Tunable all-dielectric metasurface for phase modulation of the reflected and transmitted light via permittivity tuning of indium tin oxide

Ali Forouzmand, Mohammad Mahdi Salary, Ghazaleh Kafaie Shirmanesh, Ruzan Sokhoyan, Harry A. Atwater and Hossein Mosallaei*

Abstract: We propose an electrically tunable metasurface, which can achieve relatively large phase modulation in both reflection and transmission modes (dual-mode operation). By integration of an ultrathin layer of indium tin oxide (ITO) as an electro-optically tunable material into a semiconductor-insulator-semiconductor (SIS) unit cell, we report an approach for active tuning of all-dielectric metasurfaces. The proposed controllable dual-mode metasurface includes an array of silicon (Si) nanodisks connected together via Si nanobars. These are placed on top of alumina and ITO layers, followed by a Si slab and a silica substrate. The required optical resonances are separately excited by Si nanobars in reflection and Si nanodisks in transmission, enabling highly confined electromagnetic fields at the ITO-alumina interface. Modulation of charge carrier concentration and refractive index in the ITO accumulation layer by varying the applied bias voltage leads to 240° of phase agility at an operating wavelength of 1696 nm for the reflected transverse electric (TE)-polarized beam and 270° of phase shift at 1563 nm for the transmitted transverse magnetic (TM)-polarized light.

Independent and isolated control of the reflection and transmission modes enables distinctly different functions to be achieved for each operation mode. A rigorous coupled electrical and optical model is employed to characterize the carrier distributions in ITO and Si under applied bias and to accurately assess the voltage-dependent effects of inhomogeneous carrier profiles on the optical behavior of a unit cell.

Keywords: all-dielectric nanoantennas; phase modulators; metasurfaces; electro-optical materials; silicon; indium tin oxide.

1 Introduction

The realization of compact and fast spatial light modulators, which have been unavailable to date, will enable a wide range of photonic functions and device applications in the near-infrared (NIR) wavelength range. Recently, metasurfaces (i.e. two-dimensional forms of metamaterials) have emerged as a transformational concept in the field of photonics [1–5]. In contrast to complex three-dimensional metamaterials, metasurfaces are ultrathin patterned surfaces that can offer greater ease of fabrication via a single planar nanolithography step. Passive metasurfaces, which are composed of nonuniform elements with different geometries, physical dimensions, and orientations, can impart spatially varying amplitude and phase delay to scattered light. However, passive metasurfaces are unable to achieve temporal tunability, as the arrangement of their subwavelength nano-scale elements is fixed. The criteria to realize ideal adaptive beam tailoring include a short response time, low energy consumption, wide tuning range, and control over each individual metasurface element (pixel-by-pixel reconfiguration). The first steps toward accomplishing this have led to several intriguing techniques such as mechanical actuation and usage of elastic forces [6, 7], optical pumping to generate...
nonlinear and superconducting effects [8], external stimulation of phase-transition materials [9–14], ionic transport [15], and deployment of electro-optical field effect modulation [16–37]. Among them, metasurfaces hybridized with tunable CMOS-compatible materials, including vanadium dioxide (VO$_2$) [9–11], germanium-antimonytellurium (GST) alloys [12–14], highly doped semiconductors, and indium tin oxide (ITO) [16–27] have enabled an approach to achieve strong light-matter interaction with ultrafast control over the optical properties of individual unit cell elements, which in aggregate enable realization of tunable spatially varying phase/amplitude gradients. Phase-change materials such as VO$_2$ and GST enable large-volume index modulation, but their switching is limited in speed (table 1 in Ref. [9]) and generally requires larger power consumption [38] compared to ITO with a field effect modulation scheme.

ITO has attracted interest as an active electro-optical material because of a large refractive index change and tuning range, possessing an epsilon-near-zero (ENZ) frequency regime, where its permittivity alters sign from positive to negative, and short response time at NIR wavelengths including particularly the 1.55 μm telecommunication wavelength [28, 29]. The spatial area in which the dielectric permittivity of ITO is modulated under external bias voltage is very small, requiring resonant excitation of highly confined fields in the active layer for tunable ITO metasurfaces. This places stringent requirements on the properties of the constituent materials and the geometrical shape of the unit cell in order to achieve the desired tunability function. In Refs. [28–32], it was observed that the integration of ITO into plasmonic nanostructures (e.g. metal-insulator-metal topologies) can offer wide phase modulation in the reflection mode where the geometrical resonance of the structure coincides with a zero crossing in the permittivity of ITO active layer (so-called on-resonance operation). In Ref. [33], it was reported that employing a metal-insulator-semiconductor (MIS) structure hybridized with ITO can improve the reflection efficiency of reflective tunable metasurfaces by electrically merging the excited electric and magnetic resonances, also enabling operation in an off-resonant regime. Although there has been a remarkable progress toward active phase modulation in the reflection mode [28–33], there has not been any demonstration of real-time phase control in the transmission mode via an ultrathin flat metasurface based on ITO permittivity modulation. In our earlier work [34], the proposed transmissive metasurface is based on ITO-integrated multimaterial nanowires, which is non-planar, and its fabrication may not be very favorable. Also, there is a recent work utilizing metallic slits filled with ITO [35], which has the drawback of several wavelengths thickness. It is worth mentioning that reflection/transmission amplitude modulation with a high modulation depth was reported in ITO-integrated nanostructures by coupling the incident beam to guided mode resonances [36, 37] or employing Huygens modes in dielectric resonators [38]; but these structures are not able to perform phase modulation.

In this paper, we present for the first time, an ITO-integrated all-dielectric metasurface for phase modulation in the transmission mode. In addition, while tunable metasurfaces were only able to perform phase modulation over a single reflection or transmission mode at a predesigned operating wavelength, this tunable ultrathin metasurface can separately control the transformations of both reflected and transmitted beams in real time at two different wavelengths (dual-mode operation). To realize such a design, the selection of constituent materials, the geometrical shape of nanoresonators, and optimization of structural parameters of nanostructure need to be carefully addressed. The proposed configuration is based on the integration of an ITO layer into an all-dielectric metasurface. This ITO-loaded semiconductor-insulator-semiconductor unit cell is composed of silicon (Si) nanodisk antennas located on top of an alumina-ITO stand-off layer followed by a Si back slab and a silica substrate. In order to electrically control the metasurface elements, the nanodisks are connected together via Si nanobars (Figure 1A). We performed a parametric study to define a fixed set of structural parameters that can support strongly confined electromagnetic modes coupled with the carrier density modulation of ITO at operating wavelengths of 1696 nm and 1563 nm in reflection and transmission modes, respectively, under illumination with transverse electric (TE) and transverse magnetic (TM)-polarized incident light. In the reflection mode, the high-index Si nanobars strongly interact with the TE-polarized incident beam (parallel to the nanobars) and induce a significant field enhancement below the biasing lines at the location of the ITO layer. In the transmission mode, the Si nanodisk is responsible for exciting the desired optical mode leading to field confinement in the ITO layer under illumination with TM-polarized impinging light. To assess both optical and electrical performances of the unit cell in a unified manner, the voltage-dependent carrier distribution profiles corresponding to ITO and Si are calculated via electrostatic device physics simulations, and the results are taken as input into the optical simulation to precisely attain reflection/transmission responses under applied bias voltages. Depending on the carrier density in the ITO active layer, the reflection and transmission phases can
be tuned in the vicinity of the resonant wavelengths of 1696 nm and 1563 nm, respectively, leading to phase variations of more than 240° and 270°. This is a key mechanism for the realization of the first tunable dual-mode phase modulator. Our design features a relatively small unit cell period (<λ/2) and a simple one-dimensional biasing network to realize actively controllable phase profiles at the metasurface interface. Based on this principle, the conventional single-mode (reflective or transmissive) metasurfaces capable of wavefront engineering [39–42],
manipulation of polarization states [43–45], and holography [46–48] can undergo a paradigm shift to operate as dual-mode structures in both reflection and transmission. This tunable platform can offer independent control over the phase distributions of reflected and transmitted beams, which can also attain interesting functions, such as beam steering and dynamic focusing.

2 Results and discussions

2.1 Geometry and structural parameters of the ITO-loaded dual-mode unit cell

Figure 1 illustrates the proposed electrically tunable dual-mode metasurface capable of performing as an active phase modulator. The design features a semiconductor-insulator-semiconductor (SIS) structure consisting of an array of subwavelength Si nanodisks with radius, height, and periodicity of $R=320$ nm, $h_{\text{disk}}=310$ nm, and $P=700$ nm, surrounded by air and placed on an alumina ($\text{Al}_2\text{O}_3$) gate-dielectric layer ($h_{\text{Al}_2\text{O}_3}=12$ nm), an ITO electrooptical active layer ($h_{\text{ITO}}=24$ nm), a high-index Si back slab ($h_{\text{Sub}}=200$ nm), and a substrate of silicon dioxide ($\text{SiO}_2$). The Si nanodisks are connected to each other by Si nanobars (which serve as biasing lines) with a width of $w=100$ nm oriented along the $x$-axis. This facilitates implementation of a biasing network capable of controlling the bias voltage over each subwavelength element from the side without a complex two-dimensional biasing scheme. Also, the Si nanobars have an important role to enhance the light-matter interaction when the light polarization is along the biasing lines (further discussion is provided in Section 2.2). As shown in Figure 1A, B, the $\text{Al}_2\text{O}_3$-ITO stand-off layer is placed between the high refractive index Si nanodisk/nanobar antennas and the Si back slab.

The ITO layer is degenerately n-type doped with a background carrier concentration of $3 \times 10^{20}$ cm$^{-3}$. The top Si nanodisks and nanobars are also highly n-type doped with a background carrier concentration of $1 \times 10^{19}$ cm$^{-3}$ in order to serve as bias electrodes for applying voltage and forming a carrier accumulation at the ITO-$\text{Al}_2\text{O}_3$ interface. The Si-$\text{Al}_2\text{O}_3$-ITO heterostructure can be regarded as a non-ideal parallel plate capacitor with degenerately doped ITO and highly doped Si acting as semiconductive plates in which the induced charges are equal and of opposite signs in accordance with the charge conservation. When an external bias voltage is applied between the highly doped Si nanodisks/nanobars and ITO layer, electrons accumulate at the ITO-$\text{Al}_2\text{O}_3$ interface (as depicted in Figure 1D), forming an accumulation layer in which carrier density increases with the applied voltage. The variation of charge carriers in highly doped Si nanodisks/nanobars has two regimes, which are demarcated by a threshold voltage ($V_T$). For applied voltages $V < V_T$, electron charge carriers are depleted away from the Si-$\text{Al}_2\text{O}_3$ interface, forming a depletion layer whose thickness increases with the applied bias prior to reaching $V = V_T$. By increasing the applied bias voltage beyond this threshold voltage, the thickness of the depletion layer remains almost constant, while holes start to accumulate at the Si-$\text{Al}_2\text{O}_3$ interface, and the Si surface conductivity is inverted from n-type to p-type. The spatial distribution of carrier densities within the accumulation and depletion layers of ITO and Si can be calculated for different applied voltages by self-consistently solving the Poisson and drift-diffusion equations, using electrostatic simulation solvers, e.g. the Lumerical device simulator [49]. Further details regarding the electrostatic device physics simulations are provided in Section 1S of the Supplementary Materials.

The electron distribution in the ITO active layer is plotted in Figure 1E as functions of applied bias voltage and distance from the ITO-$\text{Al}_2\text{O}_3$ interface. As shown in Figure 1E, increasing the external bias voltage from 0 V to 12 V enables the carrier concentration to be increased from $3.25 \times 10^{20}$ cm$^{-3}$ to $9.86 \times 10^{20}$ cm$^{-3}$ at the ITO-$\text{Al}_2\text{O}_3$ interface. The slight accumulation at the ITO-$\text{Al}_2\text{O}_3$ interface at no bias is due to the band-bending effect induced by the difference between work functions of doped Si and ITO. It should be noted that the band bending is negligibly small at the interface of ITO with Si back slab due to the fact that the Si back slab is undoped, and its work function is quite close to the work function of degenerately doped ITO. The accumulated carrier densities exhibit an exponentially decaying profile as a function of distance from the ITO-$\text{Al}_2\text{O}_3$ interface such that the effective thickness of the accumulation layer is less than $\approx 1$ nm. The electron and hole carrier distributions inside the Si active layer are also shown in Figure 1F and G, respectively, as functions of applied bias voltage and distance from the Si-$\text{Al}_2\text{O}_3$ interface. The threshold voltage of $V_T=4$ V can be clearly identified from the results by the voltage at which the depletion layer thickness has reached a maximum and beyond which the hole carriers are generated and accumulated at the Si-$\text{Al}_2\text{O}_3$ interface. Similar to electron densities inside the ITO active layer, the hole densities inside Si also attenuate exponentially by moving further from the Si-$\text{Al}_2\text{O}_3$ interface. Our electrostatic calculations show that when the applied bias voltage exceeds $V_{\text{th}}=10.5$ V, the electric field inside the $\text{Al}_2\text{O}_3$ layer exceeds
the breakdown field of 74 MV/cm reported in Ref. [32]. The capacitance per unit area for this unit cell is extracted from the Numerical device simulations, which gives a maximal value of \( C_{\text{unit cell}} = 6 \, \text{fF}/\mu\text{m}^2 \) (corresponding to the highest applied bias voltage). Moreover, the resistivity of n-doped Si with a background carrier concentration of \( N = 1 \times 10^{19} \, \text{cm}^{-3} \) is calculated as \( \rho = 0.0076 \, \Omega \, \text{cm} \) [50], which results into a resistance of \( R_{\text{unit cell}} = 505.7 \, \Omega \) per unit cell. Assuming a total length of \( L = 21 \, \mu\text{m} \) for the metasurface along the x-axis (consisting of 30 unit cells), it will yield a total capacitance of \( C = 58.98 \, \text{fF} \) and a total resistance of \( R = 15.17 \, \text{K\Omega} \). As such the metasurface will have an \( RC \) time constant of \( \tau = RC = 0.89 \, \text{ns} \) and a frequency bandwidth of \( 1/(2\pi RC) = 177 \, \text{MHz} \). It should be emphasized that implementing the metasurface with a smaller footprint is expected to enable larger modulation frequencies. For instance, when \( L = 7 \, \mu\text{m} \), the modulation frequency increases up to \( 1.6 \, \text{GHz} \).

Following the device physics simulations, the spatial carrier distributions inside the active regions of the structure are translated to the spatial distributions of permittivity via dispersion models to be considered in the optical simulations. In this work, the refractive indices of undoped Si, Al\(_2\)O\(_3\), and SiO\(_2\) are derived from the experimentally measured data by Palik [51]. The dielectric permittivity of ITO is quantitatively described by a Drude model as

\[
\varepsilon_\infty = \varepsilon_\infty^0 + \frac{\omega_p^2}{\omega^2 + \omega_0^2},
\]

where \( \varepsilon_\infty \) is the infinite frequency permittivity, \( \omega \) is the angular frequency, \( \omega_0 \) is the collision frequency, and \( \omega_p \) is the plasma frequency. The plasma frequency can be related to the electron carrier density \( N \), electronic charge \( e \), and electron effective mass \( m^* \) as \( \omega_p = \sqrt{Ne^2/(\varepsilon_\infty^0 \epsilon_0 m^*)} \). Throughout this manuscript, the ITO optical parameters are defined to be \( \varepsilon_\infty = 3.9, \Gamma = 180 \, \text{THz}, \omega_p = 0.35 \times m^* \omega_0 \), where \( m_0 \) is the electron rest mass, and \( \omega_0 = 1.6517 \times 10^{15} \, \text{rad}/\text{s} \) (the background carrier concentration is \( 3 \times 10^{20} \, \text{cm}^{-3} \)), which are consistent with Refs. [29–33]. The carrier-induced changes of the complex dielectric permittivity of highly doped Si nanodisks/nanobars are taken into account using a Plasma-Drude model as

\[
\varepsilon_{\text{doped-Si}}(\omega) = \varepsilon_{\text{undoped-Si}}(\omega) - \frac{e^2}{\varepsilon_\infty^0} \left( \frac{N}{m^*_{\text{Dr}} + i\epsilon_\mu_N} + \frac{P}{m^*_{\text{Dr}} + i\epsilon_\mu_P} \right),
\]

where \( N(P) \) is the electron (hole) density, \( \mu_N(\mu_P) \) is the electron (hole) mobility, and \( m^*_{\text{Dr}}(m^*_{\text{Dr}}) \) is the electron (hole) conductivity effective mass. Here, we have considered \( \mu_N = 80 \, \text{cm}^2 \, \text{V}^{-1} \, \text{s}^{-1}, \mu_P = 60 \, \text{cm}^2 \, \text{V}^{-1} \, \text{s}^{-1}, m^*_{\text{Dr}} = 0.27 \times m_0, \) and \( m^*_{\text{Dr}} = 0.39 \times m_0 \) [52].

It should be remarked that despite the fact that the active layer of Si has a larger thickness compared to that of ITO, the rate of changes in the complex permittivity of Si is much smaller in comparison with that of ITO. In particular, the permittivity of ITO at the ITO-Al\(_2\)O\(_3\) interface changes from \( \varepsilon_{\text{ITO}} = 1.95 + 0.29i \) at no bias to \( \varepsilon_{\text{ITO}} = -2.04 + 0.88i \) at 12 V, at a wavelength of 1.55 \( \mu\text{m} \) (an almost two-fold decrement in the real part of permittivity in a region narrower than 3 \( \mu\text{m} \) thick). Moreover, the formation of ENZ region in the ITO layer when the real part of the permittivity is between +1 and –1 enhances the light-matter interactions substantially. This is when the depletion of electrons in Si changes its complex permittivity from \( \varepsilon_{\text{Si}} = 12.03 + 0.005i \) to \( \varepsilon_{\text{Si}} = 12.11 \) at a wavelength of 1.55 \( \mu\text{m} \) (less than 1% increment of permittivity in a region smaller than 20 nm thick). Moreover, the hole accumulation after surface inversion changes the complex permittivity of Si at the Si-Al\(_2\)O\(_3\) interface from \( \varepsilon_{\text{Si}} = 12.11 \) at \( V_s = 4 \, \text{V} \) to \( \varepsilon_{\text{Si}} = 7.65 + 0.28i \) at 12 V, at a wavelength of 1.55 \( \mu\text{m} \) (less than 40% decrement in the real part of permittivity in a region narrower than 1.5 nm thick). Accordingly, the main contributing factor to the tunable optical performance of the structure remains to be charge accumulation inside the ITO layer.

The optical response of the tunable metasurface is investigated utilizing a custom-developed solver based on a rigorous coupled wave analysis (RCWA) [53, 54] method throughout the main manuscript. The active regions of ITO (≈3 nm thick) and Si (≈20 nm thick) are discretized into several homogenous sub-layers to carefully resolve the inhomogeneous permittivity profiles and capture the accurate optical response of the unit cell. The thickness of the sub-layers is chosen to be uniform for ITO, while a non-uniform discretization is adopted for Si depletion layer to minimize the computational cost while yielding a great accuracy. Further details regarding the discretization approach are brought up in Section IS of the Supplementary Materials.

The fabrication of the proposed metasurface can be carried out using multiple depositions and standard e-beam lithography [29, 32, 38]. The Si back slab can be deposited on a quartz glass substrate via sputtering, followed by sputtering of ITO using an RF magnetron sputtering in an oxygen/argon plasma in which the flow rate of argon/oxygen gas can be controlled to achieve different background carrier concentrations in the ITO layer. The alumina gate dielectric film can be then grown on top of the ITO layer using the atomic layer deposition process. The top silicon layer can be subsequently sputtered, and then, boron dopants can be applied to the layer in order to achieve an n-doped silicon layer with desired background carrier concentration. Afterward, the silicon layer can be patterned using an e-beam lithography system by developing a photoresist film to define the pattern followed by reactive ion etching to etch through the silicon layer. The photoresist layer can be eventually removed via lift-off in
acetone. Each row of silicon nanodisks, which are connected by silicon nanobars, can be assigned to a separate external contact pad that allows individual addressing to the silicon nanodisk row by bias application between the Si nanobars and the ITO layer from the side. These integrated electrode pads can be wire bonded to a compact multi-chip module for electrical gating [32]. To simplify the external connections and avoid any complexity of the biasing network, Si nanobars can also be connected electrically in groups (two, three, etc.). Therefore, each group is addressed by a different external electrode [29]. In addition, it is worth mentioning that the distribution of biasing voltages applied to elements can be controlled using a field-programmable gate array (FPGA) in which storing different pre-designed coding sequences allows for switching between different functionalities in real time. Such FPGA-based digital metasurfaces were previously implemented for real-time and dynamic polarization and focusing control [55]. The contact pads and wire bonding to compact chip and circuit board for electrical gating can also be fabricated by deposition of aluminum/gold using e-beam evaporation and using standard e-beam lithography to define the pattern.

2.2 Performance of a dual-mode unit cell in reflection mode: optical modes and physical mechanism

In assessing the optical performance of the proposed unit cell in the reflection mode, we assume normally incident TE-polarized plane waves where the electric field vector is normal to the plane of incidence as illustrated in Figure 1A. The sample is illuminated from the back, that is, from the SiO$_2$ substrate. The reflection amplitude and phase of the ITO-integrated SIS unit cell are plotted in Figure 2A, B in the absence of gate bias. The unit cell exhibits a geometrical resonance with the high-quality factor (Q-factor) at the operating wavelength of $\lambda_{\text{refl}} = 1696$ nm, leading to a large, almost $2\pi$, spectral phase variation in the reflection. The near-field profile, including the y-component of the magnetic field distribution ($\mathbf{H}_y$, color bar) and the normalized electric displacement currents ($E_x$ and $E_y$, white arrows) are sketched in the x-z plane in Figure 2C. Rotation of the electric displacement current loop around the normal component of the magnetic field is a well-known characteristic of magnetic resonances. The confinement of the magnetic field occurs both at the center of the Si nanodisk far from the location of the ITO layer and below the Si biasing lines at the vicinity of the ITO layer. As the strongly confined field inside the nanodisk is located far from the ITO active layer, it will not lead to a notable spectral shift or large phase modulation of the excited geometrical resonance. The electromagnetic fields confined below the Si biasing lines at the vicinity of the ITO layer are able to enhance the effect of slight modification of carrier concentration within the ultrathin ITO accumulation layer and drastically alter the reflection phase response. Figure 2C also indicates that the strength of the geometrical resonance depends on the dimensions of the biasing lines. To see this, the reflection response in the absence of the biasing lines is also shown in Figure 2A, B, which illustrates a weak resonance (low Q-factor with negligible light-matter interaction) accompanied by a smooth spectral phase variation of $\approx \pi$ at the operating wavelength of 1691 nm. In addition, the near-field distribution in the absence of biasing lines is presented in Figure 2D. To summarize, the Si nanobars that electrically interconnect the nanodisk antennas break the geometrical symmetry of the unit cells and give rise to the required resonant reflection in a critical spectral regime for phase modulation (i.e. excitation of strongly confined optical fields at the place of ITO and occurrence of the ENZ phenomenon). Comparing the near-field distributions in Figure 2C and D also confirms the role of the biasing lines in the formation of stronger confined magnetic fields at the position of the ITO active layer. A parametric study of the effects of the structural parameters including the biasing line width (varying from $w = 0$ to 140 nm) on the reflection response is provided in Figure S3 and Section 2.1S of the Supplementary Materials; briefly, we observe that the geometrical resonance becomes stronger (i.e. smaller reflection and wider spectral phase change) when increasing the biasing line width up to $w = 100$ nm, but it has opposite trend for biasing line widths larger than 100 nm. An extended discussion on the optical modes supported by the ITO-loaded SIS unit cell is presented in Section 2.2S of the Supplementary Materials.

The spatial distribution of the dielectric permittivity of the ITO and top Si nanodisk/nanobar active layers can be calculated utilizing their corresponding dispersion models outlined in Section 2.1 as shown in Figure 2E–H for different applied voltages at an operating wavelength of $\lambda_{\text{refl}} = 1696$ nm. The optical characteristics of ITO can be electrically tuned from positive relative permittivity and dielectric nature ($\varepsilon_{\text{ITO}} = 0.57 + 0.38i$ at 0 V) to negative relative permittivity and plasmonic nature ($\varepsilon_{\text{ITO}} = -3.18 + 1.15i$ at 12 V) at the ITO-Al$_2$O$_3$ interface. The zero-crossing of the ITO permittivity ($\varepsilon_{\text{ITO}} = 0 + 0.63i$) occurs when the plasma frequency is equal to $2.224 \times 10^{15}$ rad/s at 1696 nm (equivalent to the carrier concentration of $5.44 \times 10^{20} \text{ cm}^{-3}$), which results in a significant enhancement of light-matter interaction. It should be noted that the intrinsic material loss...
Figure 2: Optical performance of the dual-mode ITO-integrated metasurface unit cell in reflection mode.

Reflection (A) amplitude and (B) phase of the proposed ITO-integrated metasurface illuminated by x-polarized incident beam for the unbiased case in the presence (solid blue lines) and absence (dashed red lines) of the biasing lines. The near-field spatial distributions at the resonant wavelengths of (C) 1696 nm and (D) 1691 nm in the presence and absence of biasing lines, respectively. The color bar shows the y-component of the magnetic field (|H_y|), and the white arrows are related to the normalized electric displacement currents at the x-z plane. The location of the Si nanobars, which serve as biasing lines, is annotated in (C). (E–H) The real and imaginary parts of ITO and Si permittivities within their corresponding active layers at an operating wavelength of 1696 nm as functions of applied bias voltage and distance from their interfaces with Al_2O_3. The region denoted by dashed lines in (E) corresponds to the ENZ region within the ITO active layer. (I, J) Color maps of reflection amplitude and phase as functions of voltage and wavelength. The dashed lines denote the points for which the real part of ITO permittivity, \text{Re}(\varepsilon_{ITO}) at the ITO-Al_2O_3 interface is equal to –1, 0, and 1. Reflection (K) amplitude and (L) phase versus external applied voltage at an operating wavelength of \lambda_{Refl} ≈ 1696 nm. The dielectric breakdown of the insulator occurs for the applied bias values higher than 10.5 V depicted by shadowed regimes in (K, L).
of ITO increases proportionally with increasing accumulation layer carrier density as illustrated in Figure 2F. This dynamic absorption control is the physical mechanism behind the ITO-based amplitude modulator proposed in Ref. [38]. On the other hand, the complex permittivity of Si changes due to the electron carrier depletion and surface inversion processes as shown in Figure 2G, H. The former process leads to a slight increment in the real part of permittivity and decrement of loss within the depletion region formed at the Si-Al$_2$O$_3$ interface (complex permittivity of Si at Si-Al$_2$O$_3$ interface changes from $\varepsilon_{si} = 11.88 + 0.007i$ to $\varepsilon_{si} = 11.97$), which has a negligible impact on the optical response of the unit cell. Conversely, the accumulation of holes at the Si-Al$_2$O$_3$ interface for the applied bias voltages $V > V_T$ decreases the real part of the dielectric permittivity while increasing the loss in the narrow spatial region of up to 1.5 nm away from the Si-Al$_2$O$_3$ interface (complex permittivity of Si at Si-Al$_2$O$_3$ interface changes from $\varepsilon_{si} = 11.97$ at $V_T = 4 \text{ V}$ to $\varepsilon_{si} = 6.64 + 0.36i$ at 12 V).

Figure 2I, J represents color-coded maps of reflection amplitude and phase as functions of wavelength and externally applied voltage, varying from 0 to 12 V. The geometrical resonance shifts to shorter wavelengths with increasing applied bias voltage up to $\approx 4.7 \text{ V}$, and the resonance shifts in the opposite manner (i.e. moving to longer wavelengths) owing to the ENZ phenomenon and formation of negative permittivity associated with the ITO active layer in higher voltages. The carrier modulation at the ITO accumulation layer mainly governs the supported geometrical resonance, and the effect of permittivity decrement in the Si active layer is negligibly small on the optical response due to the non-critical permittivity change (not experiencing ENZ) within an extremely small spatial region. In other words, this distinct spectral swing of the geometrical resonance originates from the large local field enhancement in the ITO accumulation layer, which undergoes the ENZ transition under applied bias. As demonstrated in Figure 2K, L, the proposed active unit cell can achieve a large phase modulation of $\approx 240^\circ$ at an operating wavelength of $\lambda_{opt} = 1696 \text{ nm}$ with applied gate bias lower than the breakdown threshold of 10.5 V. The largest phase swing is obtained around the applied voltage of $\approx 4.7 \text{ V}$ by flipping of the spectral phase profile at the ENZ region (annotated by dashed lines in Figure 2I, J) specifically nearby the zero crossing of the real part of the dielectric permittivity of ITO. This performance is comparable to simulated results for reflection from MIM-based active metasurfaces reported in Refs. [28] and [29]. In Section 3S of the Supplementary Materials, we also investigated the phase modulation of the proposed unit cell with an effective ultrathin homogeneous ITO active layer where the optical properties are only dictated by varying the plasma frequency instead of inhomogeneous accumulation layer.

2.3 Performance of a dual-mode unit cell in transmission mode: optical modes and physical mechanism

We also examined the transmission performance of the ITO-integrated SIS unit cell under illumination of the TM-polarized incident beam where the magnetic field vector is normal to the plane of incidence as shown in Figure 1A. To achieve relatively large electrical modulation of the optical transmission, the confined geometrical resonance should overlap with the ITO accumulation layer, so that a slight modification of the refractive index within the ITO active layer in the ENZ regime will be intensified, and the transmitted beam can experience a large phase pickup. The solid blue lines in Figure 3A, B represent the transmission response of an ITO-loaded SIS unit cell with the same structural parameters as in Section 2.1 ($R = 360 \text{ nm}$, $h_{AlO} = 310 \text{ nm}$, $P = 700 \text{ nm}$, $h_{AlO} = 12 \text{ nm}$, $h_{ITO} = 24 \text{ nm}$, and $h_{Sub} = 200 \text{ nm}$). The transmission response experiences two resonances at the resonant wavelengths of $\lambda_1^{Tran} = 1563 \text{ nm}$ and $\lambda_2^{Tran} = 1641 \text{ nm}$, which have different optical characteristics. Compared to $\lambda_2^{Tran}$, the operating wavelength of $\lambda_1^{Tran}$ contributes to a resonance with high Q-factor and a rapid transmission phase variation as illustrated in Figure 3A, B. The near-field distributions (i.e. the normal component of the magnetic field and the tangential components of the electric field) are calculated in the y-z plane and demonstrated for $\lambda_1^{Tran}$ and $\lambda_2^{Tran}$ in Figure 3C, D, respectively. At both wavelengths, the circulation of the electric displacement currents around the strengthened magnetic field at the center of the unit cell denotes the magnetic nature of the corresponding resonances. In Figure 3C, the confinement of the magnetic field occurs between the Si nanodisk and the finite-thickness Si back slab in the vicinity of the ITO active layer. The comparison of Figure 3C and D can reveal the fact that the strength of the confined magnetic field at $\lambda_1^{Tran}$ is nearly 2.2 times larger than the one at $\lambda_2^{Tran}$. Therefore, the concurrence of a large local field enhancement and the ENZ effect makes $\lambda_1^{Tran}$ a promising candidate to achieve a large phase modulation. Furthermore, the requirement to satisfy the continuity of the normal displacement field at the ENZ active layer leads to significantly enhanced fields. A parametric study of the impact of the different structural parameters on the transmission response and a more detailed analysis of the supported optical modes by the unit cell can be found in...
Figure 3: Optical performance of the dual-mode ITO-integrated metasurface unit cell in transmission mode. Transmission (A) amplitude and (B) phase as a function of wavelength for the unbiased case. The near-field distributions of tunable ITO-loaded SIS unit cell at (C) $\lambda_1^{\text{Tran}} = 1563$ nm and (D) $\lambda_2^{\text{Tran}} = 1641$ nm corresponding to the excited geometrical resonances. The color bar shows the normal component of the magnetic field ($|H_x|$), and the white arrows are related to the normalized electric displacement currents at the $y$-$z$ plane. (E–H) The real and imaginary parts of ITO and Si permittivities within their corresponding active layers at an operating wavelength of 1563 nm as functions of applied bias voltage and distance from their interfaces with Al$_2$O$_3$. The region denoted by dashed lines in (E) corresponds to the ENZ region within the ITO active layer. Transmission (I) amplitude and (J) phase as functions of wavelength and various bias voltages from 0 to 12 V. The dashed lines denote the points for which the real part of ITO permittivity, $\text{Re}(\varepsilon_{\text{ITO}})$, at the ITO-Al$_2$O$_3$ interface is equal to −1, 0, and 1. (K, L) The transmission amplitude and phase at an operating wavelength of $\approx 1563$ nm versus the external DC bias voltage with consideration of the inhomogeneous active layers of ITO and Si.
Sections 2.1S and 2.2S of the Supplementary Materials, respectively.

The spatial distributions of the relative permittivity corresponding to the active layers of ITO and Si are calculated based on the carrier density distributions in Figure 1E–G using their corresponding dispersion models and plotted in Figure 3E–H at the operating wavelength of ≈1563 nm. A remarkable change in the complex permittivity of ITO from \( \varepsilon_{\text{ITO}} = 1.91 + 0.3i \) to \(-2.13 + 0.91i\) can be observed at the ITO-Al\(_2\)O\(_3\) interface by varying the applied bias from 0 to 12 V. At the same time, the complex permittivity of Si maximally changes from 12.02 + 0.005i to 12.1 due to the carrier depletion and alters from 12.1 to 7.56 + 0.28i at the Si-Al\(_2\)O\(_3\) interface by varying the applied bias from \( V_T \approx 4 \) V to 12 V due to the surface inversion. The calculated transmission amplitude and phase of the metasurface as functions of wavelength and applied bias voltage are plotted in Figure 3I, J. The supported resonance slightly shifts to longer wavelengths when increasing the applied bias voltage up to ≈5.5 V. As a consequence of entering the ENZ region (denoted by dashed lines in Figure 3I, J), the zero crossing of the real part of the ITO permittivity, and the dielectric-to-plasmonic transition of the ITO active layer by applying higher voltages, the corresponding resonance behaves in the opposite manner and shifts to shorter wavelengths. It should be mentioned that the spectral shift and phase swing of the transmission resonance exhibits an opposite trend to the reflection resonance as a function of the applied bias voltage. This opposite trend can also be observed as a function of geometrical parameters of the unit cell in the results brought in Section 2.1S of the Supplementary Materials. As shown in Figure 3K, L, the phase modulation coverage of \(-270^\circ\) is realized at the operating wavelength of ≈1563 nm when the amplitude of the applied electrical bias is within the physically possible range of 0 V to 10.5 V. As expected, the maximum phase modulation is realized around the applied bias of ≈5.5 V, where spectral phase profile is flipped because the strong field confinement generated by the geometrical resonance coincides with the ENZ transition of the ITO permittivity (ENZ region is denoted by dashed lines in Figure 3I, J). It should be mentioned that the level of transmission is comparable to the previous works such as Refs. [34] and [35] in which the amplitude of the transmission coefficients was reported around 0.017 at the external voltage of +2 V and nearly 0.006, respectively. In Ref. [34], the geometry is non-planar and more complex compared to the presented design in this work. In Ref. [35], the ITO-loaded metallic slits are geometrically thick (several times larger than the wavelength) with an aspect ratio of 1:1000 (width to height).

In Section 4S of the Supplementary Materials, we also studied the phase modulation of the proposed unit cell at the transmission mode with an effective ultrathin homogeneous ITO active layer instead of an inhomogeneous accumulation layer. We note that the presence or absence of the biasing lines and choice of their dimensions have negligible impacts on the transmission response of the unit cell when the illumination is assumed to be a TM-polarized beam (perpendicular to the biasing lines), as the biasing lines will be effectively transparent in the case of a y-polarized beam. Because the desired optical mode in transmission mode primarily originates in the Si nanodisk, it is expected that almost the same function could be obtained from the unit cell in transmission mode even when the excitation is a TE-polarized beam, but the transmission response will be remarkably perturbed for different assumed biasing line widths, as the electric field is parallel to the Si nanobars and will strongly interact with them.

As an alternative perspective for revealing the underlying mechanism of tunability in reflection and transmission modes of the proposed unit cell, the role of intrinsic material loss, which is mainly related to the ITO is investigated through hypothetically varying the collision frequency (\( \Gamma \)) of the entire ITO layer. It is shown that the modification of \( \text{Im}\{\varepsilon_{\text{ITO}}\} \) has an important role in achieving a considerable wide phase modulation through satisfying the critical coupling condition (so-called impedance matching) [56–61]. In reality, this condition can be provided by strong confinement of the field within the lossy ITO active layer upon undergoing ENZ transition under applied bias voltages. In addition, this study supports the observed opposite trends of spectral phase variation as a function of voltage for reflection and transmission modes. Further discussion is provided in Section 5S of the Supplementary Materials.

Although the metal-insulator-metal structure is the most well-known motif, which can be hybridized with active electro-optical materials in the reflection mode, the proposed SIS unit cell in this manuscript not only can achieve a similar tunable optical performance in the reflection mode but can also act as a transmission phase modulator. As shown in Figure 1, the proposed all-dielectric metasurface is mainly composed of Si, which is of particular interest for on-chip photonic integrated devices. Compared to a gold back mirror, which reflects the incident beam (transmission is zero when the thickness of the back mirror is larger than the skin depth of the constituent material, e.g. gold), a finite high-index Si slab can effectively perform in reflection and transmission modes at two distinct operating wavelengths.
In Section 6S of the Supplementary Materials, the absorption of the proposed unit cell is quantified at both operating modes as an informative optical parameter. In order to clarify the importance of the finite-thickness high-index Si back slab, supplemental studies were carried out for the SIS unit cell in the absence of a Si slab in Section 7S of the Supplementary Materials. Briefly, we observed that although such a configuration can support both electric and magnetic geometrical resonances, the field confinement at the center of the Si nanodisk is weaker, and it is not able to serve as a functional active device.

Also, we need to mention that loss and ultrathin thickness of the active layer are known as common explanations for justifying the imperfections in the ITO-based spatial phase modulators (low and inconstant reflection/transmission level as well as limited phase coverage, less than $2\pi$). Potentially, the reflection/transmission amplitude can be improved by substituting ITO by another transparent conducting oxide (TCO) with lower intrinsic optical loss as discussed in Section 8S of the Supplementary Materials. The inconstant level of amplitudes in reflection and transmission modes is a result of off-resonance operation in order to attain maximal phase modulation. It would be possible to slightly adjust the operating wavelength in both reflection and transmission modes to achieve amplitudes with less considerable variations at the expense of a lower phase agility. The off-resonance operation also helps to obtain a phase modulation profile with smooth gradual change with respect to the applied voltage. This is discussed in Sections 3S and 4S of the Supplementary Materials. The reflection/transmission phase coverage can be improved by applying multigate biasing. For this purpose, the proposed unit cell can be redesigned in a dual-gate biasing configuration similar to the one reported in Ref. [32] by placing two Al$_2$O$_3$ layers on both sides of the ITO layer, which enables the formation of active layers at both the ITO-Al$_2$O$_3$ interfaces and increment of interaction volume of the resonant mode and ITO active layers. In addition, it is possible to utilize Al$_2$O$_3$/HfO$_2$ nanolaminate (HAOL) as the gate dielectric with superior electrostatic characteristics (low leakage current, large DC permittivity, and high breakdown field) compared to Al$_2$O$_3$ in order to achieve higher carrier density at the ITO accumulation layer leading to a larger phase modulation [32].

In Section 9S of the Supplementary Materials, as potential applications of the proposed unit cell, two dynamic dual-mode metasurfaces are studied in-depth in which the functionalities of both transmitted and reflected beams can be separately controlled in real time. It is demonstrated that the phase front of an array of ITO-integrated SIS unit cells can be controlled through the adjustment of the biasing network distribution leading similar or different bending angles in transmission and reflection modes. As another proof-of-concept demonstration, a tunable dual-mode focusing metalens is proposed with the capability of control over the focal lengths of both reflected and transmitted light. Further details are provided in Section 9.1S (beam steering) and Section 9.2S (focusing).

### 3 Conclusion

We report the design of an active all-dielectric phase modulator with independent operation in both reflection and transmission modes. This design, to the best of our knowledge, not only is the first demonstration of transmission phase modulation via an ultrathin flat metasurface employing ITO but also can offer the independent reflection phase modulation (dual-mode operation). To realize this design, an SIS unit cell was hybridized with an ultrathin ITO electro-optical active layer. The proposed unit cell is composed of Si nanodisk antennas integrated on an Al$_2$O$_3$-ITO stand-off layer followed by a Si back slab and a SiO$_2$ substrate. The neighboring Si nanodisks are connected to each other by Si nanobars. The high-index Si nanobars excite the desired geometrical resonance for TE-polarized reflected beam (along the nanobars) and Si nanodisks give rise to the required confined optical mode for TM-polarized transmitted light. The overlap of the confined resonant optical mode with the ITO active layer while experiencing ENZ transition is critical for attaining considerable phase tunability. A coupled electrical and optical modeling is utilized by means of linking the Lumerical device and RCWA solvers in order to quantify the inhomogeneity and voltage dependency of accumulated/depleted carriers in ITO and Si under influence of applied bias voltages and take their effects into consideration for optical characterizations (reflection/transmission response) of the unit cell. By applying an external voltage, this ITO-integrated SIS unit cell can separately achieve phase modulations as large as 240° and 270° in the reflection and transmission modes at wavelengths of 1696 nm and 1563 nm, respectively. The high isolation and independent control of these operating modes enable this tunable optical platform to exhibit distinct functions in reflection and transmission modes. The array architecture is carried out in such a way that the biasing network can address individual elements. This can pave the way toward the realization of large arrays with electrically phase-tunable elements, capable of dynamic beam manipulation and wavefront engineering.
Acknowledgments: A. F., M. M. S., and H. M. would like to acknowledge financial support by the U.S. Air Force Office of Scientific Research (AFOSR), #FA9550-14-1-0349, Funder Id: http://dx.doi.org/10.13039/100000181 and #FA9550-18-1-0354, Funder Id: http://dx.doi.org/10.13039/100000181. G. K. S., R. S., and H. A. A. would like to acknowledge the funding by the U.S. Air Force Office of Scientific Research (AFOSR), #FA9550-16-1-0019, Funder Id: http://dx.doi.org/10.13039/100000181. The authors gratefully acknowledge useful discussions with Phillip Jahelka from the California Institute of Technology.

References


**Supplementary Material:** The online version of this article offers supplementarymaterial(https://doi.org/10.1515/nanoph-2018-0176).