text{m}$ , yielding a negligible insertion loss (IL) for devices of  $\mu\text{m}$  length. The initial free-carrier concentration in TCOs can be tailored during the fabrication process by means of dopant concentration, temperature, or oxygen pressure.<sup>34</sup> Opting for suitable deposition techniques that offer the opportunity to carefully control the deposition conditions, high-quality TCO films can be prepared with desired properties.<sup>35</sup> In general, though, deviations from the exact material specifications should be anticipated, compensable by properly adjusting the applied bias in the fabricated devices.

As ITO is gradually shifted from its initial dielectric to the ENZ regime by inducing changes in its free-carrier concentration, the polarization-dependent ENZ effect manifests itself through the dominant transverse  $E_x$  and  $E_y$  components of the TE and TM modes, respectively. Intuitively, a TE-operating Si-rib modulator is bound to

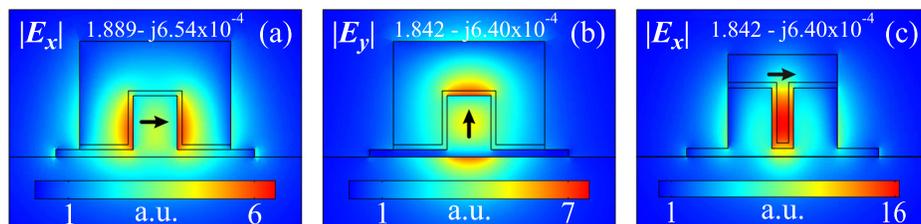


FIG. 3. Field distributions and effective indices ( $n_{\text{eff}}$ ) for the guided modes of the considered EO modulators at  $1.55 \mu\text{m}$ . (a) TE and (b) TM mode for the Si-rib waveguide with a 200 nm thick ITO cladding. (c) TE slot mode for the Si-slot waveguide topped by a 100 nm thick ITO layer. The geometric parameters for both structures are  $w = 180 \text{ nm}$ ,  $h = 220 \text{ nm}$ , and  $t_{\text{ins}} = 20 \text{ nm}$  as denoted in Fig. 2. The slot width  $g$  of the Si-slot waveguide is 40 nm.

TABLE I. Material refractive indices at  $1.55 \mu\text{m}$ . The Si donor concentration is considered equal to  $10^{18} \text{ cm}^{-3}$ .

Material	Refractive index
Silicon dioxide ( $\text{SiO}_2$ )	1.45
Hafnium dioxide <sup>36</sup> ( $\text{HfO}_2$ )	2.07
Silicon <sup>37,38</sup> (Si)	$3.48 - j1.11 \times 10^{-4}$

outperform its TM counterpart due to the double-sided interaction of the TE mode with the ITO cladding, Fig. 3(a). Moreover, superior performance is anticipated for the proposed Si-slot configuration due to the increased overlap between ITO and the tightly confined slot mode, justifying the investigation of this platform in terms of its performance. In the following Sections III and IV, a consistent modeling framework is established for the rigorous evaluation of the presented modulators.

### III. SOLID-STATE ANALYSIS

The description of the free-carrier dynamics in the proposed EO modulators can be reduced to the study of the primitive SIS junction presented in Fig. 1(b). In order to describe the carrier behavior under a  $V_a$  bias, the drift-diffusion (DD) model,<sup>39,40</sup> a well-established classical semiconductor model, is utilized. Despite its complexity, the proposed framework gains in generality and physical rigor compared to the TF approximation, which is more frequently encountered in the literature; a thorough discussion on the comparison of the two models can be found in the accompanying Appendix.

#### A. Mathematical framework

The DD model is a set of partial differential equations, consisting of the conservation laws for charge

$$-\nabla \cdot (\epsilon_0 \bar{\epsilon}_r \nabla \phi) = q(p - n + N_D^+ - N_A^-), \quad (1)$$

and for current

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{J}_n - U_n, \quad (2a)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \mathbf{J}_p - U_p, \quad (2b)$$

with the unknown variables being the electrostatic potential ( $\phi$ ), and the electron and hole concentrations ( $n$  and  $p$ ,

respectively). In Eq. (1),  $\bar{\epsilon}_r$  denotes the *static* material permittivity (distinguished from the NIR permittivity using the bar instead of the tilde symbol),  $q$  the elementary charge, and  $N_D^+/N_A^-$  the positively/negatively charged donor/acceptor sites, which are considered uniformly distributed in the material. In Eq. (2),  $\mathbf{J}_n, \mathbf{J}_p$  are the electron and hole currents, while  $U_{n,p}$  express the net carrier recombination rates, resulting from all the recombination ( $R_{n,p}^i$ ) and generation ( $G_{n,p}^i$ ) carrier mechanisms,  $U_{n,p} = \sum_i R_{n,p}^i - \sum_i G_{n,p}^i$ . These mechanisms restore the equilibrium conditions in case of externally induced perturbations. At *steady-state conditions* ( $\partial/\partial t \equiv 0$ ), it will be proven that the SIS junction is always at equilibrium allowing for setting  $U_n = U_p = 0$ .

According to the classical semiconductor theory, the carrier concentrations follow in general a Fermi-Dirac energy distribution:

$$n = N_c F_{1/2} \left( \frac{E_{f_n} - E_c}{k_B T} \right), \quad (3a)$$

$$p = N_v F_{1/2} \left( \frac{E_v - E_{f_p}}{k_B T} \right), \quad (3b)$$

where  $F_{1/2}(\cdot)$  denotes the Fermi-Dirac integral of order 1/2

$$F_{1/2}(x_0) \equiv \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2} dx}{1 + e^{x-x_0}}, \quad (4)$$

with  $T$  being the lattice temperature and  $k_B$  the Boltzmann constant. The admittance of a Fermi-Dirac distribution for the carrier concentration, instead of the simplified and commonly used Maxwell-Boltzmann statistics, is mandatory to account for the degenerately doped nature of the TCO materials.

In Eq. (3),  $E_c$  and  $E_v$  stand for the conduction and valence band levels, which are characterized by the respective effective densities of states  $N_c$  and  $N_v$ , calculated by

$$N_c = 2 \left( \frac{m_{n,\text{dos}}^* k_B T}{2\pi\hbar^2} \right)^{3/2}, \quad (5a)$$

$$N_v = 2 \left( \frac{m_{p,\text{dos}}^* k_B T}{2\pi\hbar^2} \right)^{3/2}, \quad (5b)$$

with  $m_{n,\text{dos}}^*$  and  $m_{p,\text{dos}}^*$  being the effective electron and hole masses for density of states calculations, and  $\hbar$  the reduced Planck constant. The energy levels  $E_{f_n}, E_{f_p}$  are the quasi-Fermi levels for electrons and holes, and are related to the electron and hole currents  $\mathbf{J}_n, \mathbf{J}_p$  through

$$\mathbf{J}_n = n\mu_n \nabla E_{f_n}, \quad (6a)$$

$$\mathbf{J}_p = p\mu_p \nabla E_{f_p}, \quad (6b)$$

where  $\mu_n, \mu_p$  are the carrier mobilities. After solving Eq. (3) with respect to  $E_{f_n}, E_{f_p}$

$$E_{f_n} = E_c + k_B T F_{1/2}^{-1}(n/N_c), \quad (7a)$$

$$E_{f_p} = E_v - k_B T F_{1/2}^{-1}(p/N_v), \quad (7b)$$

and substituting in Eq. (6), the conventional drift-diffusion current expressions are ultimately derived

$$\mathbf{J}_n = \mathbf{J}_{n|\text{drift}} + \mathbf{J}_{n|\text{diff}} = -q\mu_n n \nabla \phi + qD_n \nabla n, \quad (8a)$$

$$\mathbf{J}_p = \mathbf{J}_{p|\text{drift}} + \mathbf{J}_{p|\text{diff}} = -q\mu_p p \nabla \phi - qD_p \nabla p, \quad (8b)$$

where  $D_n, D_p$  are the resulting diffusion coefficients. For the derivation of the current expressions, the energy levels  $E_c$  and  $E_v$  were expressed as functions of the electrostatic potential  $\phi$ . Specifically, by selecting  $\phi$  as the *vacuum* electrostatic potential of an electron at the *vacuum* energy level  $E_0$ , the electron potential energy  $-q\phi$  can be expressed as  $E_0 - E_{\text{ref}} = -q\phi$ . The vacuum energy level  $E_0$  corresponds to the minimum energy required by an electron to completely free itself from the material, while  $E_{\text{ref}}$  is generally an arbitrarily chosen potential energy reference level. Similarly, it follows that

$$E_c - E_{\text{ref}} = -q(\phi + \chi), \quad (9a)$$

$$E_v - E_{\text{ref}} = -q(\phi + \chi) - E_g, \quad (9b)$$

as illustrated in the energy band diagrams in Figs. 4(a)–4(c) for the considered n-Si/HfO<sub>2</sub>/ITO junction. In Eq. (9),

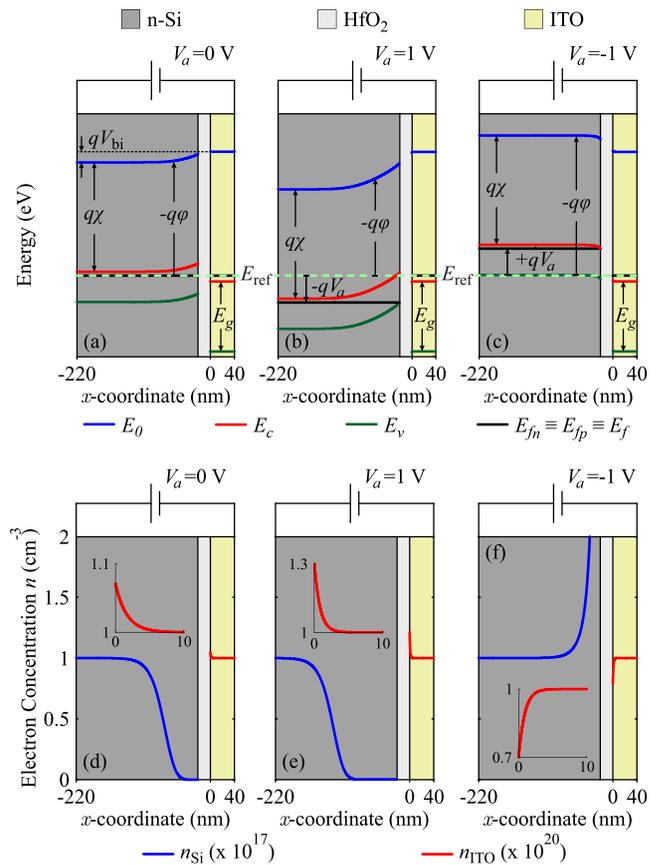


FIG. 4. (a)–(c) Energy band diagrams for the n-Si/HfO<sub>2</sub>/ITO junction under  $V_a = 0, 1$  and  $-1$  V. (d)–(f) Electron concentration corresponding to the (a)–(c) biasing conditions, normalized with respect to the dopant concentration of each material. A detailed view of the ITO electron concentration near the HfO<sub>2</sub>/ITO interface is given in the inset figures. In order to aid the energy level discrimination, the n-Si and ITO dopant concentration are selected equal to  $10^{17} \text{ cm}^{-3}$  and  $10^{20} \text{ cm}^{-3}$ , respectively, while the HfO<sub>2</sub> thickness is set equal to 20 nm.

$q\chi = E_0 - E_c$  is the electron affinity, an invariant fundamental material property, while  $E_g = E_c - E_v$  stands for the energy band gap of the semiconductor. The material parameters of the considered semiconductors are summarized in Table II.

The introduction of the potential energy reference level  $E_{\text{ref}}$  enables the calculation of relative energy differences, frequently encountered in semiconductor relations, e.g., Eq. (3), since the exact energy level values are unknown in an absolute scale. A common choice for  $E_{\text{ref}}$  is the level of the constant and coincident quasi-Fermi levels  $E_{f_n} \equiv E_{f_p}$  of the unbiased junction, illustrated in Fig. 4(a), which are also conveniently aligned along the structure as a direct consequence of the lack of external forces. The constant value of  $E_{f_n}, E_{f_p}$  stems from the absence of a net current flow,  $\mathbf{J}_n = \mathbf{J}_p = 0$  in Eq. (6), while their coincidence indicates the prevailing equilibrium conditions.

Additionally, such an  $E_{\text{ref}}$  choice conveniently allows to express the  $E_{f_n}, E_{f_p}$  levels of a  $\pm V_a$ -biased junction as  $\mp qV_a$  shifts, Figs. 4(b)–4(c), directly introducing the effect of an external voltage source. Irrespective of the applied bias, the quasi-Fermi levels are maintained constant due to the insulating effect of the intermediate  $\text{HfO}_2$  layer. This means that the steady-state SIS structure is always at equilibrium,<sup>39</sup> allowing for setting  $E_f \equiv E_{f_p} \equiv E_{f_n}$  and  $U_{n,p} = 0$  as initially claimed. At this point, the carrier concentrations  $n, p$  in Eq. (3) can be finally expressed, in conjunction with Eq. (9), as functions of the unknown variable  $\phi$ . The resulting set of equations is highly nonlinear and an iterative scheme<sup>40</sup> is employed for obtaining the solution.

Concluding the description of the mathematical framework, the effect of the insulating layer in the SIS junction is introduced through the Laplace equation, elegantly derived by setting the right side of Eq. (1) equal to zero. This approach complies with the case of an ideal SIS junction, where no charge centers or free carriers are considered in the insulating layer. The semiconductor/insulator boundaries are assumed free of surface charges, obeying the conditions of electrostatic potential  $\phi$  continuity and electrical insulation  $\hat{\mathbf{n}} \cdot \mathbf{J}_n = \hat{\mathbf{n}} \cdot \mathbf{J}_p = 0$ , with  $\hat{\mathbf{n}}$  being the outward pointing vector, normal to the semiconductor/insulator interface.

TABLE II. Solid-state parameters for the Si and ITO semiconductors. The  $N_c$  and  $N_v$  values are either explicitly reported or implicitly calculated using the respective effective masses and Eq. (5). Hole data for ITO are not reported since it is, by definition, a degenerately n-doped semiconductor.

Parameter	Si (Ref. 41)	ITO
$\bar{\epsilon}_r$	11.9	8.9 (Ref. 35)
$E_g^a$ (eV)	1.12	2.6 (Refs. 35, 42)
$\chi_0$ (eV)	4.05	4.8 (Ref. 43)
$m_{n,\text{dos}}^*/m_0$	...	Ref. 44
$m_{p,\text{dos}}^*/m_0$	...	...
$N_c$ (cm <sup>-3</sup> )	$2.86 \times 10^{19}$	...
$N_v$ (cm <sup>-3</sup> )	$2.66 \times 10^{19}$	...
$\mu_n$ (cm <sup>2</sup> /V s)	Ref. 45	Ref. 44
$\mu_p$ (cm <sup>2</sup> /V s)	Ref. 45	...

<sup>a</sup>Indirect band gap values are considered in the one-band model employed.

## B. Field-effect carrier-control mechanism

The energy band diagrams illustrated in Figs. 4(a)–4(c) offer the opportunity to gain a better insight into the field-effect mechanism that enables the control of free carriers through the formation of carrier-accumulation and -depletion regions.

As implied by Eq. (3), spatial changes in the carrier concentrations  $n, p$  should be attributed to corresponding changes in the  $E_c, E_v$  levels since the quasi-Fermi levels are maintained constant. Indeed, referring to Figs. 4(a)–4(c),  $E_c$  and  $E_v$  bend in the proximity of the insulating layer (energy band bending), indicating the presence of a perturbing electric field, as predicted by  $\nabla E_c = \nabla E_v = q\mathbf{E}$ , straightforwardly derived from Eq. (9). The field gradient determines the curvature of the energy bands. The field itself originates from electric potential differences between the semiconductor parts and is mainly accommodated in the insulating layer, penetrating into the adjacent materials in an exponentially vanishing form until reaching the flat-band (FB) region ( $|\mathbf{E}| \approx 0$ ). The terminal contacts can be considered at thermodynamic equilibrium (ideal ohmic), located sufficiently far from the perturbing electric field. Interestingly, the band bending is observed even in the absence of biasing [Fig. 4(a)] originating from the field developed by the built-in potential  $V_{\text{bi}}$ , which results from the uneven workfunctions  $W = (E_0 - E_f)_{\text{FB}}$  between the considered semiconductors.

This field-induced band bending leads to the formation of accumulation and depletion layers, shown for the electron carriers in Figs. 4(d)–4(f). In direct comparison to Figs. 4(a)–4(c), a depletion region is formed when the signed difference  $E_{f_n} - E_c$  becomes algebraically smaller than its FB value, whereas an accumulation layer is otherwise realized. The perturbation in ITO spans only for a few nanometers, whereas tenths of nanometers are necessary before n-Si enters the FB region owing to its lighter doping. The corresponding hole concentration, even though not illustrated, can be easily calculated using the  $np$  product, which is maintained constant due to the preserved equilibrium conditions.

In conclusion, a positively biased n-Si/ $\text{HfO}_2$ /ITO junction results in the formation of an accumulation layer in ITO followed by a depletion region in n-Si, Fig. 4(e), whereas the trend is reversed when  $V_a < -V_{\text{bi}}$ , as shown in Fig. 4(f) for the indicative biasing case of  $-1$  V. Consequently, ITO can be driven into the ENZ region either through a depletion or an accumulation layer, depending on whether the initial free-carrier concentration lies above or below the  $n_{\text{ENZ}}$  concentration, respectively.

## C. Bandwidth and energy requirements

As mentioned in Sec. I, the only bandwidth-limiting factor of the described mechanism is the RC time constant of the SIS junction in Fig. 1(b). Thus, the ultimate speed limit can be approximated by the well-known formula  $f_{3\text{dB}} = (2\pi RC)^{-1}$  of a low-pass RC electrical circuit. It should be borne in mind, though, that this approach is capable of producing only qualitative conclusions due to the dependence of both resistance and capacitance on the applied bias. A strict calculation of the switching speed is possible only

after a rigorous temporal analysis of the carrier behavior using the time-dependent form of Eq. (2).

Nevertheless, in a first approximation, the capacitance  $C$  of the SIS junction can be estimated by

$$C \approx C_{\text{ins}} = \frac{\bar{\epsilon}_{\text{ins}}}{t_{\text{ins}}} A, \quad (10)$$

with  $A = H \times L$  being the longitudinal contact area, lying on the  $yz$ -plane in Fig. 1(b), and  $\bar{\epsilon}_{\text{ins}}$  the static permittivity of the insulating layer. On the other hand, the resistance  $R$  can be considered as the series combination of the individual semiconductor resistances

$$R = R_{\text{Si}} + R_{\text{ITO}} = r_{\text{Si}} \frac{t_{\text{Si}}}{A} + r_{\text{ITO}} \frac{t_{\text{ITO}}}{A}, \quad (11)$$

where  $r = (q\mu_n N_D)^{-1}$  is the resistivity of each n-doped semiconductor. The contacts are considered ideally ohmic and, thus, of no contribution to the overall resistance. Due to the considerably lower ITO resistivity and provided that  $t_{\text{Si}}$  and  $t_{\text{ITO}}$  are of approximately the same order, it follows that

$$RC \approx \frac{\bar{\epsilon}_{\text{ins}}}{t_{\text{ins}}} r_{\text{Si}} t_{\text{Si}}. \quad (12)$$

Evidently, the RC time constant ends up independent of the contact area  $A$ , indicating that high bandwidth designs are realizable for low values of  $r_{\text{Si}}$  and  $\bar{\epsilon}_{\text{ins}}/t_{\text{ins}}$  ratio. Reducing the silicon thickness  $t_{\text{Si}}$  is also desirable, dictated, though, in principle by the guided-wave problem.

Low  $r_{\text{Si}}$  values can be achieved by increasing the dopant concentration in silicon, provided that the accompanying decrease in mobility does not become a limiting factor. On the other hand, the optimal value for the  $\bar{\epsilon}_{\text{ins}}/t_{\text{ins}}$  ratio should be determined after considering its effect on the switching energy consumption as well.

Specifically, the energy stored in the SIS junction can be estimated by the one-dimensional (1D) model of a parallel plate capacitor as

$$W \approx \frac{1}{2} C_{\text{ins}} V_{\text{ins}}^2 = \frac{A}{2} \bar{\epsilon}_{\text{ins}} t_{\text{ins}} E_{\text{ins}}^2, \quad (13)$$

where  $V_{\text{ins}}$  stands for the voltage drop across the insulator, and  $E_{\text{ins}}$  is the respective (constant) electric field value in the insulating layer. The condition for achieving the targeted  $n_{\text{EZ}}$  concentration in ITO can be equivalently expressed in field terms as the demand for a specific dielectric displacement value at the insulator/ITO interface. Due to the continuity of the normal component of the dielectric displacement, the corresponding electric field  $E_{\text{EZ, ins}}$  in the insulator ends up inversely related to the static permittivity of the insulating layer

$$E_{\text{EZ, ins}} \propto \frac{1}{\bar{\epsilon}_{\text{ins}}}, \quad (14)$$

leading through Eq. (13) in a similar proportional relation for the necessary energy  $W_{\text{EZ}}$

$$W_{\text{EZ}} \propto \frac{t_{\text{ins}}}{\bar{\epsilon}_{\text{ins}}} A. \quad (15)$$

Evidently, by comparing Eq. (12) with Eq. (15), the demand for high bandwidth directly conflicts with the requirement for low energy consumption, raising a performance trade-off for the switching mechanism.

Based on the above conclusions, a  $t_{\text{ins}} = 5$  nm  $\text{HfO}_2$  insulating layer of  $\bar{\epsilon}_{\text{ins}} = 25$  (Ref. 46) has been selected for the EO modulators presented in Fig. 2 as a compromise between the energy and bandwidth demands. This option is also encouraged by the increasing attention that  $\text{HfO}_2$  is attracting as a novel high-k dielectric gate material in the electronics transistor research. The Si dopant concentration is set equal to the moderately high value of  $10^{18} \text{ cm}^{-3}$ , resulting in  $r_{\text{Si}} \approx 0.03 \Omega \text{ cm}$ . The ITO dopant concentration, even though of little importance to the RC value, has a detrimental effect on the IL of the resulting modulator and, thus, is restricted to  $10^{19} \text{ cm}^{-3} < n_{\text{EZ}}$ . This means that an accumulation layer will be eventually necessary to drive ITO into the ENZ region.

In Fig. 5, the electron concentration across the engineered n-Si/ $\text{HfO}_2$ /ITO junction is illustrated under the effect of five biasing values. The maximum applied voltage is limited to 5 V, a reasonable choice for preventing the  $\text{HfO}_2$  breakdown. As shown, the carrier concentration follows a non-uniform distribution along ITO, with the targeted  $n_{\text{EZ}}$  concentration achieved for every biasing value beyond the voltage threshold of  $V_{\text{th}} \approx 3$  V, though at different spatial locations, progressively shifted away from the  $\text{HfO}_2$ /ITO interface. The ON modulation state can be attained in the unbiased case  $V_{\text{on}} = 0$ , which appears as a reasonable option for maintaining low energy consumption. Alternatively, one could opt for negative biasing values for the ON-state that would result in reduced mode losses thanks to the depletion region formed in ITO, yet inducing a rise in the energy per bit consumption due to the doubly energy consuming bit-states. On the other hand, a bias  $V_{\text{off}} > V_{\text{th}}$  is necessary for realizing the OFF modulation state, with lower values being inadequate for driving ITO into its ENZ region. The exact biasing value for the OFF-state requires a rigorous quantification of the free-carrier effect on the optical loss, presented in Sec. IV.

#### IV. ELECTROMAGNETIC ANALYSIS

In order to evaluate the effect of free carriers on the optical wave propagation, carrier-dependent material models are

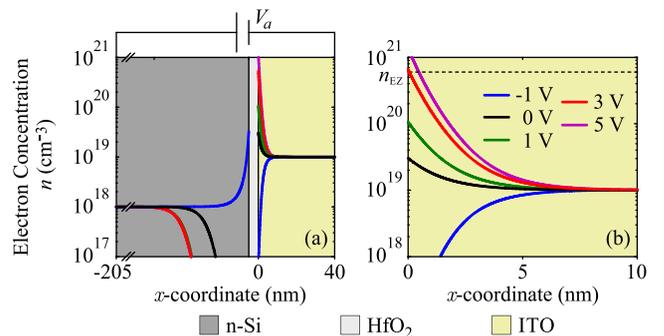


FIG. 5. (a) Spatial distribution of the electron concentration in the engineered n-Si/ $\text{HfO}_2$ /ITO junction, integrated into the EO modulators, for a set of five biasing values. (b) Detailed view of the ITO electron concentration. The  $n_{\text{EZ}}$  concentration equals  $6.17 \times 10^{20} \text{ cm}^{-3}$ .

employed to translate carrier concentration changes into infrared material permittivity variations. The resulting carrier-modulated infrared permittivity is subsequently fed into an eigenmode solver for Maxwell's equations, which translates the material permittivity changes into changes of the guided mode effective index,  $n_{\text{eff}}$ .

### A. Carrier-dependent infrared material models

The effect of free carriers on the Si permittivity is described by well-known models,<sup>37,38</sup> which are based on carrier-dependent refractive index expressions derived by a Kramers-Kronig analysis and experimentally obtained results. With the refractive index of the undoped and, thus, lossless Si set equal to 3.48, the n-Si permittivity is modulated under the effect of both electron and hole concentrations. However, its contribution to the overall modulation performance is proven negligible as it will be subsequently shown.

On the other hand, the ITO optical properties are solely determined by the behavior of its free electrons. Due to the high free-carrier concentration, a simple lossy Drude model<sup>47</sup> is adopted for the infrared permittivity, justified by the absence of interband transitions and the lack of phonon absorption in the NIR as well.<sup>35</sup> Specifically, the ITO permittivity is expressed as

$$\tilde{\epsilon}_{\text{ITO}} = \tilde{\epsilon}_{\text{opt}} \left( 1 - \frac{\omega_p^2}{\omega^2 - j\gamma\omega} \right), \quad (16)$$

where  $\tilde{\epsilon}_{\text{opt}}$  accounts for the background polarizability of the undoped semiconductor,  $\omega_p$  is the plasma frequency, and  $\gamma$  the damping rate. In this work,  $\tilde{\epsilon}_{\text{opt}}$  is set equal to 4, independent of the ITO doping,<sup>35</sup> while  $\omega_p$  and  $\gamma$  are given by

$$\omega_p = \sqrt{\frac{ne^2}{\tilde{\epsilon}_{\text{opt}}\epsilon_0 m_c^*}}, \quad (17)$$

$$\gamma = \frac{e}{\mu_n m_c^*}, \quad (18)$$

with  $n$  being the free-electron concentration,  $\mu_n$  the electron mobility, and  $m_c^*$  the effective conduction-band electron mass, coinciding with  $m_{n,\text{dos}}^*$  under the assumption of parabolic and isotropic band structure ( $m_c^* \equiv m_{n,\text{dos}}^*$ ).

Letting the typically carrier-dependent  $m_c^*$  and  $\mu_n$  values correspond to a reference ITO electron concentration,  $\tilde{\epsilon}_{\text{ITO}}$  ends up carrier-dependent solely through the plasma frequency, Eq. (17). The described model was employed for plotting the infrared permittivity of ITO in Fig. 1(a) as a function of its free carriers, considering a reference free-carrier concentration equal to  $10^{19} \text{ cm}^{-3}$ .

## B. Electro-optic modulators

### 1. Silicon rib platform

Beginning with the fundamental Si-rib platform, the effect of an applied bias  $V_a$  on the propagation losses of both TE and TM guided modes is investigated. Elementary

results, obtained by downgrading the two-dimensional (2D) waveguide cross section to equivalent one-dimensional (1D) structures, are illustrated in Figs. 6(a) and 6(b) for the TE and TM modulator, respectively. The equivalent 1D configurations result from properly selected geometry cuts, parallel each time either to the TE or TM polarization, indicated with dashed lines in the inset Si-rib cross sections.

Solving for the 1D carrier concentration, Fig. 5, and modulating accordingly the material permittivity, the sought mode is identified through an eigenmode solver as always perpendicularly polarized to ITO (transverse magnetic mode) in order for the ENZ effect to manifest itself. The respective mode loss in  $\text{dB}/\mu\text{m}$  is calculated as

$$\alpha = 10 \left| \log \left( e^{-\frac{4\pi}{\lambda} \text{Im}\{n_{\text{eff}}\}} \right) \right|, \quad (19)$$

where  $\lambda$  is the operating wavelength in  $\mu\text{m}$ , and  $\text{Im}\{n_{\text{eff}}\}$  the imaginary part of the effective refractive index.

As shown in Fig. 6, sweeping the applied bias  $V_a$  in the range 0–5 V, the initially low TE (TM) mode loss of 0.03 (0.02)  $\text{dB}/\mu\text{m}$  steeply increases around the formerly predicted threshold of  $V_{\text{th}} = 3$  V, saturating to 0.5 (0.3)  $\text{dB}/\mu\text{m}$  for a bias greater than 4 V, verifying the initially predicted inferior performance of the TM mode. The observed two-state optical switch behavior has also been described in Refs. 28 and 30 and is quite different from the non-monotonic trend reported<sup>25</sup> in the case of modeling the ENZ effect using an equivalent homogeneous accumulation layer of arbitrary thickness and averaged carrier concentration, which leads to a considerable overestimate of the ENZ effect magnitude, highly dependent on the selected thickness of the accumulation layer.<sup>30</sup>

Contrasting the field profiles between the unbiased and the 5 V-biased state [Figs. 6(c) and 6(d)/6(e) and 6(f) for TE/TM operation], a significant, almost delta-Dirac-like field enhancement is observed in the vicinity of the ITO/HfO<sub>2</sub> interface for the biased case, where the ITO permittivity crosses into the ENZ region. For biasing values beyond  $V_{\text{th}}$ , the ENZ region is formed deeper into the ITO layer, with the peak of the field being correspondingly shifted, closely

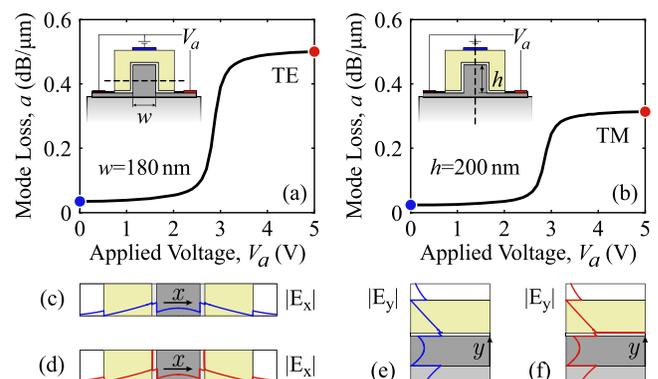


FIG. 6. One-dimensional calculations of the mode loss as a function of the applied bias  $V_a$  for (a) TE- and (b) TM-operating Si-rib modulator. The equivalent geometry cuts for each mode are indicated with a dashed line in the inset Si-rib cross sections. The 1D TE/TM field profiles at the unbiased (in blue) and the 5V-biased (in red) state are contrasted in (c) and (d)/(e) and (f), respectively.

tracking the spatially varying ENZ region. The field is so confined that the mode profile visualization becomes delicate at this scale. This highly confined field in the less than 1 nm-wide lossy ENZ region is responsible for the resulting high mode loss once the ENZ region settles in ITO. This mode is purely photonic in its essence, originating from the demand for the continuity of the normal component of the dielectric displacement across a highly varying permittivity profile. Its photonic nature is further corroborated by the fact that no plasmon modes can be supported at the ITO/HfO<sub>2</sub> interface, since the necessary condition  $\text{Re}\{\tilde{\epsilon}_{\text{ITO}}\} < -\text{Re}\{\tilde{\epsilon}_{\text{HfO}_2}\}$  is not met for  $V_a < 5$  V.

The intensity of the EO effect benefits from an enhancement in the light-matter interaction, achieved by engineering the waveguide cross section. The geometry decomposition into equivalent 1D structures allows for examining separately the effect of the dimensions lying in parallel to the TE or TM polarization from those being perpendicular, providing a computationally favorable approach for investigating the geometry effect on the modulation performance.

In Fig. 7(a), the effect of the silicon width,  $w$ , on the TE mode is examined. Reduced IL is achieved for increasing  $w$  values as a result of the increase in confinement and the resulting weaker overlap with the highly lossy ITO cladding. This, however, prevents the efficient interaction between the guided mode and ITO, resulting in poorer extinction ratio (ER) values, calculated as the ratio between the losses at  $V_a = 5$  V and those at the unbiased state. A similar behavior is observed for the TM mode for increasing  $h$  values, Fig. 7(b). The dashed lines in Fig. 7 highlight the silicon free-carrier contribution to the modulation performance, which is evidently negligible and, thus, safely ignored as previously mentioned. Interestingly, however, the n-Si losses contribute constructively to the modulation performance in the case of a weakly guided mode, offering an increase in the ER with a negligible impact on the IL at the ON-state. This positive contribution is eliminated, though, as the Si-core becomes larger and the guided mode more confined.

Both polarizations, irrespective of the Si-core dimensions, are able of achieving an ER at least an order of magnitude greater than the respective IL, as shown in Fig. 7. This implies that a 10 dB modulation is always feasible with a penalty on IL of less than 1 dB, provided that the modulator

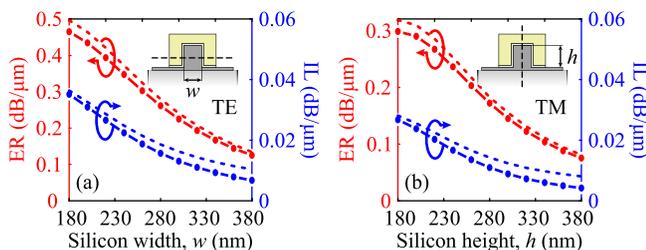


FIG. 7. One-dimensional calculations of the ER and the IL for (a) TE- and (b) TM-operating Si-rib modulator as functions of the Si width and height, respectively. The ER is calculated between the unbiased and the 5 V-biased state, while the IL refers to the ON-state of the modulator. The dashed curves incorporate additionally the effect of the Si free carriers. The studied geometry cuts and the examined parameters for each mode are denoted in the inset Si-rib cross sections.

is sufficiently long. The demand, though, for compact designs renders high ER values favorable, realizable for weakly guided modes. In this work, a Si-core of  $w = 180$  nm ( $h = 200$  nm) is selected for the TE (TM) mode, ensuring weak waveguiding conditions and providing an ER of 0.47 (0.30) dB/μm, in contrast to the strongly guided Si-wire mode reported in Ref. 21 that provided a mere 0.11 dB/μm. The optimal thickness values for the ITO cladding was found equal to 200 (300) nm for the TE (TM) mode. Thicker ITO layers have an insignificant effect on the modulation performance, while thinner films result in poor ER/μm performance due to a decrease in the confinement and, thus, in the interaction of ITO cladding with the guided mode. At this point, let it be noted that the unbiased TCO optical properties are considered thickness-independent, practically achievable by maintaining stable deposition conditions<sup>35</sup> as well as ensuring film thicknesses no less than 50 nm.<sup>13</sup>

Continuing the evaluation of the Si-rib platform, the effect of the other transverse dimension, the one perpendicular to the mode polarization, is investigated through rigorous 2D parametric studies. For the TE modulator of  $w = 180$  nm, Fig. 8(a), as  $h$  increases and the structure begins to resemble the  $y$ -invariant configuration of Fig. 6(a), the mode-loss curve converges to the calculated 1D case. Similar results are obtained for the TM modulator of  $h = 200$  nm as  $w$  increases, Fig. 8(b). The differences in the high-loss levels are attributed to the edge effects of the 2D geometry, observed for both field and carrier concentration distributions as illustrated in Figs. 8(c)–8(f) and in 9(a), respectively. These effects alter the resulting distributions from being completely uniform in the 1D case to the more elaborate 2D profiles. Thus, the 1D loss curve can serve as an estimate for the upper modulation performance limit, constituting a computationally efficient tool for a preliminary analysis of the modulation potential.

The judicious selection of the  $h$  and  $w$  values, examined in Figs. 8(a) and 8(b), should account for easy interfacing with the conventional Si-photonic waveguides, compactness, and high switching performance (determined by the bandwidth and the energy per bit requirements). Opting for a TE

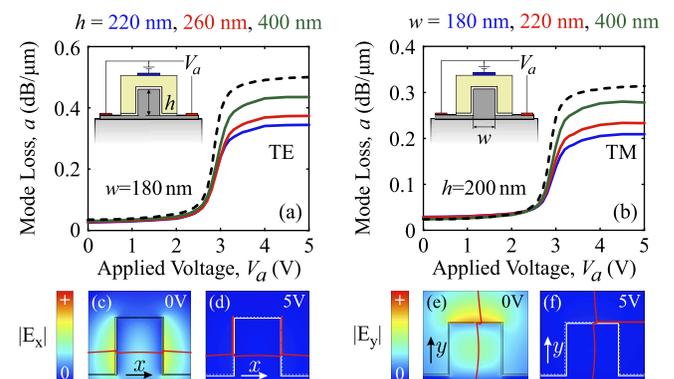


FIG. 8. Two-dimensional calculations of the mode loss as a function of  $V_a$  in (a) TE- and (b) TM-operating Si-rib modulator for a parametric set of  $h$  and  $w$  values, respectively. For the TE, TM modulator, it is  $w = 180$  nm,  $h = 200$  nm, respectively. The dashed curves correspond to the 1D mode loss presented in Fig. 6. The TE/TM field distributions at the unbiased and the 5 V-biased state are contrasted in (c) and (d)/(e) and (f), respectively.

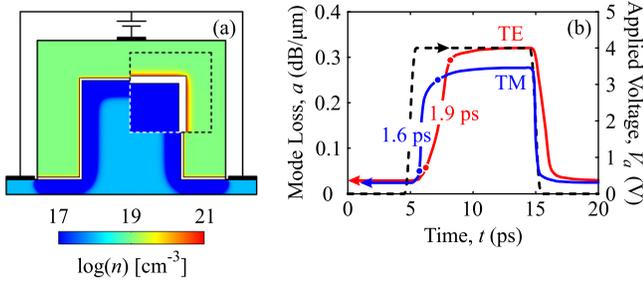


FIG. 9. (a) Logarithmic electron concentration at the cross section of the Si-rib waveguide under a 5 V bias. A narrow accumulation layer is observed in ITO whereas the respective electron depletion region in Si is more apparent. A detailed illustration of the electron concentration at the upper right corner is included as well. (b) Temporal response for the Si-rib modulator operating either in TE or TM mode under the application of a 0–4 V rectangular pulse.

modulator of  $h = 220$  nm facilitates the interface with standard Si-photonic waveguides, even though it results in poorer  $ER/\mu\text{m}$  values and, thus, less compact designs, Fig. 8(a). The increase in the device length, however, does not imply an energy consumption rise since a decrease in  $h$  reduces the capacitance per unit length, compensating the larger device footprint. This trend is reflected in the relation

$$W = \frac{L}{4} \int \int_S \mathbf{E} \cdot \mathbf{D}^* dS, \quad (20)$$

which estimates the averaged energy consumption for achieving a reference ER level through calculating the energy stored in the structure, with  $L$  being the necessary interaction length and  $S$  the waveguide cross section. Decreasing  $h$  and, thus, the total structure capacitance has an additional beneficial effect on the speed of the switching mechanism, highlighting once again the selection of lower  $h$  values as favorable.

On the contrary,  $w$  variations have a negligible impact on the bandwidth and energy per bit requirements of the switching mechanism, letting  $w = 400$  nm serve as a reasonable option for the TM modulator implementation. Apart from reducing the necessary length, as shown in Fig. 8(b), such an option enables the interfacing with a standard  $400 \times 200$  nm<sup>2</sup> TE photonic waveguide through the use of a conventional polarization rotator.

Table III summarizes the performance of the Si-rib modulator under both TE and TM operation. Demanding a 10 dB ER, the length and the energy per bit consumption of the modulators are listed, where  $V_{\text{off}}$  is restricted to a 4 V value in order to limit the energy demands while maintaining a high modulation performance. The bandwidth of the modulators is estimated using the relation  $B = 0.35/t_r$  of a low-pass RC circuit, with  $t_r$  being the 10%–90% rise time, which is calculated through a rigorous temporal analysis, Fig. 9(b),

TABLE III. Geometry options and performance metrics for a 10 dB Si-rib modulator for a bias switching between 0–4 V under both TE and TM operation.

Mode	$w \times h$ (nm)	$L$ ( $\mu\text{m}$ )	IL (dB)	pJ/bit	GHz
TE	$180 \times 220$	33	1	1.6	184
TM	$400 \times 200$	40	1	2.4	219

after employing the time-dependent form of Eq. (2). The TE modulator, despite its higher efficiency due to the mode overlap with two ITO/HfO<sub>2</sub> interfaces, has an inferior switching speed, resulting from the larger RC time constant of the double SIS junction parallel circuit.

Concluding the study of the fundamental Si-rib design, its integration potential in standard photonic circuitry is evaluated by investigating its efficient interfacing with conventional low-loss photonic waveguides. For this reason, the interfacing of the suggested  $180 \times 220$  nm<sup>2</sup> TE modulator with a standard  $400 \times 220$  nm<sup>2</sup> TE photonic waveguide is examined through the use of tapering sections, frequently employed in standard photonic circuitry. Such an implementation is depicted in Fig. 10(a), where the Si-core width of a TE-operating Si-rib waveguide is adiabatically tapered from 400 nm to 180 nm and vice versa. The lengths of the tapering and main modulation sections equal 10  $\mu\text{m}$  each, amounting to a 30  $\mu\text{m}$ -long component. The functionality of this *proof-of-concept* design was verified by means of finite-difference beam-propagation method (FD-BPM) numerical simulations in a 2D propagation analysis scheme including a horizontal cut plane, parallel to the propagation direction. The cross section of the waveguide on the  $xy$ -plane was appropriately reduced to one dimension, as shown in the inset of Fig. 10(a). The ITO permittivity profile was modulated in accordance to the applied voltage based on the carrier concentration curves presented in Fig. 5. The fundamental  $x$ -polarized mode of the waveguide was launched at the left port, and the optical field was propagated across the component until reaching the right port, where the mode loss was calculated for each applied voltage. The results presented in Fig. 10(b) are in agreement with the behavior predicted in Fig. 6(a). Specifically, the mode loss significantly increases once the applied voltage exceeds the threshold of 3 V, resulting in an ER of almost 10 dB between the low- and high-loss states, with the corresponding field distributions visualized in Figs. 10(c) and 10(d). In agreement with the calculations in Fig. 7(a), an ER of 0.45 dB/ $\mu\text{m}$  is calculated for the main

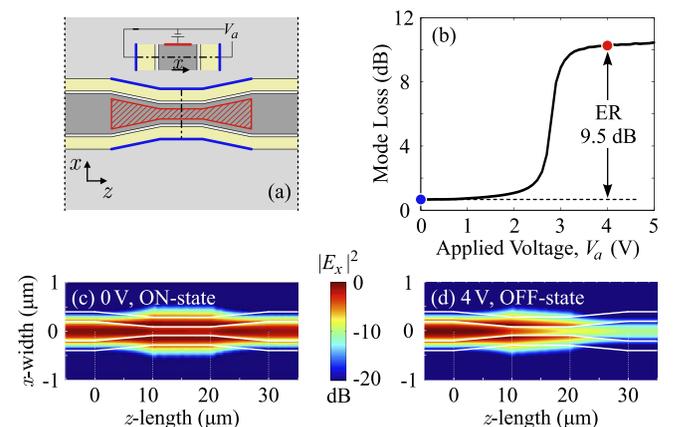


FIG. 10. FD-2D-BPM simulation of the interfacing between the proposed TE Si-rib modulator and standard silicon photonic waveguides. (a) Top view of the component. The inset depicts its cross section including the corresponding biasing circuitry. (b) Mode loss over the 30  $\mu\text{m}$ -long component as a function of the applied voltage  $V_a$ . A top view of the beam propagation through the component is provided at (c) the ON-state and (d) the OFF-state.

modulation section, with the tapers providing an average ER of  $0.25 \text{ dB}/\mu\text{m}$  each. The total IL of the component does not exceed 1 dB. The FD-BPM simulations also verified that the tapering sections do not perturb the wave propagation, ensuring efficient mode matching even in the limit case of Fig. 10(d), where the mode profile in the modulator becomes highly discontinuous.

## 2. Silicon slot platform

Despite the simplicity of the Si-rib modulator configuration, its pJ/bit consumption raises the question for more efficient modulation designs. Decreased energy requirements are in principle feasible by reducing the necessary device length, Eq. (20), through enhancing the overlap between the guided mode and ITO in order to achieve higher  $\text{ER}/\mu\text{m}$  values. In this work, the Si-slot waveguide is proposed as a novel platform for developing high-performing TCO-based modulators, being the only design capable of achieving such a high field enhancement, Fig. 2(b), while enjoying the low-loss merits of the SOI platform.

Following a similar to the Si-rib case evaluation procedure, the preliminary 1D loss calculations for a Si-slot waveguide of  $w = 180 \text{ nm}$  and  $g = 20 \text{ nm}$ , Fig. 11(a), yield an ER equal to  $1.6 \text{ dB}/\mu\text{m}$  with an IL of  $0.01 \text{ dB}/\mu\text{m}$ , constituting a significant improvement over the respective Si-rib performance. The 1D field profiles at the unbiased and the 5 V-biased state are given in Figs. 11(c) and 11(d), respectively. Higher  $w$  values loosen the interaction between the slot mode and ITO, further reducing the already sufficiently low IL, but limiting at the same time the achievable ER/ $\mu\text{m}$ , Fig. 12(a). Regarding the slot width  $g$  examined in Fig. 12(b), small values are favorable for high performance, suggesting strong waveguiding conditions for the Si-slot platform in contrast to the weakly guided mode scheme adopted for the Si-rib waveguide, with  $g = 20 \text{ nm}$  being a fabricationally tolerable option. Similar to Fig. 7, the dashed lines in Fig. 12

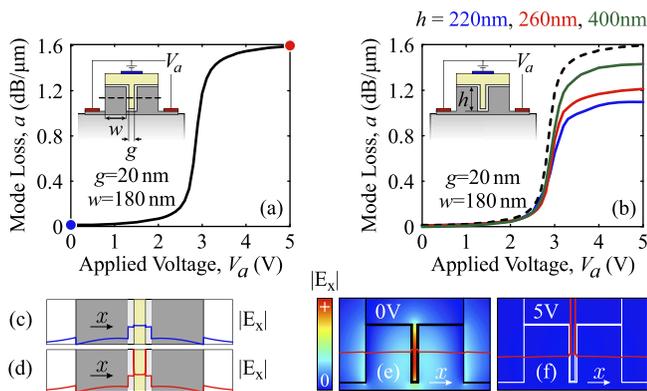


FIG. 11. Calculations of the mode loss as a function of the applied bias  $V_a$  for the Si-slot waveguide. (a) One-dimensional calculation based on the geometry cut indicated with a dashed line in the inset Si-slot cross section. (b) Two-dimensional calculations for a Si-slot waveguide of  $w = 180 \text{ nm}$  and  $g = 20 \text{ nm}$  and a parametric set of  $h$  values. The dashed curve can be identified as the 1D loss curve in (a). The 1D/2D field distributions at the unbiased and the 5 V-biased state are contrastively given in (c) and (d)/(e) and (f), respectively.

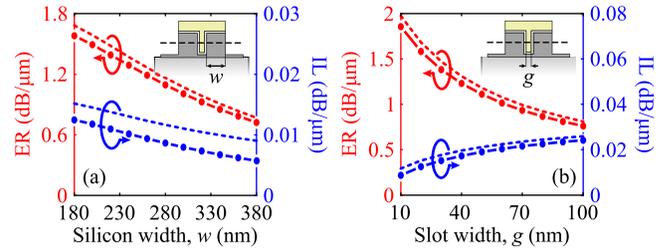


FIG. 12. One-dimensional calculations for the ER and the IL of a Si-slot modulator as functions of the (a) silicon width and (b) slot width. The ER is calculated between the unbiased and the 5 V-biased state, with the IL referring to the ON-state of the modulator. The corresponding geometry cut and the examined parameters are denoted on the inset figures. The dashed curves additionally incorporate the effect of the Si free carriers.

incorporate the Si free-carrier contribution, revealing a weak impact on the Si-slot platform as well.

The effect of silicon height  $h$  is evaluated in Fig. 11(b) through rigorous cross-sectional calculations, including waveguide eigenmode studies, illustrated in Figs. 11(e) and 11(f), following rigorous 2D carrier concentration calculations, Fig. 13(a). Similar to the TE Si-rib modulator case, increasing  $h$  values shift the high-loss levels towards the 1D loss curve, which constitutes an upper limit for the realizable mode loss. With an ER per  $\mu\text{m}$  exceeding  $1 \text{ dB}/\mu\text{m}$ , the proposed Si-slot configuration is capable of achieving  $\mu\text{m}$ -long modulator scales and a sub-pJ/bit of energy consumption. Opting for the minimum examined, yet conventionally used, value of  $h = 220 \text{ nm}$ , an ultra high bandwidth, well exceeding 100 GHz, can be additionally ensured. This is verified in Fig. 13(b), where a 2 ps rise time is calculated for the engineered Si-slot modulator under the application of a 0–4 V rectangular pulse, identically to the Si-rib calculations presented in Fig. 9(b).

Table IV clearly highlights the enhanced performance achieved by the  $\mu\text{m}$ -long Si-slot modulator, which, apart from the reduction in the switching energy, achieves an order of magnitude lower IL compared to the previously studied Si-rib configuration, maintaining additionally an ultra high bandwidth. The proposed Si-slot modulator exhibits performance metrics that meet the demands of contemporary integrated photonic circuitry, suggesting a viable solution for the on-chip modulation of the optical carrier.

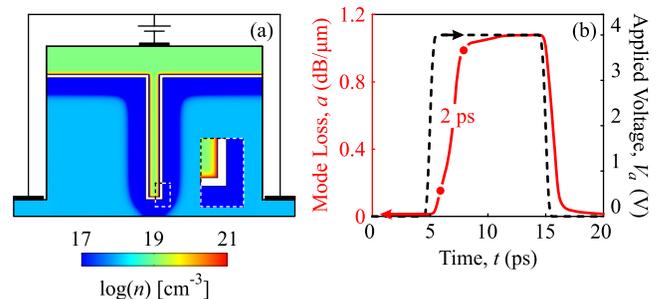


FIG. 13. (a) Logarithmic electron concentration at the cross section of a Si-slot waveguide under a 5 V bias, with a narrow accumulation layer observed in ITO and a more pronounced electron depletion region in Si. A detailed illustration of the electron concentration in the slot is included as well. (b) Temporal response for the Si-slot modulator under the application of a 0–4 V rectangular pulse.

TABLE IV. Geometry options and performance metrics for a 10 dB Si-slot modulator for a bias switching between 0–4 V.

$w \times h$ (nm)	$g$ (nm)	$L$ ( $\mu\text{m}$ )	IL (dB)	pJ/bit	GHz
$180 \times 220$	20	9	0.1	0.6	175

## V. SUMMARY AND CONCLUSIONS

Electro-optic modulators comprising ITO and developed on two of the most representative SOI platforms, the Si-rib and Si-slot waveguides, were comprehensively studied and evaluated using a rigorous multi-physics modeling framework. Employing the drift-diffusion model for the description of the free-carrier dynamics in semiconductors, the electro-optic effect was straightforwardly introduced using a unified FEM implementation, where changes in the concentration of free carriers are directly mapped to changes in the infrared material permittivity. This approach manages to consistently represent the underlying physics, seamlessly linking the solid-state material behavior with the Maxwellian wave principles, avoiding, thus, the frequently employed definition of an artificial homogeneous accumulation layer that lacks physical interpretation. After a thorough review of the fundamental Si-rib design, a novel Si-slot modulator was proposed, outperforming in all metrics, its Si-rib counterpart. Exhibiting a potential for sub-pJ energy consumption, a trivial IL, and a bandwidth well exceeding 100 GHz (obtained from rigorous temporal simulations), the performance of the  $\mu\text{m}$ -long Si-slot modulator is comparable to that of high-IL plasmonic implementations, highlighting the SOI platform as the leading candidate for TCO-based modulators, encouraged additionally by the scalability of the CMOS technology.

## ACKNOWLEDGMENTS

This work was supported by the ‘‘Research Projects for Excellence IKY/Siemens.’’

## APPENDIX: COMPARISON TO THE THOMAS-FERMI APPROXIMATION

In this Appendix, a direct comparison between the frequently employed Thomas-Fermi (TF) approximation and the drift-diffusion (DD) model is provided. Initiating from a reduced majority-carrier and steady-state form of the DD model, the conditional derivation of the electron-only and time-independent TF equations is outlined, highlighting the underlying conventions and the inherent limitations of the TF approach, as encountered in the literature (supplementary material of Ref. 28).

Beginning with the reduced form of the DD model, the conservation law for charge is simplified as

$$\nabla \cdot (\varepsilon_0 \bar{\varepsilon}_r \nabla \phi) = q(n - N_D^+), \quad (\text{A1})$$

with the electron concentration,  $n$ , given by

$$n = N_c F_{1/2} \left( \frac{E_f - E_c}{k_B T} \right), \quad (\text{A2})$$

where  $E_f$  substituted  $E_{f_n}$  in Eq. (3a), as discussed in Sec. III A.

In the case of a highly degenerately doped semiconductor (or metal), the condition  $E_f - E_c \gg k_B T$  holds, allowing for the asymptotic expansion  $F_{1/2}(x) \rightarrow 4x^{3/2}/(3\sqrt{\pi})$ ,<sup>48</sup> which reduces Eq. (A2) to

$$n = \frac{4N_c}{3\sqrt{\pi}} \left( \frac{E_f - E_c}{k_B T} \right)^{3/2}. \quad (\text{A3})$$

Solving Eq. (A3) with respect to  $E_f$  and using Eq. (5a), the Fermi level is expressed with reference to  $E_c$  as

$$E_f = \frac{(3\pi^2)^{2/3} \hbar^2 n_0^{2/3}}{2m_{n,\text{dos}}^*}, \quad (\text{A4})$$

where  $n_0$  denotes the equilibrium free-electron concentration. By additionally setting  $E_c = -q\phi$  in Eq. (A3), resulting from Eq. (9) with  $\phi$  referring to the *local potential* at the Fermi-level, the familiar expression of the TF electron concentration is derived

$$n = \frac{4N_c}{3\sqrt{\pi}} \left( \frac{E_f + q\phi}{k_B T} \right)^{3/2}. \quad (\text{A5})$$

The latter substitution,  $E_c = -q\phi$ , assumes a homogeneous structure, composed of a single semiconductor material, since no energy-band related semiconductor quantities (e.g., the electron affinity,  $\chi$ ) are considered. Thus, the n-Si/HfO<sub>2</sub>/ITO heterojunction considered in this work, Fig. 1(b), cannot be in principle modelled using Eq. (A5). More importantly, though, the relative  $E_f$  positioning with respect to  $E_c$  does not justify the use of the TF approximation in the first place, since the doping levels for n-Si ( $10^{18} \text{ cm}^{-3}$ ) and ITO ( $10^{19} \text{ cm}^{-3}$ ) result in  $E_f < E_c$  and  $E_f - E_c \approx 3k_B T/2$ , respectively, failing to meet the necessary  $E_f - E_c \gg k_B T$  condition.

Despite the aforementioned shortcomings, the TF approximation is applied to the case of the n-Si/HfO<sub>2</sub>/ITO junction, with the resulting carrier-concentration curves illustrated in Fig. 14(a), after artificially considering the n-Si layer as perfectly conductive. Comparing to the respective DD calculations, it can be evidenced that the built-in potential at the unbiased state is neglected, while the positive-bias curves converge to the DD-calculated distributions as the bias, and, thus, the radicand in Eq. (A5), increases. It should be pointed out that the TF approximation is incapable of evaluating negative biasing values as straightforwardly deduced from Eq. (A5).

Next, the effect on the electromagnetic response of the EO modulators is investigated employing the case of the 1D Si-slot modulator. As illustrated in Fig. 14(b), the bias threshold for the onset of the mode losses shifts to lower values owing to the overestimate of the field-effect magnitude, resulting from the TF-approximation inability to account for the electric potential drop across the n-Si layer. The mode loss at the ON and OFF state is in agreement with the DD calculations, with the discrepancies observed in the solid-state analysis greatly suppressed by the wave physics of the switching mechanism.

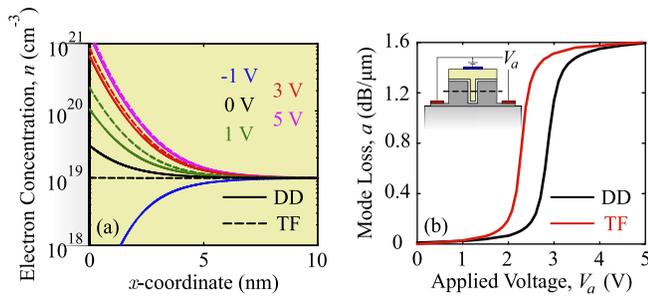


FIG. 14. (a) Bias-dependent electron concentration curves across the ITO layer of the n-Si/HfO<sub>2</sub>/ITO junction, calculated using the DD model and the TF approximation. Despite some curve pairs (for instance at 3 V) appear very close in logarithmic scale, they are, nevertheless, sufficiently differentiated when linearly scaled, translating in the non-negligible differences observed in the (b) mode loss, examined in the case of the 1D Si-slot modulator, first introduced in Fig. 11(a).

Overall, the DD framework is valid for semiconductor configurations of arbitrary energy-band diagrams and biasing conditions in contrast to the TF approximation that is only applicable to positively biased and highly degenerately doped single semiconductor structures. In addition, the DD model, contributing with the continuity equations, Eq. (2), and the expressions for current, Eq. (8), can be used to study a wide range of carrier-controlled schemes, including both carrier-injection and field-effect mechanisms of either steady-state or transient carrier dynamics. This constitutes a major comparative advantage over the exclusively field-effect and steady-state formulation of the TF approximation.

- <sup>1</sup>A. Chen and E. Murphy, *Broadband Optical Modulators: Science, Technology, and Applications* (Taylor & Francis, 2011).
- <sup>2</sup>D. Chatzidimitriou, G. Sinatkas, T. Christopoulos, A. Pitolakis, E. E. Kriezis, and O. Tsilipakos, in *2016 5th International Conference on Modern Circuits and Systems Technologies (MOCASST)* (2016), pp. 1–4.
- <sup>3</sup>O. Tsilipakos, T. V. Yioultsis, and E. E. Kriezis, *J. Appl. Phys.* **106**, 093109 (2009).
- <sup>4</sup>Y.-H. Kuo, Y. Lee, Y. Ge, S. Ren, J. Roth, T. Kamins, D. Miller, and J. Harris, *Nature* **437**, 1334 (2005).
- <sup>5</sup>P. Chaisakul, D. Marris-Morini, M.-S. Rouified, J. Frigerio, D. Chrastina, J.-R. Coudevyille, X. L. Roux, S. Edmond, G. Isella, and L. Vivien, *Sci. Technol. Adv. Mater.* **15**, 014601 (2014).
- <sup>6</sup>G. T. Reed, G. Mashanovich, F. Gardes, and D. Thomson, *Nat. Photonics* **4**, 518 (2010).
- <sup>7</sup>M. Liu, X. Yin, E. Ulin-Avila, B. Geng, T. Zentgraf, L. Ju, F. Wang, and X. Zhang, *Nature* **474**, 64 (2011).
- <sup>8</sup>L. Alloatti, R. Palmer, S. Diebold, K. P. Pahl, B. Chen, R. Dinu, M. Fournier, J.-M. Fedeli, T. Zwick, W. Freude, C. Koos, and J. Leuthold, *Light Sci. Appl.* **3**, e173 (2014).
- <sup>9</sup>T. Christopoulos, G. Sinatkas, O. Tsilipakos, and E. E. Kriezis, *Opt. Quantum Electron.* **48**, 128 (2016).
- <sup>10</sup>O. Tsilipakos, T. Christopoulos, and E. E. Kriezis, *J. Lightwave Technol.* **34**, 1333 (2016).
- <sup>11</sup>D. Chatzidimitriou, A. Pitolakis, and E. E. Kriezis, *J. Appl. Phys.* **118**, 023105 (2015).

- <sup>12</sup>A. Pitolakis, D. Chatzidimitriou, and E. E. Kriezis, *Opt. Quantum Electron.* **48**, 243 (2016).
- <sup>13</sup>G. V. Naik, V. M. Shalaev, and A. Boltasseva, *Adv. Mater.* **25**, 3264 (2013).
- <sup>14</sup>K. Ellmer, *Nat. Photonics* **6**, 809 (2012).
- <sup>15</sup>P. West, S. Ishii, G. Naik, N. Emani, V. Shalaev, and A. Boltasseva, *Laser Photonics Rev.* **4**, 795 (2010).
- <sup>16</sup>D. C. Zografopoulos and K. P. Prokopidis, *Phys. Rev. Appl.* **2**, 064009 (2014).
- <sup>17</sup>E. Feigenbaum, K. Diest, and H. A. Atwater, *Nano Lett.* **10**, 2111 (2010).
- <sup>18</sup>S. Zhu, G. Q. Lo, and D. L. Kwong, *Appl. Phys. Lett.* **99**, 151114 (2011).
- <sup>19</sup>Z. Lu, W. Zhao, and K. Shi, *IEEE Photonics J.* **4**, 735 (2012).
- <sup>20</sup>V. E. Babicheva, A. Boltasseva, and A. V. Lavrinenko, *Nanophotonics* **4**, 165 (2015).
- <sup>21</sup>A. P. Vasudev, J.-H. Kang, J. Park, X. Liu, and M. L. Brongersma, *Opt. Express* **21**, 26387 (2013).
- <sup>22</sup>H. Zhao, Y. Wang, A. Capretti, L. D. Negro, and J. Klamkin, *IEEE J. Sel. Top. Quantum Electron.* **21**, 192 (2015).
- <sup>23</sup>C. Huang, R. J. Lamond, S. K. Pickus, Z. R. Li, and V. J. Sorger, *IEEE Photonics J.* **5**, 2202411 (2013).
- <sup>24</sup>V. E. Babicheva, N. Kinsey, G. V. Naik, M. Ferrera, A. V. Lavrinenko, V. M. Shalaev, and A. Boltasseva, *Opt. Express* **21**, 27326 (2013).
- <sup>25</sup>C. Lin and A. S. Helmy, *Sci. Rep.* **5**, 12313 (2015).
- <sup>26</sup>V. E. Babicheva and A. V. Lavrinenko, *Opt. Commun.* **285**, 5500 (2012).
- <sup>27</sup>A. Melikyan, N. Lindenmann, S. Walheim, P. M. Leufke, S. Ulrich, J. Ye, P. Vincze, H. Hahn, T. Schimmel, C. Koos, W. Freude, and J. Leuthold, *Opt. Express* **19**, 8855 (2011).
- <sup>28</sup>A. V. Krasavin and A. V. Zayats, *Phys. Rev. Lett.* **109**, 053901 (2012).
- <sup>29</sup>H. W. Lee, G. Papadakis, S. P. Burgos, K. Chander, A. Kriesch, R. Pala, U. Peschel, and H. A. Atwater, *Nano Lett.* **14**, 6463 (2014).
- <sup>30</sup>U. Koch, C. Hoessbacher, J. Niegemann, C. Hafner, and J. Leuthold, *IEEE Photonics J.* **8**(1), 4800813 (2016).
- <sup>31</sup>T. Amemiya, E. Murai, Z. Gu, N. Nishiyama, and S. Arai, *J. Opt. Soc. Am. B* **31**, 2908 (2014).
- <sup>32</sup>S. Zhu, G. Q. Lo, and D. L. Kwong, *Opt. Express* **22**, 17930 (2014).
- <sup>33</sup>V. J. Sorger, N. D. Lanzillotti-Kimura, R.-M. Ma, and X. Zhang, *Nanophotonics* **1**, 17 (2012).
- <sup>34</sup>H. Kim, M. Osofsky, S. M. Prokes, O. J. Glembocki, and A. Pique, *Appl. Phys. Lett.* **102**, 171103 (2013).
- <sup>35</sup>I. Hamberg and C. G. Granqvist, *J. Appl. Phys.* **60**, R123 (1986).
- <sup>36</sup>D. L. Wood, K. Nassau, T. Y. Kometani, and D. L. Nash, *Appl. Opt.* **29**, 604 (1990).
- <sup>37</sup>R. Soref and B. Bennett, *IEEE J. Quantum Electron.* **23**, 123 (1987).
- <sup>38</sup>M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, *IEEE Photonics J.* **3**, 1171 (2011).
- <sup>39</sup>R. Pierret, *Semiconductor Device Fundamentals* (Addison-Wesley, 1996).
- <sup>40</sup>S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer Science & Business Media, 2012).
- <sup>41</sup>S. M. Sze and M.-K. Lee, *Semiconductor Devices: Physics and Technology*, 3rd ed. (Wiley, 2012), p. 548.
- <sup>42</sup>Y. Ohhata, F. Shinoki, and S. Yoshida, *Thin Solid Films* **59**, 255 (1979).
- <sup>43</sup>Y. Shigesato, in *Handbook of Transparent Conductors*, 1st ed., edited by D. Ginley, H. Hosono, and D. C. Paine (Springer US, 2011), Chap. 5, p. 160.
- <sup>44</sup>A. K. Kulkarni and S. A. Knickerbocker, *J. Vac. Sci. Technol. A* **14**, 1709 (1996).
- <sup>45</sup>N. D. Arora, J. R. Hauser, and D. J. Roulston, *IEEE Trans. Electron Devices* **29**, 292 (1982).
- <sup>46</sup>G. D. Wilk, R. M. Wallace, and J. M. Anthony, *J. Appl. Phys.* **89**, 5243 (2001).
- <sup>47</sup>M. Fox, *Optical Properties of Solids*, 2nd ed. (Oxford University Press, 2010), p. 192.
- <sup>48</sup>R. Kim and M. Lundstrom, see <https://nanohub.org/resources/5475> for “Notes on Fermi-Dirac Integrals (3rd edition)” (2008).