Three-dimensional concentration of light in deeply sub-wavelength, laterally tapered gap-plasmon nanocavities

Giulia Tagliabue,1,2 Dimos Poulikakos,1 and Hadi Eghlidi1,a)
1Laboratory of Thermodynamics in Emerging Technologies, ETH Zurich, Zurich 8092, Switzerland
2Thomas J. Watson, Sr. Laboratories of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA

(Received 29 March 2016; accepted 20 May 2016; published online 3 June 2016)

Gap-plasmons (GP) in metal-insulator-metal (MIM) structures have shown exceptional performance in guiding and concentrating light within deep subwavelength layers. Reported designs to date exploit tapered thicknesses of the insulating layer in order to confine and focus the GP mode. Here, we propose a mechanism for the three dimensional concentration of light in planar MIM structures which exploits exclusively the lateral tapering of the front metallic layer while keeping a constant thickness of the insulating layer. We demonstrate that an array of tapered planar GP nanocavities can efficiently concentrate light in all three dimensions. A semi-analytical, one-dimensional model provides understanding of the underlying physics and approximately predicts the behavior of the structure. Three-dimensional simulations are then used to precisely calculate the optical behavior. Cavities with effective volumes as small as $10^{-5} \lambda^3$ are achieved in an ultrathin MIM configuration. Our design is inherently capable of efficiently coupling with free-space radiation. In addition, being composed of two electrically continuous layers separated by an ultrathin dielectric spacer, it could find interesting applications in the area of active metamaterials or plasmonic photocatalysis where both electrical access and light concentration are required. Published by AIP Publishing.

Plasmonic structures have become essential in efficiently coupling light to entities with deep subwavelength characteristic dimensions such as nanoscale electronic devices or quantum systems.1 An increasing number of designs have been proposed to confine, guide, manipulate, and focus light at the nanoscale, relying on both localized2 and propagating surface plasmons.3 Recently, a combination of both has also been demonstrated.4 In particular, metal-insulator-metal (MIM) structures supporting odd surface plasmon polariton (SPP) modes that do not exhibit any cut-off thickness have been shown to be more effective than single metal-dielectric interfaces in achieving subwavelength light confinement.5 Given the proportionality between the gap-plasmons (GP) mode group velocity and the gap thickness,6 a number of designs have employed MIM waveguides with tapered gap thicknesses to successfully slow down and concentrate the light.6-9 However, realizing three-dimensionally tapered structures requires cumbersome fabrication procedures.8 On the other hand, MIM nanocavities supporting gap plasmons have been used for extreme light confinement.10 Indeed, film-coupled nanoantennas form nanoresonators,11 where MIM waveguide modes are reflected at the terminations of the front metal patch creating a standing wave pattern inside the underlying nanocavity. Contrary to waveguides, GP nanocavities can efficiently couple the far-field radiation to modes in the deep subwavelength region10 with fabrication facility (even with simple colloidal techniques11). However, existing structures to date can efficiently confine the light only in one dimension. Furthermore, the lack of electrical control due to the discontinuity of the front metal layer can be limiting for certain applications, e.g., lasing.12

Here, we propose to exploit exclusively the lateral tapering of the front metallic layer for realizing three-dimensional (3D) concentration of the GP mode. After demonstrating how the dispersion curve of an MIM waveguide can be modified solely by controlling the width of the front metal layer, we develop a semi-analytical model to describe the 3D focusing effect of a tapered waveguide. Subsequently, we demonstrate an electrically continuous, symmetric, and periodic planar design where this concept is used to create deep sub-wavelength nanoresonators which can be directly and efficiently excited by far-field radiation. Being able to confine light into GP nanocavities with effective volumes as small as $10^{-5} \lambda^3$, these structures could find application in situations where strong field confinement is required, such as third-harmonic generation,13 and the control of optical processes in quantum emitters, such as fluorescent molecules14 and 2D semiconductors.15 In addition, enabling electrical access to the system, the proposed approach could be employed in such emerging fields as active metasurfaces16 and plasmonic photocatalysis.17

In an MIM waveguide, the odd GP mode does not present a cut-off thickness of the dielectric spacer,5 and extremely large mode wave-vectors ($\beta$) (as well as extremely low group-velocities, $v_g = \partial \omega / \partial \beta$) can be obtained for dielectric spacers a few nanometers thick. Therefore, most MIM waveguide designs achieve light confinement and focusing exploiting dielectric spacers with tapered thicknesses.5,7,9,10 In addition, it is known that lateral confinement can also be responsible for an increased mode wave-vector.7,18 Nevertheless, only few works in the literature have analyzed and exploited the

a)eghlidim@ethz.ch
latter effect to focus light.\textsuperscript{8,19} We used the Mode Analysis solver in COMSOL to compute the fundamental (odd) GP mode of an MIM waveguide consisting of an infinitely thick silver (Ag) metallic layer, an ultrathin (h$_{SiO_2} = 13$ nm) glass (SiO$_2$) layer and a 100 nm thick gold (Au) stripe of width W$_g$ (Figure 1(a)). The combination of Ag and Au allows to minimize losses while avoiding oxidation issues of the layer exposed to air. The dispersion curves, $\omega$ (frequency) versus $\beta$, for different values of waveguide width (W$_g$) are plotted in Figure 1(b). We observe that a reduction of W$_g$ modifies the dispersion curve of the waveguide, in a similar manner as in reducing the dielectric spacer thickness.\textsuperscript{12} Indeed, at a given frequency, the narrower the waveguide the larger the propagation wavevector $\beta$ and the smaller the group velocity $v_g$. This effect becomes more pronounced for W$_g$ values comparable or smaller than h$_{SiO_2}$. For W$_g$ larger than 50 nm, the dependence of the dispersion curve on W$_g$ becomes negligible, confirming the validity of the infinitely wide strip approximation which is implicitly assumed in the majority of previous reports (e.g., see Refs. 10 and 11). To better quantify the increase in mode confinement for decreasing waveguide width, we computed the effective area (A$_{eff}$) of the GP mode as a function of W$_g$ (Figure 1(c)). Indeed, we observe a reduction of nearly two orders of magnitude in A$_{eff}$ as W$_g$ goes from 100 nm to 2 nm. Moreover, A$_{eff}$ drops dramatically for W$_g$ smaller than 20 nm, and the obtained curve resembles previous results reported in the literature for more complex confinement geometries.\textsuperscript{8} This result suggests the possibility of achieving light focusing in a planar MIM structure by varying solely the waveguide width, an approach which could significantly simplify the fabrication process of these structures.

Figure 1(d) shows the schematics of such a planar, tapered, and ultra-thin MIM waveguide. Upon decreasing the lateral size of the waveguide, three phenomena take place. First, as mentioned above, the propagation wavevector, $\beta$, increases and the mode group velocity decreases. Second, the transversal wavevector, $k_y$, which has a real value in the dielectric and is approximately inversely proportional to W$_g$, increases along the tapered waveguide. This effect can also be observed in Figure 1(a) where the lateral extension of the field profiles decreases with W$_g$. Finally, due to momentum conservation in the dielectric, $k_{0z}^2 = \beta^2 + k_y^2 + (ik_z)^2 = \beta^2 + k_y^2 - k_z^2$, where $\varepsilon$ is the dielectric permittivity of the insulator layer, the magnitude of the purely imaginary wavevector in $z$-direction, $k_z$, grows. The three aforementioned effects, i.e., the increase in the magnitude of $\beta$, $k_y$, and $k_z$, result in a three dimensional compression of the mode. Nevertheless, light coupling into such a waveguide remains a significant issue,\textsuperscript{9} which we will address in the following.

In order to ensure electrical accessibility as well as efficient coupling to free-space radiation, a continuous and periodic MIM structure is considered. Nanosphere lithography (NSL) is a potentially facile-method to create two-dimensional hexagonal networks of tapered MIM waveguides\textsuperscript{20} whose top-view is schematically shown in light orange in Figure 2(a): the two geometrical parameters which define the top pattern.\textsuperscript{21} Based on the calculations reported in Figure 1, we considered a gold top pattern, a silver back reflector, and a thin SiO$_2$ spacer (h$_{SiO_2} = 13$ nm).

We first analyze the tapered waveguide corresponding to the dark orange section in Figure 2(a) as a representative example of the structure discussed in Figure 1(d). From the results of Figure 1(b), we first calculate the odd GP mode effective refractive index as $N_{eff}(\lambda, W_g) = \beta/k_0$, where $\beta$ is the propagation wave-vector (Figure 2(b)) and $k_0$ is the light wavenumber in vacuum. Combining $N_{eff}(\lambda, W_g)$ with the chosen tapered waveguide profile, W(x), it is possible to calculate the increase in $N_{eff}$ along the waveguide for two different values of W$_{neck}$ (solid lines) (Figure 2(c)). While the case with W$_{neck} = 10$ nm shows only 28.5% variation in the effective refractive index along the waveguide, the case with W$_{neck} = 2$ nm experiences a dramatic variation in the effective refractive index of 287%, demonstrating the potential of lateral tapering for nano-focusing.

Contrary to a waveguide which presents a continuum of allowed modes, a two-dimensional symmetric and periodic pattern has discrete resonance modes.\textsuperscript{22} An allowed GP mode consists of counter-propagating waveguide modes\textsuperscript{22} which experience a phase accumulation of a multiple of $2\pi$ along one period of the structure and of $\pi$ along half a period (symmetry). If L/2 is half the length of a spatial period (Figure 2(a)), the phase accumulation of a propagating mode over this distance is given by

$$\int_0^{L/2} \beta(x)dx = 2\pi \int_0^{L/2} N_{eff}(\lambda_x, x)dx. \quad (1)$$

![FIG. 1. Role of width of the metal strip in a GP-based waveguide. (a) Schematic of the simulated MIM waveguide and coordinate system; the odd GP mode in-plane magnetic field (H$_y$) profiles (left) and out of plane electric field (E$_z$) profiles (right) calculated for an MIM waveguide with finite width W$_g$ equal to 50 nm (upper row) and 5 nm (lower row) using the Mode Analysis solver in COMSOL; propagation is along the x axis. (b) Dispersion curves for MIM waveguides with different finite widths W$_g$. (c) Effective area (A$_{eff}$) of the odd GP mode as a function of the waveguide width at $\lambda = 700$ nm. The insets show $|E|^2$ profiles on a logarithmic scale for three waveguide widths (from top to bottom: W$_g$ = 50 nm, 10 nm, and 2 nm). (d) Schematic representation of a tapered MIM waveguide illustrating the three-dimensional concentration effect due to lateral confinement of the GP mode.](image-url)
shown in Figure 2(b). For a fixed value of $L$ equals to
in $W_{\text{neck}}$ translates to an increase in the effective optical
GP mode substantially red-shifts. In other words, a decrease
length $(\lambda)$ remains unchanged in all cases, this effect is to be
attributed solely to the increase in effective refractive index
attributed solely to the increase in lateral confinement of the wave-
the neck sections of the waveguide. In addition, the increase in the antinode along both the $y$- and $z$-directions decreases (due to the increase in $k_y$ and $k_z$) significantly from the broad to the narrow waveguide sections of the waveguide network along all three $x$-, $y$-, and $z$-directions (from left to right, respectively). In agreement with the concept presented in Figure 1(d), the spatial extent of the antinode along both the $y$- and $z$-directions decreases (due to the increase in $k_y$ and $k_z$) significantly from the broad to the neck sections of the waveguide. In addition, the increase in the effective refractive index along the propagation direction (see Figure 2(c)) is responsible for the shrinkage in the $x$-extension of the antinode under the neck.

Given the magnetic nature of the fundamental resonance mode of the MIM network, it thus occurs at a wavelength where this phase is equal to $\pi$. From Equation (1), we observe that the resonance wavelength ($\lambda_{\text{res}}$) depends both on the period of the structure $L$ and on the effective refractive index $N_{\text{eff}}(x)$, the latter being controlled by the waveguide width profile $W(x)$, as shown in Figure 2(b). For a fixed value of $L$ equals to 173 nm, Figure 2(d) shows the integral of Equation (1) for various $W_{\text{neck}}$ values, and the extreme color scale highlights the wavelength at which the integral is equal to $\pi$ (border between white and black areas). We observe that for $W_{\text{neck}}$ values below 10 nm, as $W_{\text{neck}}$ decreases $\lambda_{\text{res}}$ of the excited GP mode substantially red-shifts. In other words, a decrease in $W_{\text{neck}}$ translates to an increase in the effective optical length of the formed waveguides. As the period of the structure, $L$, remains unchanged in all cases, this effect is to be attributed solely to the increase in effective refractive index induced by the increase in lateral confinement of the waveguide (tapering). In this respect, the proposed concept is very general, and one could arbitrarily design a width profile $W(x)$ to engineer a given phase accumulation within a chosen distance $L$. On the other hand, the period $L$ (together with the dielectric spacer thickness) would determine the blue-most wavelength accessible under the infinitely wide strip approximation (where, $W = W_{\text{neck}} \gg h_{\text{SiO}_2}$). In our case, we chose the periodicity $L$ so that the initial resonance condition was located around 670 nm, at a wavelength close to the visible regime but longer that the onset of gold interband transitions.

Figure 3(a) shows the absorption spectra for different values of $W_{\text{neck}}$ upon plane wave illumination obtained from full 3D numerical simulations. The absorption peak identifies the resonance condition of each structure at which light coupling to the structure is most efficient and maximal field concentration can be obtained. We observe that the maximum of the absorption at resonance increases for smaller $W_{\text{neck}}$ values. This is because photons have a higher chance to be absorbed as the effective optical length of the waveguide increases. The change in the resonance wavelength as a function of $W_{\text{neck}}$ obtained from the numerical simulations is also compared to the analytical one (dashed curve in Figure 2(d) represents the numerically obtained resonance wavelengths). The two solutions present a very similar trend apart from a nearly constant offset, the numerical result being blue-shifted compared to the analytical one. Such discrepancy can be attributed to a polarization-induced reduction in the effective value of $L$ in the 3D simulations (see Figure 3(b), the broader magnetic antinode is not centered to the triangle center-of-mass).

Figure 3(b) shows profiles of magnetic field distributions for the resonances indicated with 1, 2, and 3 in Figure 3(a), corresponding to $W_{\text{neck}}$ equal to 10 nm, 6 nm, and 2 nm, respectively. For all cases, due to the constructive interference of two counter-propagating GP modes, within one period of the structure, the considered fundamental resonance mode has three magnetic field antinodes. From the side ($x$-$y$) and front ($x$-$z$) views, it is apparent that, contrary to the case of a uniform width GP nanocavity, these antinodes have different spatial extensions and that under the neck the field is more tightly confined in all three directions compared to the broad section of the structure. Additionally, the degree of confinement increases for narrower necks.

For the case of $W_{\text{neck}}$ equal to 2 nm, Figure 3(c) allows a quantitative comparison of the extension of the magnetic field antinodes under the broad (red curves) and neck (blue curves) sections of the waveguide network along all three $x$-, $y$-, and $z$-directions (from left to right, respectively). In agreement with the concept presented in Figure 1(d), the spatial extent of the antinode along both the $y$- and $z$-directions decreases (due to the increase in $k_y$ and $k_z$) significantly from the broad to the neck sections of the waveguide. In addition, the increase in the effective refractive index along the propagation direction (see Figure 2(c)) is responsible for the shrinkage in the $x$-extension of the antinode under the neck.
$H_{\text{field}}$ and along the $y$- and $z$-axes by the positions at which the $H$-field decays to $1/e$ of its maximum value at the metal-dielectric interface (see Figure 3(c)). The inset of Figure 4 shows the change in the mode extension along the $x$-, $y$-, and $z$-directions as a function of the $W_{\text{neck}}$. We observe that for the $W_{\text{neck}}$ equal to 40 nm, the mode is almost symmetric in the $x$-$y$ plane. Moreover, we notice that the mode extension decreases steadily along all three spatial directions with the most dramatic reduction happening along the $y$-axis. This result confirms the three-dimensional focusing effect which can be obtained through lateral tapering of the MIM pattern.

Figure 4 shows that overall the effective mode volume decreases with $W_{\text{neck}}$, with a one order of magnitude reduction in $V_{\text{eff}}$ when going from $W_{\text{neck}} = 10$ nm to $W_{\text{neck}} = 2$ nm. Smaller dimensions have not been considered due to the emergence of non-local effects as well as fabrication issues. These small features are difficult to achieve with nanosphere lithography; however, they could be potentially realized using planar fabrication techniques such as e-beam lithography or ion tailored mask lithography eventually combined with the use of a mono-crystalline gold film. For a value of $W_{\text{neck}} = 5$ nm, whose fabrication feasibility has already been demonstrated, we obtain $V_{\text{eff}} \approx 3 \times 10^{-4}$ $\lambda^3$. For the more experimentally challenging case of $W_{\text{neck}} = 2$ nm, we obtain $V_{\text{eff}} \approx 10^{-5}$ $\lambda^3$. We foresee that even smaller values could be obtained by reducing the dielectric spacer thickness to sub-10 nm, similarly to the designs reported recently in isolated patch antennas.

So far, in this paper, we assumed a bulk permittivity for all the parts of the top gold structure. However, it is known that the damping constant in the dielectric permittivity in structures with feature sizes smaller than approximately 10 nm increases. A precise estimate of this effect in narrow lines is very challenging, and experimental determination of the dielectric constant of ultra-thin films is not easy due to the breakup of the films into isolated islands. On the other hand, measurements on gold nanoparticles, which are available from the literature, are likely to overestimate the damping constant for a thin film as the confinement in nanoparticles occurs three-dimensionally rather than along a single dimension, which is the case for thin films. In order to approximately assess the effect of increased losses in the presented...
results, we estimated the increase in the damping constant in the thin parts of the tapered structure (which are confined only along one dimension) as one-third of this increase in a spherical particle with a diameter similar to the narrowest $W_{\text{neck}}$. Furthermore, in order to consider a conservative case, we used the dielectric permittivity of the thinnest part (which is the lossiest part) for the whole section of the structure whose width is below 10 nm. From the performed simulation for a structure with $W_{\text{neck}} = 2$ nm, we obtained a new $V_{\text{eff}} \approx 1.4 \times 10^{-5} \lambda^3$. We observe that the increased losses do not change the order of magnitude of the calculated $V_{\text{eff}}$ even for such a conservative calculation and therefore does not invalidate the proposed concept and the reported calculations.

In conclusion, combining the extreme coupling in ultrathin MIM structures with the field concentration capability (both along the direction of propagation and perpendicular to it) of laterally tapered metal strips, and including the symmetry constraints, we achieved deep subwavelength ($V_{\text{eff}} \approx 10^{-5} \lambda^3$), three dimensional light concentration in a planar geometry. Due to the GP nature of the excited modes, the incoming light can be directly coupled into the ultrathin dielectric spacer, the coupling frequency strongly depending on the width profile of the structure. The proposed structure is inherently capable of trapping light in a very small volume within two continuous metal layers. Therefore, substituting the employed glass layer with a gain material or a catalyst would provide an interesting, electrically accessible system where incoming radiation is strongly concentrated in the active material. At the same time, the high thermal conductivity of the metals and the low thermal capacity of the entire structure provide interesting thermal characteristics for thermally demanding applications (e.g., steady state lasers). More generally, the planar tapered ultrathin waveguides can represent a useful paradigm to design and fabricate planar highly integrated optical components for light guiding and focusing at the nanoscale, avoiding the often cumbersome fabrication procedures required to achieve three dimensional designs.

The authors would like to thank Ruzan Sokhoyan for useful discussion over the simulations and Artur Davoyan and Ravishankar Sundararaman for useful discussions over the manuscript. G.T. acknowledges support from the Swiss National Science Foundation, Early Postdoc Mobility Fellowship No. P2EZP2_159101.